Ethics and the Nature of Action

Heine Alexander Holmen

PhD / Dr.Philos.

Institutt for filosofi, idé- og kunsthistorie og klassiske språk / Department of Philosophy, Classics, History of Arts and Ideas

Faculty of Humanities / Humanistisk Fakultet

Universitetet i Oslo / University of Oslo

May / 2011

If there is one question I dread, to which I have never been able to invent a satisfactory reply, it is to the question of what I am doing.

Samuel Beckett, *Molloy*

We are still far from pondering the essence of action decisively enough.

Martin Heidegger, *Letter on Humanism*

Contents

Acknowledgements

First and foremost I would like to express my outmost gratitude to my supervisors, Carsten M. Hansen and Jennifer Hornsby, whose support, advice and philosophical stings have been of tremendous importance. I am very grateful for the Ph.D. scholarship granted me by the Centre for the Study of Mind in Nature (CSMN) that made it possible to work on this thesis and for providing excellent working facilities at the University of Oslo in due cooperation with the Department of Philosophy, Classics, History of Arts and Ideas. I am also grateful to the Faculty of Humanities for the travelling grant that made it possible to visit Birkbeck College, University of London from January to July 2008 and to the Norwegian Research Council (NFR) for the generous grant that supported my research visit at Rutgers University from February 2009 to January 2010.

Thanks are due to my family, friends and colleagues. I would also like to thank everyone participating at the Ph.D. seminar at the philosophy programme in Oslo and the Language and Rationality Seminar at the CSMN. My thoughts are also going to academic staff, administrative staff and graduate students at Rutgers University and at Birkbeck College that involved me in their activities, seminars and discussions, thereby making my visits to these institutions a wonderful experience. More people than I can remember have contributed with their friendship, thoughts and love during my work on this thesis. In particular, I would like to mention Anders Strand, Kari Refsdal, Timothy Chan, Julian Fink, Anders Nes, Nick Allott, Mathias Sagdahl, Eline Busck-Gundersen, Larry Tempkin, Alison Jaggar, Barry C. Smith and Olav Gjelsvik who have all in their own way chipped in at crucial occasions on this journey. Special thanks go to Georg Kjøll for invaluable discussions and for encouragement when needed, and I should not forget to send my thoughts to Chris Saunders for his proofreading towards the end of this process. Last, but not least, I am indebted by the love, encouragement, patience and hard work that Trine Antonsen has generously dedicated to help me complete this thesis making it all bearable.

Abbreviations

1. PLATO

G	*Gorgias*
L	*Laws*
M	*Meno*
R	*Republic*
S	*Sophist*
St	*Statesman*
Th	*Theaetetus*

2. ARISTOTLE

EE	*Eudemian Ethics*
M	*Metaphysics*
MM	*Magna Moralia*
NE	*Nicomachean Ethics*
OS	*On the Soul (De Anima)*
PA	*Posterior Analytics*

3. THOMAS AQUINAS

ST	*Summa Theologica*

4. THOMAS HOBBES

L	*Leviathan*

5. BARUCH SPINOZA

E	*Ethics*

6. DAVID HUME

EPM	*Enquiry Concerning the Principles of Morals*
T	*A Treatise of Human Nature*

7. IMMANUEL KANT

CPR	*The Critique of Pure Reason*
G	*Groundwork of the Metaphysics of Morals*

8. ARTHUR SCHOPENHAUER

FR	*On the Fourfold Roots of the Principle of Sufficient Reason*
WN	*On the Will in Nature*
FPE	*The Two Fundamental Problems of Ethics*
WWR	*The World as Will and Representation*

9. LUDWIG WITTGENSTEIN

OC	*On Certainty*
PI	*Philosophical Investigations*

1

Ethics and the Metaphysics of Action

> [*M*]*ore than one philosopher has held the view that vitally important philosophical consequences can be reached by derivation from the idea of a rational being.* (...) *Now I do not know whether or not any such grand conclusion can be derived* (...) *though I must confess that I have a sneaking hope that they can, and a nagging desire to try to find out.*
>
> H.P. Grice[1]

§1 INTRODUCTION

In Plato's *Republic* it becomes clear, as Socrates puts it, that its inquiry 'concerns no ordinary topic, but the way we ought to live' (*R* 352d).[2] Challenged by the moral indifference and cynicism of the sophist Thrasymachus, Socrates had been asked by Plato's brothers, Glaucon and Adeimantus, to answer why one should be a just person when there are so many apparent advantages to having an unjust disposition. In order to defend justice, Socrates argues that one has reason to embody this virtue and be just. To a large extent moral philosophy – or philosophical ethics (I will use the terms interchangeably) – is still primarily devoted to such a task. A moral philosopher applies reason and argument to answer such worldview-shaping questions about how life should be lived, what is to be done, and how to live together, and the aim is to show, as Robert Nozick says, 'that and how we are better off being moral' (1981, p. 403). Our aim here seems natural: morality may, after all, ask us to make large sacrifices and risk our own well-being. It is only fair to ask for a reason for why one should follow suit. If morality has a legitimate claim on us, we should therefore expect an answer to the question: 'Why should I be moral?' Moral philosophy thus depends on our answering 'whether our more unreflective moral beliefs and motives can withstand the test of reflection,' as Christine Korsgaard puts it (1996, p. 47). Korsgaard dubs this *the Normative Question*. Roughly speaking, it asks why – or, in a sceptical frame of mind, whether – one

[1] Grice (2001, p. 4).
[2] Unless I mention otherwise, all references and citations from Plato are from the Cooper edition of *Plato: Complete Works* (1997).

should or has reason to lead a life of moral virtue.[3] That *is* no ordinary topic. In fact, it can be argued, as Bernard Williams does, that the 'aims of moral philosophy, and any hopes it may have of being worth serious attention' are bound up with the result of such questioning (1985, p. 1). The normative authority of morality depends on there being something in it for us.

Our inquiry in what follows starts from a concern about a specific strategy in dealing with the *Normative Question*, a strategy Kieran Setiya refers to as *Ethical Rationalism* (2007b, p. 14-5). The bearing idea of Rationalism is to derive our commitment to morality from an account of what it is to be rational. For rationalists, the *Normative Question* requires an independent ground or justification for the normative authority of morality and the normative authority of such a justification, they think, must hold independently of the authority of those moral and ethical virtues. The rationalist is a philosophical optimist with respect to this task and believes that we can provide a philosophically satisfactory answer by employing material from action theory. In particular, the rationalist hopes there are features of our practical rationality that allow us to defend the normative authority of morality. The rationalist thus hopes to infer what Paul Grice calls a 'grand conclusion' – which Grice thinks can be 'derived from the concept of a rational being' – by calling attention to what we have practical reasons for doing (2001, p. 4). By so doing the rationalist hopes to answer in the positive the question which, according to Philippa Foot, 'is as old as Plato but which still haunts moral philosophy today' (1994, p. 159), namely, whether our rational capacity grounds a commitment to morality and an ethical life. Their aim is, in a word, to show that the moral 'should' is a secondary reflection from the independent 'should' of practical reason.

What I shall be discussing in this thesis is whether this particularly ambitious way of grounding the normative authority of morality depends on an image of human agency and practical reason which, on reflection, turns out to be nothing more than 'a philosophical mirage', as Setiya puts it (2007b, p. 1). As I shall argue in this chapter, Rationalism may only succeed if the nature of actions and human agency is rich enough to allow us to derive independent normative standards from practical reason. In particular, the rationalist strategy depends on there being something in the nature of what it is to act intentionally that allows us to see what a good or rational action is independently of what it is for an action to be morally good. If that can be done and those standards recommend leading a moral life, one may argue that agents like us are committed to lead a moral life because we are rational agents endowed with practical reason. In this way, we must be moral agents just as much as we are

[3] Bernard Williams discusses this as 'Socrates' question' in his *Ethics and the Limits of Philosophy* (1985, ch. 1). Also see Michael Smith's rendering of the 'moral problem' (1994, ch. 1). Korsgaard discusses the *Normative Question* in *The Sources of Normativity* (1996, ch. 1).

'condemned to choice and action', as Korsgaard puts it (2009, p. 29). A rationalist must give us reason to think that there is something in the essence and nature of intentional action – such as a constitutive component that an action could not lack without failing to be an action in this sense – that commits an agent to good practical standards. In a word, it will be necessary to find something in the nature of actions that commits an agent to the standards that recommend morality. The hope of the rationalist thus depends on what I shall call a normative interpretation of human agency: it must allow a derivation of good standards for action and practical thinking whose normative authority are independent of concerns about morality.

One way of doing this is by accounting for intentional actions as what we do insofar as our behaviour is properly governed and guided by a constitutive aim, principle, or norm, *C*, whose normative authority does not depend on the authority of moral virtue. The relationship between an activity and its constitutive aim, *C*, is such that 'unless you are guided by the [aim] in question, you are not performing that activity at all' (Korsgaard 2009, p. 29). If such a *constitutive aim theory* (or CAT) is correct for intentional agency, it follows that nothing could be an intentional action unless the activity is guided, patterned, or structured by an agent's concern for behaving in accordance with *C* (Setiya 2007b, p. 85). *C* will thus function as a constitutive condition for intentional actions and no actions are intentional unless its agent is guided by a concern for *C*. A *CAT* will therefore treat it as a fundamental constraint on acting intentionally that its agent takes her behaviour to be a good action in the sense that it provides an adequate implementation of *C* in one's situation. The truth of a *CAT* would thus commit us to what is known in the literature as the *Guise of the Good*:

> *The Guise of the Good*: If *A* is Φ-ing for a reason (or intentionally), *A* must see some good in Φ-ing on this occasion.

The rationalist's hope is thus that acting in accordance with some constitutive aim, *C*, is a constitutive component of acting intentionally. In this way, Rationalism depends on a certain understanding of human agency: namely, that the nature of human agency is normatively rich enough to allow this kind of interpretation. As I shall argue below, this makes the rationalist's concern for the *Normative Question* a concern for the nature of action and action theory.

In what follows, I shall critically assess whether the nature of human agency warrants their assumption. The primary aim of the thesis is thus to engage in discussions of action theory and the nature of human agency in order to assess the rationalist's strategy. In particular, I shall discuss an argument by Setiya who concludes that 'the standards of practical

reason *cannot* derive, not even in outline, from the nature of agency or practical thought' (2007b, pp. 69–70). Setiya thinks that there is no room – nor need – for a normative interpretation of the nature of acting intentionally since we may account for the fundamental puzzles of intentional agency without committing ourselves to principles such as the *Guise of the Good* or to there being constitutive aims for action. According to Setiya, the main task for action theory is to address a puzzle that stems from Elizabeth Anscombe's discussion in *Intention*. Anscombe argued that we should define the class of intentional actions as those to which a reason-requesting sense of the question 'Why?' applies and which we explain by reference to an agent's reasons. She then claims that the sense of this question can be further demarcated by noticing that it applies only to actions which its agent is expected to know 'without observation' that she is performing (Anscombe 1957, pp. 13–5). Setiya takes these observations, suitably qualified, to say something true and important about human agency. However, he argues that Anscombe's discussion fails to address the question 'why [these] two marks of intentional action go together': viz. its susceptibility to the question 'Why?' – or an explanation in terms of reasons – and being subject to agents' knowledge (2007b, pp. 22–3).

Setiya believes that the necessary connection between doing something for a reason and knowing what one is doing says something important about the nature of intentional agency: namely, that the nature of acting intentionally does not lend itself to a normative interpretation. All one needs to act for reasons, Setiya argues, is to act reflexively and have a certain epistemic attitude to one's reasons for acting in a way that is sufficient for explaining the special way in which agents know what they are doing. Setiya's strategy against Rationalism is thus to argue that a normative interpretation of human agency is uncalled for since we can account for what it is to act intentionally without invoking the normatively loaded doctrines that rationalists need. If there is no need in action theory for a normative interpretation of human agency, there is little a rationalist may allude to in order to derive independent standards of practical reason. Moral philosophy will, as a result, get nothing from action theory and the rationalist project will collapse.

To reject the rationalist strategy Setiya needs to come up with his own account of acting for reasons where it becomes clear that when 'we decide upon our reasons [it] is not to say that we decide that they are good or decisive reasons, just that we decide that they will be the reasons on which we act' (2007b, p. 61). That is, he must show that there really is an alternative, non-normative interpretation of intentional agency. What Setiya suggests is what I shall refer to as *Minimalist Cognitivism* about intentional action. Minimalism consists of two

reductions: (1) a reduction of intentional actions to behaviour causally sustained by a relevantly qualified intention to so act; and (2) an account of this intention as a special kind of belief whose formation allows an agent to take some *R* as her reason for Φ-ing. This special kind of belief is, according to Setiya, a desire-like belief that one is hereby Φ-ing because of one's belief that *R* (2007b, p. 48).[4] On this story, it is the fact that an agent's desire-like belief sustains her Φ-ing that makes it a case of intentional agency. The first reduction thus commits Setiya to a causal-motivational reductive account of acting intentionally; whereas the latter commits him to what is known as a *Cognitivist* account of intentions.

Setiya motivates this view as the unique best way to account for the puzzle in Anscombe's *Intention*. Allegedly, the two reductions allow him to explain why the two marks of intentional action – viz. its susceptibility to reason-giving explanations and its being subject to agents' knowledge – go together. According to Setiya, Minimalism explains an agent's knowledge by guaranteeing the presence of a suitably qualified belief about what one is doing intentionally and why. This belief is not always knowledgeable, but he takes it to be known often enough to account for the range of cases where agents really know what they are doing. Since intentional actions must include a relevant intention and the intention just is a belief about what one is doing and why, Setiya may hope to home in on the kernel of truth behind Anscombe's original claim that agents must have knowledge in action. If this works, Minimalism can also account for why actions are susceptible to reasons-giving explanations. All agents must have an intention to Φ insofar as she is Φ-ing intentionally. Since having an intention is either to take some *R* as one's reason – or else to be Φ-ing while knowing that one is Φ-ing for no reason in particular[5] – on can derive the application of the 'Why?' question.

The bottom line of his argument against Rationalism is thus the sufficiency of an account of what it is to act intentionally that does not invoke a normative interpretation of human agency. The function of Minimalism is thus to provide Setiya with a crucial lemma for his rejection of Rationalism since its sufficiency counts against any conception that invokes normative principles to explain the nature of agency. Within the minimalist framework there is no need to think that agents must take their reason to be a good reason for Φ-ing. It is thus compatible with the idea that an agent may Φ for a reason *R* without seeing anything good about her Φ-ing. If Minimalism is sufficient, we may therefore conclude that normative

[4] The fact that the belief in question is desire-like is meant to make sure that the attitude has a motivational function much like ordinary conative attitudes such as desires and wants. In other words, practical and epistemic features join together in the desire-like beliefs so that they exhibit opposing directions of fit (both from mind-to-world (conative) and world-to-mind (cognitive)). That some state can exhibit both these directions of fit has been criticized by Michael Smith (1987, pp. 54–6). I return to this in chapter 3.

[5] The qualification is meant to cover the possibility that agents may act intentionally with literally speaking no reason, a possibility that some readers of Anscombe take to be seriously considered in *Intention* (1957, §17). See Setiya (2007b, p. 52).

doctrines – such as the *Guise of the Good* – are not only redundant but false as principles of rational agency. In the wake of Minimalism, there is thus little room for Ethical Rationalism.

Setiya's strategy against Rationalism makes it imperative to occupy ourselves with action theory and, in particular, with the relationship between the two marks of intentional actions. Setiya aims to show that there is no ground in the nature of human agency for conceiving good standards of practical thinking independently of being good as a trait of character. That verdict depends on what it is correct to say about intentional agency. To reach any conclusion on this matter we must know more about what the rationalist needs and what she can in fact get from action theory. I shall spend chapters 2 to 6 discussing what we can get from an account of intentional agency that sufficiently explains Anscombe's two marks of intentional actions. In chapter 2 I discuss agent's knowledge as a condition of adequacy for action theory and argue that this condition favours *Anscombean Theories* of action, or accounts that may derive the necessity of agent's knowledge. In chapters 3 and 4 I discuss Cognitivist versions of such Anscombean Theories – in particular, Setiya's Minimalism – which try to account for agent's knowledge by conceiving intentions in action as a sort of belief. This view, I argue, runs into trouble in a number of ways and should be rejected. In chapter 5 I discuss and criticize a non-Anscombean way of accounting for agent's knowledge. The upshot of these discussions is that there can be neither a cognitivist nor a non-Anscombean answer to our puzzle. I will therefore support a framework where agent's knowledge is distinctively practical in chapter 6 in and then spend chapter 7 discussing its foundation in the metaphysics of knowledge. In a word, I side with Anscombe by arguing that the correct metaphysics of knowledge allows a notion of *practical knowledge*. The result is a thorough rejection of Setiya's Minimalism and his lemma against Rationalism. I devote sections of chapter 6 to discussing where this leaves action theory and the rationalist strategy.

While ethical material does occasionally intrude, most of my discussions will be action-theoretic. I shall therefore spend the rest of this chapter clarifying the rationalist project, what it requires from action theory and locate its main disagreements with Setiya. I begin by attending to ethics and some connections with action theory (§2) before describing the rationalist project in more detail (§3). I go on to expound Setiya's objection to Rationalism (§4) before locating the dispute between Setiya and rationalists as a disagreement about the nature of intentional actions (§5–§7). Finally, I expound Setiya's strategy against Rationalism (§8). Setiya's strategy will then set the course for what is to come.

§2 ETHICS and the TURN to ACTION

Back in 1958 Anscombe, in a now classic paper, recommends her fellow philosophers to lay moral philosophy aside for awhile, arguing that it is 'not profitable (…) to do moral philosophy' and that we should 'banish ethics totally from our minds', at least 'until we have an adequate philosophy of psychology, in which we are conspicuously lacking' (1958, p. 169). Her idea is to back out of ethics to regroup in moral psychology or action theory since there is 'philosophically (...) a huge gap' which she thinks 'needs to be filled by an account of human nature, human action, the type of characteristic a virtue is, and above all of human flourishing' (1958, p. 193). Anscombe thereby signalled a realignment of intellectual focus from ethics to moral psychology and action theory. Moral philosophy without a sound base in psychology is blind, she warns, and we must reconstitute ethics from philosophy of action to understand and evaluate what different ethical theories, such as Kantianism, Utilitarianism, Consequentialism, Virtue Ethics and so on, really say.[6] In this section, I will grant that Anscombe's recommendation has some point to it. Not that I think we should abstain from doing moral philosophy; however, we should not pursue this inquiry without close attention to moral psychology and action theory.

Suppose we take moral philosophy and ethics as efforts to apply reason to answering worldview-shaping questions, such as how life should be lived, what is to be done, how to live together, and so on.[7] Let us then define ethics as a concern for how one should, all things considered, lead one's life and let us treat the ethical virtues as virtues of character. Assume also that morality is a part of ethics and that the moral virtues are a sub-set of ethical virtues.[8] Ethics is then defined as a putative rational inquiry into the question of how we should lead our lives and it seeks to answer what one should do in one's life as a human being. Towards such an inquiry action theory will prove relevant if its principles and ideas are decisive factors to the issue of how a creature such as us should live. In deciding that, facts about our nature become relevant given that it is a fundamental constraint on the inquiries of ethics that its results must be something that can usefully be applied to creatures like us and thus practised. The normative criteria we work out in ethics – and which are our basis for ascribing moral praise and blame – must in some way or another apply to creatures like us. There must be

[6] Alvin Goldman agrees: 'Ethical theories, such as utilitarianism, cannot be properly assessed without an understanding of the relationships between acts, consequences, circumstances and motives' (1970, p. v). Anscombe, for one, thought that such an inquiry would reveal that the implicit moral psychology of Consequentialism is 'a shallow philosophy' (1958, p. 185). See Michael Stocker (1976) for similar criticism.

[7] See Timothy Chappell (2009, p. 3) and Michael Thompson (2008, p. 5). Socrates, as we saw, pursued the first question (Plato *R* 352d).

[8] I do not have a principled demarcation between morality and ethics and will not take a stand on whether morality is a proper part of ethics or whether morality exhausts ethics. However, I will follow Setiya in thinking that morality is that part of ethics which is essentially other-regarding (2007b, p. 2). Paradigmatic moral virtues are justice and benevolence, whereas courage and chastity may qualify as ethical virtues. For a critical discussion of morality and the moral virtues, see Anscombe (1958), Susan Wolf (1982) and Williams (1985, ch. 10; 2008).

something about the nature of the things we do that make us accountable to these criteria. As Martha C. Nussbaum puts it, our movements are assumed to be constituted in such a way that they 'meet our criteria for the ascription of praise and blame' (2001, p. 265). In this way, ethics as a meaningful enquiry depends fundamentally on implicit assumptions about the nature of our activities and thus relies on moral psychology and action theory.

In this regard, however, ethics as a philosophical discipline echoes our moral practices. When we encounter morally loaded phenomena and events, our evaluations and judgements work against the background of – if not something as sophisticated as a conception or folk theory – so at least an everyday understanding of agency that allows us to see how the agent's knowledge, intentions, motives, decisions, thoughts, and hopes relate in complex manners to the way we evaluate them and their deeds.[9] In our everyday lives, we often try to figure out whether someone's actions and deeds are 'signs of habitual character,' as Schopenhauer said, and thus seek grounds for turning from the deed 'to establish the qualities of the doer' (*FPE* p. 113).[10] An assumption here seems to be that these deeds – as they bespeak our character – are something for which an agent feels responsible and accountable. Being a rascal or a scoundrel is perhaps one's own mistake, as it were. In so doing, we must recognize that the agent's thoughts pertain to the nature of their action in a complex way and that the cognitive state of the agent matters for what we can truly say that they did. In a word, evaluating someone's deeds and character thus requires an adequate way to carve out the nature of what someone did and what they truly are accountable for doing.

In light of this, Anscombe's recommendation can be read as a healthy correction of moral philosophy and a (re-)reminder that our ethical inquiries must be grounded in a sound moral psychology that allows us to see relevant nuances. That gives a certain importance to action theory: given that it is a philosophical investigation that pertains to some of the basic building blocks of the practical normative domain, the inquiries of the philosopher of action may turn out to have far-reaching normative implications. In particular, it will be crucial that the criteria for ascription of praise and blame – criteria that guide our ethical enquiries and our everyday encounters with other people – are met in the nature of human agency. After all, 'the perfectly clear and certain feeling of *responsibility* for that which we do, of *accountability* for our actions,' as Schopenhauer correctly points out, does seem to be 'resting on the unshakeable certainty that we ourselves are the *doers of our deeds*' (*FPE* p. 112). There is

[9] It is such a conception that Jürgen Habermas alludes to when he says that 'we experience and learn *prior to* all philosophy' and thus bring with us as background when we enlist for duty as moral philosophers (1994, p. 185).
[10] For Schopenhauer's *The Two Fundamental Problems of Ethics* I will refer to the 2010 translation of D.E. Cartwright and E. Erdman.

thus clearly, as Nussbaum says, 'a close connection between an account of action and the ethical assessment of persons and lives' and the ethical implications for what we say about the nature of actions – or the theory's 'tendency to support or undermine our evaluative practices' – should be taken into account (2001, p. 265). It then becomes pertinent for philosophers to try, as Jennifer Hornsby says, to 'account for the connection between ethical species of normativity and practical capacities of human beings' (2004a, p. 1). At the very least, the intimate relationship between action theory and ethics thus requires us to say something about the nature of our activities and the way they stand out so as to meet the relevant criteria for ascriptions of praise and blame. There must be something about the way we interact with the world that makes us the doers of our deeds and thus accountable. Without saying something about the nature of this capacity of ours ethics will – however elaborate and sophisticated – hang in thin air. Anscombe was right, then, and it thus seems true, as Alvin Goldman points out, that 'in ethics, most of all, questions about action arise on every front' (1970, p. v).

As will become clear as we move along, to pursue this task is largely about trying to capture something that is unique for creatures like us. It seems plausible to say, as Korsgaard does, that 'both human beings and the other animals act' (2009, p. xi). It also seems plausible that praise and blame apply to creatures insofar as the creature itself is the origin (*archē*) of the motion or action and that this allows room for evaluating the lives of animals and small children (Nussbaum 2001, p. 282).[11] Yet, there seems to be a significant difference here between the way a mature, rational human being is responsible for her life and the accountability of other agents; indeed, as Korsgaard goes on to note, 'human actions can be morally right or wrong, while the actions of the other animals cannot' (2009, p. xi). Korsgaard concludes that '[t]his must be because of something distinctive about the nature of human action, about the way in which human beings make choices'. Creatures like us thus seem unique in being *morally accountable*.[12] This uniqueness in turn ties in with our distinctive capacity for being agents. Question is: what is this capacity? Well, it appears, as Thomas Scanlon says, to be a fact that what we do and think can be evaluated as praiseworthy or not as a function of the 'appropriateness of demanding reasons' (Scanlon 1998, p. 22). There must thus be something about the way we exercise our capacity for agency which, unlike most

[11] As Nussbaum points out, Aristotle grants that animals and small children act voluntary (*hekousion*) (Aristotle *NE* 1111a24-6) and that so acting is enough to warrant a certain notion of praise and blame (Nussbaum 2001, p. 489n62). A *hekousios* action is defined as an action caused by the animal's own desire (*orexis*) for the action and cognitive states concerning the action (Nussbaum 2001, p. 282). Unlike intentional actions – or chosen actions (*prohairesis*) – reason is not necessarily involved in *hekousios* actions. For criticism of this 'dangerous' extension of the voluntary and responsible, see Terrence H. Irwin (1980, p. 124).

Unless I mention otherwise, references to Aristotle are made to the 1984 Barnes edition of Aristotle's *Complete Works*.

[12] Note that this uniqueness in being morally accountable is not equivalent to uniqueness in being someone or something that has a moral status. In my view, there is no need to equate grounds for moral accountability with grounds for direct moral concern.

other creatures, makes us accountable to reason. Such accountability is, after all, a striking feature of the way in which we differ from most other animals; as Brandom says, '[r]eason is as nothing to the beasts of the field' before adding '[w]e are the ones on whom reasons are binding, who are subjects to the peculiar force of the better reason.' (1994, p. 5) Also, Kant emphasized our responsiveness to reasons: our rational capacities allow us to deliberate and act not only in accordance with natural law but in accordance with our conception of such a law (*G* 4:412).[13] Aristotle entertains a similar thought when he ascribes to us a capacity for deliberation, reason and choice (*prohairesis*) while denying it to animals and small children (*EE* 1225b26-8; *MM* 1189a1-4). He then adds that such choices (*prohaireton*) are 'most intimately bound up with' character (and thus to be the ground for judgements of character) (*NE* 1111b5-6, 1112a1-2).[14] A natural suggestion is therefore that we are subject to requirements of morality because we are deliberators, choice makers, or *zoon logikons*. In a word, we are moral beings because we inhabit what Wilfrid Sellars metaphorically dubs 'the space of reasons' and can, at least when we rise to the occasion, choose what we do and think for reasons (Sellars 1991, p. 169). Our actions thus make us morally responsible because they are the actions of 'a creature capable of effective deliberation' (Nussbaum 2001, p. 283).

Again, this suggests a deep connection between ethics and action theory. In action theory a primary object of inquiry – or so I shall argue below – is the kind of action that shows the agent's rationality in operation by being done for reasons. In other words, its focus is on the actions that are chosen (*prohaireton*). Depending on how one accounts for this operation these actions may be taken to reflect, at least in most cases, the agent's judgement that 'a certain reason is worth acting on' (Scanlon 1998, p. 23). The mere fact that such actions get performed may thus be taken to involve judgements of worth and value, at least implicitly, which in turn will make them accountable to 'the canons relevant to them' (Scanlon 1998, p. 22). Thus, the philosopher of action's primary object of inquiry appears to be the performance by which an agent makes herself a legitimate target for moral praise and blame. In a word, its focus is the performance that grounds moral accountability.

In a sense, this is well-known terrain for ethicists. For them it is somewhat customary to connect questions of human rationality in complex ways with questions of value and human good, an idea that goes back as far as Plato's effort to ground ethical disputes in reason and human nature.[15] Most moral philosophers nowadays seem to accept at least some such

[13] References to Kant's *Groundwork on the Metaphysics of Morals* are made to *Practical Philosophy* in *The Cambridge Edition of the Works of Immanuel Kant* (1996), edited and translated by M.J. Gregor.
[14] See Nussbaum (2001, p. 489n62).
[15] See Foot (1994, p. 159) and Warren Quinn (1992, p. 211).

connection. A particularly ambitious way to make this connection, however, is if it turns out that the nature of agency provides ground or justification for ethical normativity. If so, action theory would help answer what is known in the literature as the *Normative Question*. Roughly speaking, the *Normative Question* asks why – or whether – one should be a moral person or live a life of moral virtue. That is, whether it is really true that we have reason to be moral. Socrates in the *Republic* pursued a version of that very question when he concerned himself entirely with 'the way we ought to live' who is called upon to defend why one should lead an ethically virtuous or just life in response to a question posed by Plato's brothers, Glaucon and Adeimantus, about why one should lead an ethically virtuous or just life (*R* 352d). They seem to accept that justice is a moral virtue but nevertheless want an argument in its defence.[16] They demand a reason for being just that pertains to justice itself and not mere fortunate externalities. As Glaucon says, 'I want to know what justice and injustice are and what power each itself has when it's by itself in the soul' (*R* 358d). Socrates deals with this challenge by calling on the nature and proper function of our soul. The idea is to argue that injustice is a sort of cognitive impairment and failure to organize one's soul in accordance with reason and the proper function of its parts (*R* 443d). One has thus reason to be just insofar as one has a human soul. A virtue of this answer is that it allows Socrates to offer a formally egoistic reply to the question of 'Why should *I* be just?' since it would show 'to each person that he has good reason to live ethically' (Williams 1985, p. 32). At the same time, it grounds ethical normativity on what we all can agree on is valuable.[17]

[16] Glaucon and Adeimantus thus play the role of a moral sceptic, or one who questions and remains 'skeptical about the force of ethical considerations' (Williams 1985, p. 25).The sceptic asks why he should be a *Good Dog*, one that 'always does what he ought to do spontaneously and with tail-wagging cheerfulness and enthusiasm', as Korsgaard says (2009, p. 3). It is, however, hard to find such a sceptic. What we usually find are sceptics disputing the authority of specific virtues. A case in mind is Callicles, as he is portrayed by Plato in *Gorgias*, who mockingly disputes Socrates' cherished virtues of character by displaying complete disregard for Socrates' recommendations. Callicles cherishes 'wantonness, lack of discipline, and freedom, if available in good supply, are excellence and happiness'; and as 'for these other things' – viz. 'these fancy phrases, these contracts of men that go against nature' – he claims that 'they're worthless nonsense!' (*G* 492c) According to Foot, we can read Nietzsche's critique of morality similarly as disputing the validity of specific virtues of character. Our cherished virtues are too crude, according to Nietzsche, since they underestimate that 'the true nature of an action depend[s] on *the nature of the individual who did it*' (Foot 2001, p. 110). However, as Foot sardonically remarks, it seems as if Nietzsche would want us to 'probe deep into the psychology of a Mengele or an Eichman before we could evaluate their actions' (2001, p. 113).

A moral sceptic will, on the other hand, question the authority even of something she recognizes as a moral virtue. An advantage of letting Glaucon and Adeimantus *stand in* on behalf of a genuine moral sceptic is that it is controversial whether one really could acknowledge some, *M*, as a genuine virtue without thereby bowing to its authority. John McDowell, in particular, has defended the Socratic idea that knowledge is sufficient for virtue and that consequently 'the attractions of a virtuous life', though certainly real, may be 'recognizable only from within a commitment to that life' (1996, p. 76). A consequence of that view is that one cannot fully recognize and know what the right thing to do is without also possessing this virtue. Since virtues are connected with action and motivation, one seems unable to have this kind of knowledge without also having motivational propensities towards the morally right kinds of actions and ways of being. One could therefore argue that a genuine moral sceptic cannot fully recognize something as a moral virtue without already having motivational propensities – which McDowell describes as 'singleness of motivational focus' (1998e, p. 47) – towards it. However, then one must also ask: how can a moral sceptic possibly challenge the normative authority of this virtue? If he knows it, he already possesses it and is motivated to live in accordance with this virtue. And if he fails to know what it is, then how can he challenge its authority? After all, the sceptic's perception may be just a 'flawed approximation', as McDowell puts it (1998e, p. 47). By standing in for the sceptic Glaucon and Adeimantus can overcome this difficulty by asking the relevant question on behalf of the sceptic. For further defence of the Socratic idea and some of its consequences, see McDowell (1979; 1996; 1998b).

[17] As Williams points out, 'Plato's aim (...) is to give a picture of the self of such a kind that if people properly understood what they were, they would see that a life of justice was a good not external to the self but, rather, an objective that it *must* be rational to pursue' (1985, pp. 33–4). Later Williams adds that 'Plato felt it necessary to discover the ethically significant categories inside human nature, and at the most

My focus in this thesis will be given by the kind of strategy that Socrates pursues when he aims to show that we have reason to lead an ethical life. I shall follow Setiya in referring to this strategy as Ethical Rationalism (2007b, pp. 14–5). Although rationalists differ on many issues, they share the idea that practical reason and intentional agency provide grounds or justification for ethical normativity. If philosophy of action bears on this question, it will thus move beyond mere scholarly disputes to engage with the questions that really count and make us tremble, as Nozick puts it (1981, p. 1). Perhaps we may even come to terms with some of the 'facts the heart can feel', as Camus says, by addressing 'the fundamental question of philosophy' – viz. the question of worth, meaning, value and 'whether life is or is not worth living' (1955, p. 11). Let us see what kind of strategy it is.

§3 ETHICAL RATIONALISM

Ethical rationalists, as I will understand them, seek to justify or ground the normative authority of ethics and morality by showing that our moral accountability is sanctioned by rational requirements and good standards of practical reason. The rationalist's suggestion is that underlying the normative authority of moral virtue are independent conditions of practical rationality. When we come to understand what actions are and the distinctive evaluative standards they stand up to, we are therefore in position to appreciate why we should, all things considered, lead an ethical life. In this respect, the rationalist suggests that philosophy of action is a foundation of ethics. For rationalists, the turn to action and action theory thus becomes acute, but it carries the hope that one may appeal to reason in order to finally make something straight out of the crooked timber of humanity.[18] In this section, I shall expound two important aspects of this project. First, the idea that justifying the normative authority of morality requires sanctioning morality on independent standards of practical reason. Second, the idea that these independent standards spring from the nature of intentional agency so that the recommendation of an ethical life pertains to all intentional agents as a volitional necessity in 'response to an independent normative reality' (Frankfurt 2006, p. 32).[19] I shall also briefly mention some of the merits of this view to see why it deserves serious consideration.

According to Derek Parfit, there are formal aims of moral theories and theories about rationality that most philosophers agree to. On 'moral theories,' Parfit says, 'we ought to try

basic level (...) [namely] in the structure of the soul, at the level of the theory of action itself' in order to respond to the waning normative authority of 'the gods and fate and assumed social expectations' (2008, p. 43). The *Republic* is thus a prime example of how the Socratic concern for the human soul (*epiméleia tes psychés*) goes hand in hand with his Delphic emphasis on self-knowledge (*gnōthi seauton*).

[18] According to Isaiah Berlin, Kant wrote that 'out of the crooked timber of humanity (...) no straight thing was ever made' (1969, p. 39).

[19] For similar characterizations, see Kit Fine (2001) and Ralph Wedgwood (2007, p. 1).

to act morally. According to all theories about rationality, we ought to try to act rationally' (1984, p. 3). Where theories differ, he goes on to say, is in the substantive aims concerning what one must do insofar as one is moral/rational. In philosophy, one often talks about reason in contexts where one wants to draw attention to the active rather than the passive aspect of the human life. This is particularly fitting when it comes to practical reason and reasons for acting because the mental activity needed to initiate and guide intentional activity is active in a way that goes far beyond our capacities for being receptive.[20] If we take human reason to be the faculty or power we use to engage in thinking in general, we may understand the property of being rational as describing the extent to which one is disposed to reason in accordance with the normative standards that pertain to this faculty. Rationality is thus a measure of excellence of this faculty when it embodies virtues of reason and dispositions for good thinking. Without implying compartmentalization in any interesting sense we may then let practical reason – or practical thought – denote the part of our thinking engaging specifically with practical questions and concerns with a view to action. One is practically rational when one's faculty is disposed to function in accordance with good standards of practical thought.[21]

What the rationalist suggests is that the substantive aims of practical reason or rationality, say $R_1,R_2,...,R_n$, that determine the extent to which one is rational, are independent of the normative authority of the substantive aims of morality, $M_1,M_2,...,M_n$, whose satisfaction determines the extent to which one is moral. They then go on to argue that the former standards ground or explain the authority of the latter. That is no minor claim. It may, of course, seem evident that we should do what a morally good person does. As Josef Pieper says, there might be 'no one who needs to be told that he ought to be just and brave and temperate. This is self-evident, and calls for no deliberation' (1966, p. 33). However, what the rationalist commits to is not just that we should live as a virtuous person would, but that the 'should' connected with ethical and moral virtue is grounded in or a secondary reflection from the 'should' of practical reason. Rationalism is thus a view about the relationship between practical rationality, so conceived, and the 'should' that pertains to moral virtue. Like Aristotle, the ethical rationalist therefore links the moral 'should' with the 'should' of practical reason and wisdom (*NE* 1144b31-2). In addition, however, the rationalist believes that the standards of practical reason support the normative authority of moral virtue. The

[20] I do not mean to deny that human receptivity and/or belief-formation is active as opposed to a passive performance. Quite the contrary, I think that Korsgaard is right when she says that '[t]he perceived world does not merely enter the mind, as through an open door' (2008a, p. 207). Presumably, an adequate story about receptivity – i.e. sensation, perception, our emotional sensitivity and so on – must show how these processes and their resulting states can fall to 'the active side of the divide,' as Joseph Raz puts it (1999, p. 15).

[21] See Quinn (1992, p. 210). The excellence of human reason thus comes in degrees: one is more or less rational, more or less irrational.

rationalist thus applies a framework allowing for a possible evaluative divergence between the standards of rationality and the standards of morality, though she hopes to close that gap by showing how the moral 'should' converge with the 'should' of reason. There are thus two separable parts to the rationalist's project: one concerning the rationality of morality; the other the priority of reason over moral virtue. Let me deal with these in turn.

There are different ways one may take morality to be rational. A particularly strong claim would be to say that the morally right thing to do, in any circumstance, is also what there is most reason for doing. James Rachels, for instance, claims that '[t]he morally right thing to do, in any circumstance, is whatever there are the best reasons for doing' (2003, p. 12). However, that would entail the intolerable conclusion that you will never have reason to act against morality and it is plausible that morality is sometimes outweighed by non-moral concerns. It would be awkward if we had to say that someone failed to respond to what she has most reason to do if she lives a good moral life yet one day acts to save her endangered only child rather than two endangered strangers. Morality may require us to save two lives over one; yet it seems perfectly alright to say that in this case she has reason to save her only child, albeit not a moral one. Or consider a morally degenerated individual who, given the kind of person he has become, has reason to perform some despicable deed. It just sounds like moralizing to insist that regardless of what kind of person the agent is or what her situation is the moral option always overrides reasons for doing things incompatible with morality. What the rationalist should argue instead is that there is reason on the whole to lead a moral life:

> *Rational* (Rat): For an agent, *A*, to live a life of ethical and moral virtue is to live as one has most reason, all things considered, to live.[22]

Philosophers may agree with this part of the rationalist's conception of morality without thereby signing up for Rationalism. Philippa Foot, for instance, thinks that 'acting morally *is* part of practical rationality', but she does not think that 'a discussion of the rationality of moral action would start from some theory or other about what a reason for action must be' (2001 pp. 9–10). That brings me to the second part of the rationalist project: the idea that not only is moral conduct rational, but that practical reason grounds morality. Foot credits her change of mind on this issue to an argument run by Warren Quinn against neo-Humean conceptions of practical rationality. Let us briefly rehearse his argument.

[22] *Rat* does not make a choice on the issue of whether a life of moral virtue embodies distinct and irreducible virtues, $V_1, V_2, \ldots, V_n$, or embodies and subsumes all these virtue in one primary master virtue, V^*. See McDowell for a defence of the unity of virtues (1979).

Standard neo-Humean conceptions of rationality are end-neutral and thus morally indifferent. '[T]he goal of practical reason' is nothing but 'the maximal satisfaction of an agent's desires and preference' when these preferences are suitably corrected for effects of internal consistency, or 'misinformation, wishful thinking, and the like', as Quinn argues (1992, p. 210). Beyond such corrections there is no ground or warrant for measuring the reasonability or rationality of one's preferences. In particular, a desire is not judged as less rational if it opposes a conception of what constitutes a valuable preference. It is only when such concerns are expressions of the agent's own preferences that they could provide strictures on what it is rational for an agent to desire via requirements of internal consistency. Beyond internal consistency and misinformation there is thus no ground for arguing that rationality rules for or against some desire or preference. As Hume said, it is not 'contrary to reason to prefer the destruction of the whole world to the scratching of my finger' (*T* 2.3.3).[23]

Quinn's objection goes by way of asking: What, if anything, could be so important about practical rationality if it turns out that it could be most rational to seek to fulfil a truly shameful or despicable desire? Why give practical rationality 'center stage in normative treatments of action and choice' if standards of rationality can recommend pursuing shameful and despicable desires? The problem Quinn is raising is that the end-neutrality and moral indifference of this conception may force us to accept that truly repulsive and morally despicable agents – who are, as Williams puts it, 'horrible enough and not miserable at all but, by any ethological standard of the bright eye and the gleaming coat, dangerously flourishing' (1985, 46) – may nevertheless embody practical excellence by pursuing their despicable preferences.[24] Take, for instance, the historical case of Gustav Wagner, the SS-Oberscharführer known as 'the beast' who ran one of Hitler's death camps in Poland and without remorse contributed to the horrors of the Holocaust in the name of purifying mankind.[25] The neo-Humean conception may have to judge that Wagner instantiates practical excellence. Now it should be granted that such figures may, as Bernard Williams remarks, 'seem sleeker and finer at a distance' and that a closer look may reveal that they are not primary examples of flourishing and psychological well-being (1985, p. 46). There might

[23] For Hume's *A Treatise of Human Nature* I will refer to D.F. Norton and M.J. Norton's 2000 edition.

[24] In Quinn (1992, pp. 215–6) the emphasis is on the claim that the neo-Humean conception of rationality would sometimes judge an agent as most rational by doing something nasty or truly shameful. For this reason Quinn argues that the neo-Humean conception is latent with *shamelessness* in its indifference to the acts and preferences of the shameless character: '[N]othing insulates the shamefulness of the action from the shamelessness of the agent in doing it or – and this is the important point – the shamelessness of his reason in recommending it,' as he says (1992, p. 220). Later Quinn emphasizes similar considerations and argues that 'the only proper ground for claiming that a quality is rational to have or an action rational to do is that the quality or action is, on the whole, good' and that 'the primary questions are not what it is rational or irrational, but what it is good or bad to be, seek, or do' (1993b, pp. 254–5).

[25] See Foot (2001, p. 90). A BBC interview ('The Angel of Death') was made with Gustav Wagner in 1979 where he said that 'I had no feelings... It just became another job. In the evening we never discussed our work, but just drank and played cards.'

therefore be room for some manoeuvre even on the neo-Humean conception. Maybe it turns out that exterminating Jews was not the best way to satisfy Wagner's other desires, or that he misjudged the nature of his own preferences.[26] But however it turns out to be with Wagner, it clearly seems possible to imagine truly despicable agents who are bent on pursuing deep-seated despicable ends. For Quinn, that possibility is a serious embarrassment for a conception of practical reason since it undermines the importance and normative authority of practical excellence. If that is rationality, why bother about being rational, one might ask. The neo-Humean conception is thus in tension with the idea that rationality is a dignifying feat.

According to Foot, Quinn's objection generalizes to question 'whether it is right to think that moral action has to be brought under a pre-established concept of practical rationality' (2001, p. 10). That might be a mistake. Perhaps it is wrong to think that we must 'first come to a theory of rational action, and then try as best we can to slot in the rationality of acts of justice and charity' (2001, p. 10). Perhaps it is better to let human good and bad 'stand at the center of practical thought and not any independent ideas of rationality or reasons for action' (Quinn 1993b, p. 254). As Foot says, she sees Quinn 'as watching us all struggling to tie a horse to the back of the cart, and suggesting "Try it the other way around"' (2001, p. 63). In a word, it might be better to let the good standards of character traits – viz. moral virtue – have explanatory primacy with respect to the good standards of practical rationality.[27]

In this respect, Foot and Quinn disagree with a rationalist over the issue of primacy. For Foot and Quinn, goodness of character grounds the standards for practical thought and reason, whereas for the rationalist the grounding goes the other way around. Their disagreement thus highlights the asymmetric primacy given by rationalists to the nature and standards of practical reason over virtues of character. For rationalists it is therefore important that there is not only a nominal distinction between the 'should' of ethical virtue and the

[26] Wagner ended his own life in captivity, perhaps illustrating, as Foot says, that 'there is always a price to pay for wickedness' (2001, p. 90).

[27] Foot and Quinn thus agree with Hume's scepticism about reason as the foundation of morality, while they disagree with his claim that this undermines the rationality of morality. Whereas Foot and Quinn believe that morality is rational on a moral or virtue-based conception of practical rationality, Hume infers that since we cannot bring and support moral virtue under a pre-established concept of end-neutral rationality, there is no hope of bringing it under a concept of rationality at all. According to Hume, there is a sense in which practical rationality is a misnomer since there is no way in which actions or desires can – 'except in a figurative and improper way of speaking' – be judged reasonable or unreasonable. Hume inferred that reason alone is not adequate to yield moral evaluations: in order to yield moral evaluations a sentiment – such as 'a feeling for the happiness of mankind, and resentment of their misery' – must be added that makes us favour and prefer moral virtues (*EPM* Appendix 1). (References for Hume's *An Enquiry Concerning the Principles of Morals* are to the 1998 T.L. Beauchamp edition). That is to say, in addition to matters of fact and (conceptual) relations – which exhaust the domain of reason on Hume's view – our moral evaluations must involve the contribution of approbation or disapprobation. Gustav Wagner, given his preferences for purifying mankind, is thus just as rational as someone pursuing her preference for feeding the needy. Our reaction to someone like Wagner is therefore moral and sentiment-based rather than a measure of his rationality. In a word, Hume replaced the rationality of morality with a sentiment-based conception of moral virtue. Thereby, Hume is in continuity with modern moral philosophy in its search for a non-moral ground of morality; while be breaks sharply with tradition in giving up the idea that the ground is rational (Rachel Cohon 2010). Foot and Quinn, on the other hand, give up non-moral ground, while retaining faith in the rationality of moral virtue. As Quinn remarks (1992, p. 212n5), the historical Hume may thus be interpreted as accepting the consequence that rationality is potentially shameless and willing to secede its dignity and value. Contemporary proponents of the neo-Humean conception of rationality, however – such as David Gauthier (1986) – do not seem as willing to do this. They should therefore take Quinn's objection seriously.

'should' of practical reason but that this distinction corresponds to something real.[28] We can thus sum up Ethical Rationalism as adding the following claims to *Rat*:

> *Priority* (Pri): The standards of being good as a trait of character, $V_1,V_{2,...,}V_n$ (viz. ethical and moral virtue) are non-circularly justified from good standards of practical reason, *R*, or explained by the nature (or proper function) of practical reason, *R**.
>
> *Nominal Distinction* (Nom-D): *R**/*R* are conceivable independently of $V_1,V_{2,...,}V_n$.
>
> *Real Distinction* (Real-D): *R**/*R* are conceivable independently of $V_1,V_{2,...,}V_n$ and this distinction corresponds with something real.[29]

These principles are formal, so to evaluate them requires specific accounts of what the good standards of practical reason are (viz. provide Parfit's 'substantial aim'). However, the general framework can be assessed as it is. The aim is to conclude in favour of the normative authority of ethical and moral virtue (*Rat*). Following the rationalist strategy, this is done by grounding the authority of morality in practical reason (*Pri*), but to achieve that it is required that the standards of reason really are different from the ethical/moral standards (*Real-D*).

To truly sanction the normative authority of morality one must, of course, accept the rational authority of reasons. Rationalists aim to show that we have reason to be moral; however, one may question whether a commitment to reasons is integral to the subject. Now, that is a fundamental requirement which holds insofar as we are subject to any reason requirements at all. Questioning it thus comes close to absurdity.[30] After all, one cannot consistently request reasons for obeying it. I therefore take it that rationalists have come a long way if the normative authority of morality holds insofar as we accept the rule of reason.

One may criticise rationalism at any of these stages. Perhaps one doubts the existence of practical reason and reasons for acting; or maybe one thinks the standards of practical reason fail to justify the ethical life (denying *Pri*); or that morality is sanctioned on a non-rational basis. Perhaps one may question the intelligibility of non-moral standards of practical rationality (challenging *Nom-D*). Another way to disrupt Rationalism is, as Setiya tries, by

[28] A nominal distinction between *Fs* and *Gs* is a distinction in how one conceives *Fs* and *Gs* that may fail to correspond with a real distinction between them (parallels: a nominal definition may fail to correspond with a real definition that signify the essential nature of what one defines, and two ways of referring to Venus – i.e. Hesperus and Phosphorus – may fail to correspond with a real distinction).
[29] *Nom-D* is a part of *Real-D* but should be kept apart since a way to undermine *Real-D* is by rejecting *Nom-D*. See Foot (2001, p. 62-5).
[30] See David Velleman (2006, p. 19). The alleged absurdity of questioning this principle may perhaps not convince a devoted anti-rationalist. After all, isn't this just empty rhetoric to mask the concession that, as McDowell puts it, '[t]here is *nothing* outside our reasoning on which we could found confidence in its results' (2008c, p. 184)? Well, yes and no. Before celebrating, the sceptic should consider whether a 'conviction of groundlessness' in this area is enough to 'induce panic', as McDowell puts it (2008c, p. 184). The authority of reasons may be the limit at which all attempts at grounding are thoroughly misguided. Asking for a ground here may simply be a gross mistake.

arguing directly against *Real-D*. By so doing, one may still be open to the possibility that *Rat* is true; however, one rejects the existence of any real basis for the priority of reason over moral virtue. The truth or falsity of *Real-D* will thus be at stake throughout this thesis.

Whether or not *Real-D* is true depends on the nature of actions and practical reason. In particular, it requires the nature of human agency or practical thought as such to allow standards of practical reason to be derived, at least in outline, independently of good standards of ethical character. To implement this step, a rationalist must turn to action theory in order to 'find certain features of practical thinking that are essential to intentional agency,' as Michael Bratman says, 'and then to show that these essential modes of practical thinking, without further appeal to general standards of good character, support standards of practical reason that apply to all cases of intentional agency' (2009, p. 511). In the rationalist's view, there are features of our capacity to act intentionally that give rise to independent standards of practical reason. If that is right, there are fundamental standards flowing from this capacity that are independent of individual differences among agents and can be used to single out or recommend the ethical life. Any creature endowed with the faculty of practical reason and the ability to act intentionally will thus have reason to be moral. The answer to the question 'Why should I be moral?' is thus for the rationalist 'Because you are a rational agent endowed with practical reason'. In a word, one will be committed to these standards merely in virtue of having the capacity to act intentionally or for reasons (Setiya 2007b, p. 14). Since, as Kant emphasized, choosing what to do is unavoidable insofar as one is endowed with reason and a capacity for intentional agency, it follows that we cannot avoid having reasons for so doing. Our commitment to morality will be as necessary as the necessity of acting intentionally.

As we shall see in the next section, whether the nature of intentional agency allows deriving standards of practical reason is a crucial part of the rationalist enterprise and it is this point of their strategy that I will critically assess. Before we get there, however, let me briefly say what this lemma involves. One thing that may immediately strike us is that its truth would help explain the pervasiveness of morality. Following Kant, it is common to conceive of moral obligations and requirements as unconditioned (or categorical): they are inescapable as long as they apply to our situation. Such requirements are thus nothing like hypothetical obligations and responsibilities we may choose to take on. One does not sign up as a moral subject as one signs up as a volunteer or enrols as a Green Party member. A related respect in which morality is pervasive is that it leaves no room for a life unbound by morality. Morality applies to anyone insofar as they are moral subjects. As Williams says, 'moral judgement and

blame can apply to people even if, at the limit, they want to live outside that system altogether' (1985, p. 178). This phenomenon of *moral necessitation* indicates that our moral obligations run deep and pertain to what we are essentially. Rationalists can explain that since their crucial lemma, *Real-D*, ties ethical and moral virtue in with our very nature as intentional agents. We are in its grip insofar as we remain agents. In this respect, Rationalism may explain moral necessitation as 'reason's capacity to bind us' (Korsgaard 2009, p. 1).[31]

As such, rationalists pursue a particularly ambitious version of what Velleman defines as a *Kantian Strategy* in that they aim to derive 'normative conclusions in ethics from premises in the philosophy of action' (2006, p. 14). In addition, their strategy conforms to how Korsgaard characterizes the *Kantian Approach* to moral philosophy, namely as trying 'to show that ethics is based on practical reason' and that our ethical judgements are 'explained in terms of rational standards that apply directly to conduct or to deliberation' (1986, p. 5). These connections may make it sound as if Rationalism is a Kantian moral theory; however, I should emphasize that it covers wider ground. Kantians include in their numbers, of course, some of the most prominent rationalists. Velleman qualifies with his effort to explain moral necessitation as deriving from a constitutive aim for action to which everything we do intentionally must be, at least tacitly, directed (1985; 1992b; 1996; 2000a; 2000b, ch. 1; 2006, chs. 1 and 2; 2007, chs. 10 and 11; 2009). Korsgaard, on the other hand, qualifies with her effort to derive moral obligation from the claim that rational agency constitutively depends on the agent's reflective – or, rather, unconditional and categorical – endorsement of an action and on willing one's actions 'as laws for every rational being' (2009, p. xii). A consequence of that view, Korsgaard argues, is that agents acting intentionally are necessarily practically committed to respect humanity as an end in itself (1996; 2009).[32] Likewise, Thomas Nagel (1970) qualifies as a rationalist by his argument that altruism is a basic rational requirement on action and essentially connected with the nature of practical reason. Understood in a broad way, however, Ethical Rationalism not only covers such full-blooded Kantian views concerning the nature of practical reason; it also covers a broad range of neo-Humean (or instrumentalist) accounts of practical reason, Hobbesian views, hedonism, and so on. As long as one harbours the ambition to show that the standards of practical reason, as they conceive them, recommend the moral life as the most rational way to live and these standards are independent of considerations about virtuous traits of character, one is an ethical rationalist.

[31] Kant was therefore right when he talked about conformity to morality and its requirements as a kind of necessitation (*G* 4:413). An agent is not 'by its nature necessarily obedient' to morality; but becoming a moral subject requires necessitation by bowing to a categorical demand.
[32] The stress on necessity does not mean we cannot fail to commit to respect humanity as an end in itself. As Korsgaard argues, that would take away the normative force of the principle (1997). The point is that we cannot fail to so commit when we are agents acting for a reason.

David Gauthier (1986) provides an interesting example: starting with a game-theoretic, end-neutral and neo-Humean conception of practical reason he argues that one may support the rationality of morality on that conception.[33] Where Gauthier differs from full-blooded Kantians is thus not in the general approach to moral philosophy, but in his views about the substantial aims of rationality. Despite these differences, however, the two projects share something important: namely, the rationalist ambition to explain why we should be moral by pointing to independently conceived standards of practical reason.[34] Rationalism is therefore not essentially Kantian, nor is its ability to explain moral necessitation a Kantian feature of the theory; rather, it follows from the structure of the rationalist strategy. As long as we have no choice but to be rational agents and as long as the standards of practical reason recommend and ground moral virtue – whether they be Humean, Hobbesian, hedonist, or Kantian – we are necessarily bound by the rule of morality.

If any such strategy works out, a rationalist account would have a couple of additional attractive features. First of all, it promises a favourable answer to the question 'Why should I be moral?' Moreover, it can provide us with reasons for obeying the demands of morality, which may influence our ways as well as provide grounds for refuting a moral sceptic. In addition, it may offer some explanatory advantages over other philosophical replies to the *Normative Question* – such as certain versions of moral realism – in that its vindication of morality depends all and only on the ontology required to account for intentional agency.[35]

As I have said, this thesis will discuss a crucial aspect of the rationalist response to the *Normative Question*, namely the rationalist's claim that independent standards of practical reason are derivable from the nature of intentional agency. Rationalists are optimists who argue that the correct account of what it is to act intentionally enables us to appreciate that there are independent standards of practical reason and that on these standards it becomes clear that we have most reason to lead a moral life. Philosophy of action would then ground ethics and the moral 'should' be a sort of secondary reflection based on the more primary 'should' of practical reason. Their strategy thus depends on a certain understanding of human

[33] For a similar project, see Foot (1958-9). Also see James Dreier (1997) for some 'Humean doubts'.

[34] Gauthier's ambition is more akin to Thomas Hobbes's project in the *Leviathan*. Hobbes argues that anyone – even a 'Foole' whose egoistic self-interest conflicts with the rules of morality – has a reason to be just or moral. Hobbes and Hume share an end-neutral or instrumentalist conception of practical reason, so they both speculate whether an agent – a 'sensible knave' or a 'Foole' – could be rational in observing the general rules of justice and honesty, at the same time as he takes advantage of exceptions that he can make without damage to his own interests and reputation when doing so benefits him (Hume *EPM* 9.2). As Hobbes says, there might be a 'Foole' who 'hath said in his heart, there is no such thing as justice' and who 'questioneth, whether injustice (...) may not sometimes stand with that reason, which dictateth to every man his own good' (Hobbes *L* ch. 15). But whereas Hobbes sought to reject this as an error of reasoning, Hume accepts that the 'sensible knave' is a genuine case of divergence between moral and practical excellence. For Hume, this represents no threat to morality since he thinks moral virtue reflects our sentiments rather than our reasons. Gauthier, on the other hand, sides with Hobbes on this and tries to vindicate the rationality of morality from an instrumentalist or neo-Humean conception of practical rationality.

My references to Hobbes's *Leviathan* are to the 1985 Penguin Books reprint edited by C.B. Macpherson.

[35] See Korsgaard (1996; 2003; 2009).

agency. In particular, the nature of intentional actions and human agency must be rich enough to allow deriving independently conceived standards of practical reason. In the next four sections I shall discuss what this demand on action theory is and I shall begin to explore whether the nature of human agency warrants their optimism. The nature of this challenge to Rationalism will in turn give structure to the rest of the thesis.

§4 ETHICS without RATIONALISM

In *Reasons without Rationalism* Setiya argues that the ethical rationalist cannot possibly succeed in her project. His argument for drawing this conclusion is complex and I will spend significant parts of this thesis discussing aspects of it in more detail; however, the bottom line is that he thinks Rationalism depends on false assumptions about the nature of human agency. In particular, Setiya believes, the features a rationalist needs in order to derive independently conceived standards of practical reason do not exist; or, as he says:

> *My claim is that this distinction, the idea that there are standards of practical reason apart from or independent of good character, is a philosophical mirage.* (...) [*T*]*he standards of practical reason* cannot *be derived from the nature of agency or practical thought.*[36]

According to Setiya, we have reason to reject *Real-D*, or the claim that there is a real distinction between the good standards of practical thinking and the standards of character (or ethical virtue). Without *Real-D*, however, there is no reason to think there are independent standards of practical reason that could justify the normative authority of moral virtue. If Setiya is right, there will thus be no hope for Rationalism.

Setiya's argument is based on action-theoretic considerations he takes to support the conclusion that 'standards of practical reason cannot derive, not even in outline, from the nature of agency or practical thought' (2007b, pp. 69–70). The rationalist strategy, as you may recall, requires that the practical 'should' and the moral 'should' are distinct enough to leave room for a genuine recommendation or justification of morality by applying these rational standards to actions independently of their moral worth. That requires being able to distinguish and separate 'the property of being good as a disposition of practical thought from the disposition of being good as a trait of character' (Setiya 2007b, p. 69). Without this the standards of practical reason would collapse with the good standards applying to traits of

[36] Setiya (2007b, pp. 1 and 15).

character. What Setiya has noticed, however, is that this crucial lemma makes a demand on the nature of practical reason and intentional actions that may very well fail to be satisfied.

Setiya starts out by arguing that dispositions of practical thought are traits of character (2007b, pp. 70–9), before he applies this conclusion – which I shall accept – to Peter Geach's claim that 'good' is an 'attributive adjective'. Geach defines attributive adjectives as follows:

> *I shall say that in a phrase 'an A B' ('A' being an adjective and 'B' being a noun) 'A' is a* (...) *predicative adjective if the predication 'is an A B' splits up logically into a pair of predications 'is a B' and 'is A'; otherwise I shall say that 'A' is a*[n] (...) *attributive adjective.*[37]

By claiming that 'good' is attributive one denies there could be some x that is good simpliciter. In a word, 'good' makes no sense on its own and we have to be told how things are being evaluated – or what they are evaluated as – when we say they are good. An agent, *A*, is not good simpliciter but must be good as an *F*, or as a *G*, by being good as a teacher, philosopher and so on. It follows that being a good *F* (or good *as* an *F*) is not just a matter of being an *F* and being good. As Setiya points out, being a good theft is 'not a matter of being a theft and being good' (2007b, p. 80). Similarly, being a good parent is not just being a parent and being good. Finally, it follows that 'the *standards* for being a good *F may* differ from the standards for being a good *G*, even when *Fs* are a kind of *G*' (Setiya 2007b, pp. 80–1). Standards for being a good theft are for instance not the standards for being a good action despite the fact that a theft is a kind of action. Standards for being a good parent are also not necessarily the standards for being a good person, although a parent is a kind of person. If we apply this to the standards of moral virtue and practical reason, it becomes clear that it is possible for standards of practical reason – viz. the standards for being a good disposition of practical thought – to be different from the standards for being a good trait of character even though dispositions of practical thought are traits of character. Setiya concludes that since 'good' is attributive there is at least a nominal distinction between the standards of reason and the standards of character traits even if we accept that dispositions of practical thought are character traits. In a word, Setiya concedes the truth of *Nominal Distinction* (*Nom-D*).

The question remains whether this distinction corresponds with something real. The truth of *Nom-D* is not enough for rationalists; they also need the nominal distinction to correspond with a real distinction (*Real-D*). That this claim or its negation – viz. the claim this is a distinction without a difference – is substantive requires an argument. To render the

[37] Geach (1956, p. 64).

nature of what is at stake clearer Setiya asks us to consider what he takes to be a metaphysical truth: namely, that 'standards for being a good *F* are determined by the *nature* of *Fs*, as such' (2007b, p. 82). The *Fs* whose standards we want to fix must have something in common that serves as basis for fixing their criteria of goodness. In Setiya's words, 'the standards of goodness as an *F supervene intelligibly* on the nature of *Fs*' (2007b, p. 82 [original emphasis]). In order for the nominal distinction between good *Fs* (e.g. thefts or good dispositions of practical thought) and good *Gs* (e.g. acts or good traits of character) to correspond to a real distinction (even though *Fs* are *Gs*) there must be something in the nature of the *Fs* explaining this difference. This is what Setiya calls the *Difference Principle*:

> *The Difference Principle*: If *Fs* are a kind of *G*, and being a good *F* is not simply a matter of being an *F* that is a good *G*, there must be something in the distinctive nature of *Fs* to explain or illuminate the difference.[38]

The rationale behind the *Difference Principle* is a version of the widely accepted notion of normative supervenience, or the idea that there are no differences in normative standards for *Fs* and *Gs* without a corresponding difference in some relevant features of *Fs* or *Gs*.[39] To defend *Real-D* an account of practical reason must meet this challenge; conversely, one may reject *Real-D* (and thus Rationalism) if one can show that the challenge cannot be met. Setiya's strategy is thus to argue that rationalists cannot satisfy the *Difference Principle*.

Let me briefly illustrate what meeting the *Difference Principle* would involve. A good theft is arguably good or well-performed in a different way than actions are good or well-performed. Accordingly, we should be able to explain their distinct standards by pointing to a relevant difference in the nature of thefts. Setiya's explanation goes as follows:

> *A theft is,* inter alia*, the dishonest appropriation of another's property with the end of making it one's own. When we evaluate a theft, as such, we ask how well it achieves the aims that belong to the nature of theft: does the thief make off with the loot, in his safe and secure possession? Without claiming that the standards of good theft can be* deduced *from the metaphysics of theft—still less from my rough specification of what theft is—we can see how the nature of theft introduces an evaluative standard that is foreign to the evaluation of action in general.*[40]

[38] Setiya (2007b, p. 83).
[39] For a formulation of normative supervenience in terms of properties, see Wedgwood (2007, 6.5). Also, see James Griffin (1996) for arguments against normative supervenience.
[40] Setiya (2007b, p. 83).

This is just an illustration and Setiya's point does not rely on the specificities of the explanation. What the story exemplifies, however, is what an ethical rationalist must accomplish in order to support the claim that being a good disposition of practical thought is not simply a matter of being a disposition of practical thought that is good as a trait of character. The rationalist must point to something in the nature of practical reason that explains and illuminates the difference and to explain why practical thought has the standards it does by appealing to its distinctive nature or character. As Setiya puts it, the rationalist strategy requires that '[d]ispositions of practical thought must be distinctive, as traits of character, in a way that explains why they are subject to their own evaluative standard' (2007b, p. 85). Actions must be like thefts in having an end of their own, one that determines whether an action is good or bad, as such, and not just good or bad as a trait of character.

In the remainder of his argument Setiya aims to show that this explanatory burden cannot be met since there is nothing about our dispositions of practical thought that could ground the difference in evaluative standards. A significant part of that argument is spent on discussing action theory and the nature of intentional actions. What Setiya needs to do is to rule out the accounts of what it is to act intentionally or for reasons that would enable a rationalist to meet the *Difference Principle*. In this respect, Setiya's argument against Rationalism takes us into the philosophy of action. To make the connection clear we better have a firmer grip on what exactly the action-theoretic issue is and how it ties in with the issue of Rationalism. As I shall argue in the next few sections (§5–§7), the link pertains to a problem in accounting for the nature of intentional actions and acting for reasons.

§5 The PROBLEM of ACTION

Philosophy of action is defined as the field of inquiry dealing with the (metaphysical) nature of action, or what agents do when they are exercising their agency. Philosophers of action thus start from the commonsensical idea that there are agents in the world and that at least some of them are proper agents capable not only of undergoing mere events (or happenings) but contributing to the history of the world by exercising their agency and doing things actively. A cat can for instance participate as it decides to chase a mouse rather than remain seated on someone's lap. In contrast, a radioactive isotope emitting a gamma ray may only react: it is dead, passive and decides nothing. This basic distinction between activity and passivity is not only between kinds of agents; it also occurs within the life of a proper agent as a contrast between the kind of events one actively participates in and those one merely

undergoes. What philosophers of action want is to understand the nature of these differences. A theory of action is thus in a broad sense an effort to understand these active agents and the nature of their activities, and to explain the fundamental distinction between such activity and passivity. This is therefore also known simply as the *Problem of Action*.[41] The key term for making these distinctions is agency. It is the presence of agency that marks the distinction between the class of proper agents and passive (or dead) agents, such as rocks and isotopes, but it is also the exercise of agency that marks the difference between what a proper agent does and what it merely undergoes. It is when an agent exercises its agency that we suppose, as Bratman says, 'that the agent is the source of, determines, directs, governs the action, and is not merely the locus of a series of happenings, of causal pushes and pulls' (2001, p. 91). Only then – that is, *qua* being agents – do we seem able to be what Brian O'Shaughnessy calls an 'ultimate source of change in the environment in the way a river or hurricane is not' (2008b, p. 318).[42] In this section, I will single out an important part of this problem as our primary object of inquiry: namely, the class of rational or intentional actions – or actions done for a reason – which many take to be a hallmark of human existence. I shall assume that we have a working notion of what an action is and sufficient grasp of what it is to do something as opposed to the happenings one passively undergoes and concentrate on our capacity for intentional agency. Our question is thus not: 'What is the nature of action?' We shall ask: 'What is it to do something intentionally?' 'What is it to do something for a reason?'[43]

We are rational animals. What is special about us is that we possess a distinctive kind of agency, the exercise of which enables us to respond to reasons in deciding what to do and for which reasons. Now an agent's agency – its ability to actively participate in events – is

[41] The *Problem of Action* has been defined differently by philosophers. Nagel, for instance, in *The View from Nowhere* treats the problem broadly as the metaphysical problem of the nature of agency and as answering the question 'What is action?' In his mind, this question pertains to all kinds of activities – or anything that is 'not involuntary' – and thus includes the activity of a spider (1986, p. 111). Harry Frankfurt in his paper, 'The Problem of Action', defines the problem similarly as the task of explicating 'the contrast between what an agent does and what merely happens to him' (1978, p. 69). Carl Ginet, on the other hand, defines the problem more narrowly as the task of accounting for the actions and activities that apply to beings 'who have wills' (1990, p. ix). My primary focus will be on actions that are intentional or done for reasons; however, that focus should not make us blind to the fact that there exist activities that fall short of being willed actions but yet are more than mere happenings. We risk ignoring this class if we only confine the problem of action to 'the contrast between what an agent does and what merely happens to him' (Frankfurt 1978, p. 69). As David Velleman points out (2000b, p. 1), this crude version of the founding question in philosophy of action is in line with a certain arithmetical interpretations of Wittgenstein's famous question: 'What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?' (*PI* §621) (References to Wittgenstein's *Philosophical Investigations* are made to G.E.M. Anscombe's 1953 edition). If this understanding of Wittgenstein's question is taken to exhaust the issue, we risk overlooking the activities agents can participate in that are more active than mere happenings, and yet fall short of being willed actions. Answering the broad question 'What is action?', then, involves making more distinctions than between what agents do – which includes intentional actions as well as mere activities – and what merely happen to them.

[42] I will largely bypass the question whether recognizing this capacity is at odds with those who, like Plato in the *Laws*, claim that humans are no more able than other 'mortal animal[s]' to resist our 'pleasures, pains, and desires' and avoid being 'hung up dangling in the air (so to speak) in total dependence on these powerful influences' (*L* 732e). That is, I will bypass the issue of *freedom*.

[43] I work under the assumption that acting for a reason is to perform an intentional action. I thus accept the claim that what is done for a reason is done intentionally. The converse implication is less clear, as Setiya points out (2007b, p. 24), although Donald Davidson thought that the implication carries both ways (Davidson 1963, p. 6). If Davidson is right, it follows that one Φ-es intentionally if and only if one Φ-es for a reason. Some readers interpret Anscombe as being open to the possibility that one may act intentionally for literally speaking no reason (1957, p. 25). In that case, Davidson's thesis will be false. See Maria Alvarez (2009). I will not take a stand on this issue but will take it as a constraint that an account of intentional action must be able to accommodate or explain actions for which the agent had no reason.

presumably grounded in possessing certain causal powers or liabilities. Agents possess various properties that constitute parts of their agency, properties that in turn make it possible to separate the work of the agent from external influences.[44] How one should enact the separation depends on the particular nature of the kind of agent we deal with. Having a certain capacity, *C*, for active agency is a property (or a complex of properties) whose specific nature influences the nature of the activities a *C*-agent can initiate. The contribution of the rational human agent is separated differently from the way we separate a rattlesnake from its circumstances when we explain what it does.[45] When we know the nature of the kind of agent we have in front of us we also know something about the behaviour we can attribute to it and what we should rather attribute to various externalities. In this way we are able to distinguish a 'bird's progress out of the stadium where it has been mistaken for a fast ball' from 'a progress out [of the stadium] in search of better food,' as Thompson puts it (2008, p. 44). In the latter case the behaviour came about because the bird exercised its agency, whereas in the former it came about because someone else exercised their agency on it. What we want when we single out intentional actions is to understand the kind of agency that agents like us possess, the exercise of which enables us to act intentionally or for reasons. That may or may not include birds; however, it narrows down our primary interest to beings endowed with reason and will and what they are capable of when they rise to the occasion.[46]

There are many reasons for carving out this group of actions as the object of inquiry. On the action-theoretic side we find a tradition for dealing with the nature of action where doing something intentionally has been recognized as a key notion. This is at least partly due to the perceived explanatory importance of this group of actions and its significance in someone's life as an agent. As Hornsby says, 'one keeps track of what is significant in someone's life as an agent if one attends to what they intentionally do' (2004a, p. 4).[47] My reason, however, is its possible connection with practical reason and good practical thinking. Intentional actions are a hallmark of agency due to the latter's connection with reasons and reason-explanation. Anscombe famously sought to define the category of intentional actions as those actions to which a certain sense of the question 'Why?' is given application': the sense being 'that in which the answer, if positive, gives a reason for acting' (1957, p. 9).

[44] Thompson (2008) argues that live agents participate in a 'life-forms' and that this is not merely objects with special features and properties, but instances of a *sui generis* logical category on top of the conventional categories recognized by Gottlob Frege (e.g. concepts, relations and objects). See Frege (1892a; 1892b). If Thompson is right, much of what follows should be reformulated, but the substantial points remain.
[45] This example and the notion of separateness is discussed by Nomy Arpaly (2006, pp. 3–5).
[46] In what follows, I will regard practical thinking and the will as our capacity for self-determination the exercise of which enables an agent to take something as her reason for acting and act intentionally or for reasons. The relationship between this faculty and (judgements of) reason – the faculty whereby one makes judgements and evaluations – is, at least in part, what will be at issue in what follows.
[47] As Hornsby remarks, this is also one of the principal claims Davidson made in his (1971) when he asks for the events that 'reveal agency'.

Intentional actions are instances of a behaviour we may be required to explain and justify by giving reasons for what we are doing. An account of intentional agency should, among other things, explain why this is so. A natural explanation is that agents acting intentionally had reasons for so doing, at least in the paradigmatic case. As we shall see, this makes intentional actions particularly interesting for those who approach action theory from ethics and with a concern for the *Normative Question*. In particular, it may offer hope to the ethical rationalist and their effort to meet the *Difference Principle*. Their hope is kindled by a certain problem that pertains to the so-called *Standard Story* of intentional actions. It will, however, take some effort to develop what exactly their hope consists in. Let me therefore first discuss the nature of action-explanations and the Standard Story, before I go on to discuss what its problems are in the next section. Then, finally, I shall close that section discussing why suggested solutions to these problems may provide hope for the rationalist.

One thing we want in action theory is an explanation of intentional actions. An action-explanation may, however, have different goals. It may for instance, as Hornsby says, aim to tell us something about the agent that 'makes it understandable that she should have done what she did' (2004a, p. 8). We are then typically interested in knowing why someone did something. Having such an explanation enables us to know for what reason they acted as they did and why the agent chose to play some particular role. This is exemplified by:

(1) She did such and such because she desired ...
(2) He did so-and-so because he believed that *P*.
(3) I am walking down Broadway because I'm heading for Union Square.
(4) She is reading *Being and Time* because it is assumed to be Heidegger's masterpiece.

The focus in (1) to (4) is on the individual agent and why she did something in the sense of 'Why?' that operates as a request for the agent's reasons. They explain such actions by uncovering an agent's motivation in a way that 'advert[s] to their desires and to beliefs they have about ways to bring about things in line with their desires', or else by directly citing their reasons for acting (Hornsby 2005, p. 112). Their adequacy signals that the actions are intentional since some of the many descriptions that apply to these agents' behaviour reveal it as being done for a reason. Here the explanations explain by revealing facts about their situation: the fact that *Being and Time* is assumed to be Heidegger's masterpiece, the fact that its agent desired something or believed something another, or that I am heading for Union

Square.[48] But unlike other facts that explain why something occurs or happens – such as the reason why a bridge collapsed, or why tiger cubs grow teeth – they explain by revealing reasons *for which* the agent acted. In a sense, the agent takes a part in her own explanation by contributing the relevant reasons that explain why she is doing something.

A reason for acting is ambiguous between its justificatory and explanatory (or motivational) sense. This is no coincidence, for, as Michael Woods points out, 'the concept of a reason for an action stands at the point of intersection, so to speak, between the theory of the explanation and the theory of their justification' (1972, p. 189). In one sense, a reason, *R*, is a good or normative reason that counts in favour of doing something. *R* may, for instance, be a reason for me to finish my thesis and thus specify a content that justifies this act. They make 'valid claims on those who have them,' as Korsgaard says (2008a, p. 208), and so 'should be,' as Stephen Everson claims, 'such as to favour the action for which it is a reason' (Everson 2009, p. 27). In short, these are items that are conceptually related to conditions of good practical thinking as they support (or tend to justify) an attitude or course of action for a particular subject given her situation.[49] They are similar to how facts count in favour of believing something. Some fact f_1 counts in favour of believing *p*; similarly f_2 may count in favour of me being afraid, careful, considerate, or picking up a large stick. Reasons for acting are thus considerations favouring that some agent *A* Φ-es in such-and-such a situation.[50]

Good reasons, although they should be able to figure in a reason-giving explanation, may or may not be an agent's reasons for acting. When people act for reasons their reasons for acting may fail to be reasons for so doing. In this context 'reason' is thought of as one of the constituent objects of the possibly complex beast known as acting for reasons. An agent's reason need not favour the action for which it is one's reason for acting since in acting for reasons human imperfection enters with dramatic effect: we may act for a reason that is not a (good) reason to act.[51] We are capable of being quite unaware of good reasons for Φing; or we may be aware of them and yet fail to will the actions they recommend; and we Φ for

[48] As John Hyman points out, the sense of 'because' 'which introduces a reason for doing or not doing' may imply that its agent Φed intentionally while it identifies her reason for Φing (1999, p. 441).

[49] See Setiya for a similar characterization (2007b, section 1 (especially p. 9)).

[50] By considerations I do not mean the act of considering, but rather the objects *considered* (the content of an act of considering). There is a similar ambiguity in talk about beliefs: a belief might refer to the act of believing something or the object believed.

[51] The awkwardness of this way of talking is, as Dancy points out, that we may have cases where it is okay to say that the agent has a reason for acting (a motivating one) in the absence of any reason for doing so (a normative one) (2000, p. 3). In other words, there might be a case where 'there was no reason to do what he did, even though he did it for a reason.' These are not contradicting claims; rather, it is no more than 'a little local difficulty' due to the pleonastic sense of 'reason'. If we spell it out, we would say that the agent had a (motivating) reason without having a good or normative reason. Dancy advises us to take the different senses of reason in play here as two ways to answer Anscombe's question 'Why?' that have a somewhat different emphasis. This might still leave us with a problem of ontology, however. For suppose that *A* acts for the reason *p* when she Φ-es in a situation where *p* is in fact a very bad reason (or no reason at all) for Φ-ing. Perhaps *A* jumps into a river in order to save what she takes to be a drowning child but which in fact is a piece of wood with deceptive looking clothes around it. What kind of item serves as the reason for which the action was done in such a case? In other words, what kind of thing is *p* when *p* is someone's reason for Φ-ing and yet fails to denote a real or good reason for Φing? It would be premature, however, to deal with this before we have a grip on what it is to act for a reason. I will put this issue aside and think of an agent's reasons as *putative* facts.

considerations that are no reason for Φ-ing.[52] Iago, in Shakespeare's *Othello*, manipulates people with verbal artistry into a web of malice and treason because Othello failed to promote him. This is a fact but no good reason for Iago's evil schemes. Iago might even perceive reasons for behaving differently and yet resist the force of these better reasons. At other times our reasons for acting are not facts, as when Othello murdered his beloved wife, Desdemona, in a fit of jealousy because he (erroneously) believed she was an 'impudent strumpet'.[53] That *A* is Φ-ing for reason *R* does not entail that *R* is a fact or a good reason for Φ-ing:

(5) His reason for Φ-ing was that *R*, but in fact *R* is not true nor would it justify Φ-ing if it were.[54]

With (5) it is easier to see that agents may act for what fails to be (good) reasons. Now this may seem to be in tension with what some philosophers have pointed out, namely that the 'because' that operates in action-explanations seems to give rise to a factive reading.[55] If so, what it identifies as the because of a reason-explanation must be a fact. Whether that is correct or not – and it is not entirely clear to me that it is correct – it should not lead us to conclude that the agent's reasons as they are identified by (1) and (2) are facts. The facts required by the 'because' here are the facts that the agent had this-and-that desire and belief. In (3) and (4), on the other hand, the fact required by the 'because' – if it is – is the agent's reason. It follows from (4) that *Being and Time* is assumed to be Heidegger's masterpiece; whereas it fails to follow from (2) that *P* is a fact. To avoid these complications it is better to think of explanations that identify an agent's reason as saying '*A* Φ-ed for the reason that *R'* (or formulations analogous to (5)). I will then follow Dancy (2000) in taking the schematic letter '*R*' as a consideration or putative fact. The agent's reason is thus a candidate good reason. If some '*R*' is identified as someone's reason for Φ-ing, it could therefore refer to a good reason for Φ-ing if only *R* really is a fact and favours or justifies that *A* Φ-es.

The above action-explanations should not be confused with what is known in the literature as the Standard Story. Inspired by Donald Davidson, the Standard Story seeks not

[52] These clauses may be in tension with what is known as *Motivational Internalism* (or the *Motivation Requirement*), which is the claim that if some agent *A* has a good reason *R* to Φ and *A* is properly aware that *R* obtains (i.e. aware that she has a good reason to Φ), then *A* must desire to Φ (or be motivated to Φ) (Setiya 2010b p. 82). This is an issue over which *externalists* and *internalists* about reasons disagree. Internalists, such as Williams, insist that a certain motivational propensity to Φ must be there if it is true that an agent *A* has a reason *R* to Φ (1980; 1989; 1995c). Externalists, on the other hand, disagree that there is motivational requirement on good (or normative) reasons for acting and would argue that the agent's lack of motivational propensity is a mark of irrationality (since the agent then fails to respond to the reason she has for Φing) (Korsgaard 1986). In between we find McDowell arguing that lack of motivational propensity to Φ for reason *R* may happen without thereby inflating the normative authority of *R*; yet, he agrees with the internalist that the agent's failure is not a mark of irrationality (1995b). I take it that an agent may see *R* as a reason to Φ without being motivated to Φ and I think *R*'s status as a reason for Φ-ing will survive the agent's lack of motivational propensities. What is wrong with the agent in such a case is her *will*, or her capacity for self-determination. I will not consider whether this is truly a mark of irrationality, or just the result of bad upbringing. See R. Jay Wallace (1999).

[53] Othello believes this as a result of Iago's evil scheme, of course. See Shakespeare (1975, *Othello* 1.1.7–14 and 4.2.86).

[54] Similar examples are discussed by Dancy (2000, p. 132).

[55] See Setiya (2007b, p. 29) and Thompson (2008, p. 88). Hyman (1999, pp. 441–2) also thinks the 'because' entails its agent 'knew that *p*'.

merely to explain someone's intentional action by citing their reasons for doing it; rather, it seeks to explain the nature of actions in terms of the causal aetiology of the event that is one's action (viz. by citing the mental causes of an agent's bodily movements). On this story, one explains why some event is an action by showing that an agent's beliefs and desires played a specific role in that event's (prior) causal history.[56] What this story suggests is thus that we analyze actions in terms of a causal-motivational explanation that situates the event that is one's action within, as Bratman says, 'some, perhaps, complex, causal structure involving, events, states and processes of a sort we might appeal to within a broadly naturalistic psychology' (2001, p. 92). That is, the story explains the nature of intentional actions as

> *... movements of an agent's body that are suitably caused by a desire that things be a certain way and a belief on the agent's behalf that something she can just do, namely perform a movement of her body of the kind to be explained, has some suitable chance of making things that way. Movements of the body that are caused in some other way aren't actions, but are rather things that merely happen to agents.*[57]

To explain individual actions on this model would involve explanations such as:

SS_1 Her desire ... caused [an event which was] her bodily movement.[58]

SS_2 Her belief that *Being and Time* is assumed to be Heidegger's masterpiece caused her to read it.

SS_1 and SS_2 identify what Davidson coined, somewhat technically, an agent's *primary reason,* which is the mental state(s) that caused the behaviour in a certain qualified way. In addition, however, they also work to identify an agent's reasons for acting, in the sense just discussed.[59] These notions should be kept apart since a person's psychology is usually not taken into account in the agent's practical thinking. Of course, there are cases where the fact that one believes something is a consideration in practical thinking: as Dancy says, someone believing 'that there are pink rats living in his shoes may take *that he believes this* as a reason to go to the doctor or perhaps a psychoanalyst' (2000, p. 125). But this is exceptional and does not

[56] The aetiology here is not meant to include Aristotle's four *aitia* but is restricted to efficient causation. It should be noted that some variants of the Standard Story suggest that what distinguishes actions from mere events is not that the mental states that bring about the event that is one's actions are antecedents in the event's prior causal history. Rather, they suggest that the crucial component is the agent's state while the event unfolds and that it must occur under the person's guidance. See Frankfurt (1976; 1978). Velleman (1992b) develops a similar view.
[57] Smith (2009, p. 58).
[58] Hornsby (2004a, p. 8).
[59] A primary reason why some agent Φed is defined as the pair of attitudes – i.e. a *propositional attitude* (e.g. believing that *p*, knowing that *p*, seeing that *p*, etcetera) and a *pro attitude* (e.g. what the agent wanted, desired, prized, held dear, valued, found agreeable, and so on) – that together *rationalized* the action and brought it about by causing an agent's bodily movement in the right way. In order for a state to rationalize one's behaviour, it is important that it leads us to see 'something the agent saw, or thought he saw, in his action' (1963, p. 3). As Davidson also noted, one half of the pair is often implicit in the explanation and one implies the presence of the other (1963, p. 4). The desire mentioned in (SS_1) implies a relevant desire and *vice versa* for (SS_2).

change the fact that when we talk about an agent's reasons we are thinking of the content of what she considers in practical thought as opposed to the psychological states themselves.[60] As Setiya says, '[e]ven if acting for reasons is to be understood in terms of the causal role of psychological states, it would be a mistake to describe these states as "the agent's reasons"' (2007b, p. 30). That is a valid point, even if one is sympathetic to the Standard Story.[61]

By providing explanations such as SS_1 and SS_2 the Standard Story aims to explain one's action by pointing to its causal story. Its primary focus is then on why an event occurred rather than on why the agent did something.[62] SS_1 and SS_2 thus aim for something different than (1) to (5). The difference is not that (1) to (5) are not characterizing causal relationships. Rather, SS_1 and SS_2 reflect a particular concern for what I will refer to as *Davidson's Challenge*, namely that action theory must address the following question:

> *Davidson's Challenge*: When an agent, *A*, Φ-es for a reason *R*, what makes it the case that *A* Φ-es for reason *R*?

Davidson's Challenge is meant to give voice to a line of criticism that Davidson raised against Anscombe's *Intention* and action theory pursued in the tradition from Wittgenstein: namely, that this tradition fails to recognize that '[c]entral to the relation between a reason and an action it explains is the idea that the agent performed the action because he had the reason' and that this 'because' requires explanation (1963, p. 9). One may accept that (1) to (5) explain the individuals actions; however, Davidson also asks us to say what the general story is when some reason explains one's Φ-ing rather than others. Davidson thus assumes, at least as a working hypothesis, that there is 'an interesting principle at work' that we may pin down and use to explain the relationship 'between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did' (1963, p. 3; 1971, p. 43). In an important sense, this move goes beyond the challenge itself. In fact, there is nothing in the challenge suggesting that we can only meet it by providing an analysis of the 'because' in (1)

[60] Some philosophers also refer to the agent's primary reason as her 'motivating reason'. See Smith (1994, chapter 4).

[61] Davidson may have thought that his account of primary reasons reduced or analyzed the notion of an agent's reason; however, for my purposes it is wise to keep these notions nominally distinct. What an agent's reason is in the case of (SS_1) is of some dispute since it makes reference to a desire rather than a belief. Beliefs have propositions and putative facts as objects, and this suggests that an agent's reasons should be rendered accordingly. Desires may have other objects than what can be described as a that-clause. Perhaps, the objects of pro attitudes are *actional objects* rather than propositional objects, and so better rendered as 'to nibble some olives', or 'to drink a glass of sherry'. One may further speculate whether reasons for acting are better thought of as actional objects as opposed to putative facts or propositions. See E.J. Lowe (2008, pp. 205–12). I will not enter the debate at this stage, but talk as if reasons for acting are putative facts. I am open to revision if it turns out that we also need to invoke actional objects.

[62] This focus is not a peculiarity of the versions of the Standard Story that focus exclusively on the events prior causal *aetiology*. When Harry Frankfurt formulates his influential criticism of this view, he phrases the crucial point as follows: 'the state of affairs *while* the movements are occurring is far more pertinent' (1978, p. 72). Frankfurt thus retains what Hornsby calls the Standard Story's event-based character, or the view that actions are explained when we have a story that delimits which events are actions, a story that draws on the causation and/or guidance of spatiotemporal particulars that occur. For criticism of its event-based character, see Hornsby (2004a; 2004b; *forthcoming*). For defence of the Standard Story, see Smith (*forthcoming* a; *forthcoming* b).

to (5); nor does its question assume that there is a single relationship holding between reason, agent and action when some agent acts intentionally or for a reason. For many of the followers of Davidson, however, that working assumption still holds with the result that for them the *Primary Question* for action theory is something like the following:

> *Primary Question*: What is the relation between a reason, an action and its agent when the reason explains the action by giving the agent's reason for doing what he did?[63]

In what follows, I will accept that *Davidson's Challenge* asks a legitimate question. I will also discuss theories that take the *Primary Question* as their starting point and try to come up with an analysis of the relationship they take to obtain between reason, action and agent when the agent acts for a reason. Bear in mind, however, that I do not think answering the *Primary Question* is the only way to meet *Davidson's Challenge*. A possible answer is to say that what makes it the case that some agent is acting for a reason *R* is that the agent is taking *R* as her reason for so acting and thus vouch for a model where there is a plurality of ways one may take something as one's reason that does not reflect one property of agents.[64]

Meanwhile, let us focus on the way Davidson points out for us: followers of the Standard Story start out from the *Primary Question*. They then argue that one can answer this question – thereby meeting *Davidson's Challenge* – only by treating and analysing reason-giving explanations as a species of causal explanation.[65] The Standard Story is thus a candidate analysis. However, that story has run into serious problems. In particular, it struggles with what we may call *missing agent problems*. Briefly put, it seems as if 'the story leaves agents out' (Hornsby 2004a, p. 2). As we shall see, it is by addressing this issue that

[63] On Davidson's account 'reason' in the first two instances denotes an agent's primary reason and citing this primary reason is what gives the agent's reason (viz. a putative fact) for doing what he did. As I phrase this question, however, it is meant to remain neutral on whether the reason that gives an agent's reason is a primary reason, or whether some other story needs to be told to give an agent's reason for acting.

[64] See Hornsby (2004a; 2004b, *afterword*).

[65] As Kirk Ludwig points out (2006, p. 147), Davidson called into question established orthodoxy in the philosophy of action, which, under the influence of Gilbert Ryle's *The Concept of Mind* (1949), Wittgenstein's *Philosophical Investigations* (1953) and Anscombe's *Intention* (1957), resisted the thought that reasons for actions could be their causes. Davidson's claim may be ambiguous, however, between an agent's reasons and Davidson's technical notion of a *primary reason*. Primary reasons are, as I said above, pairs of mental states – a pro attitude and a propositional attitude – which, according to Davidson or the Standard Story, must rationalize and cause the event that is one's action. If a reason-giving explanation is causal, it thus cites primary reasons (i.e. mental states). However, there is a sense of giving an agent's reasons for Φing that does not provide the agent's mental states but rather directly provides the content that moved the agent to act (e.g. (3) above). Davidson does not make the claim that those reasons must cause the event that is one's action. A reason-giving explanation that cites these reasons may therefore be non-causal. *Davidson's Challenge*, however, urges action theory to explain why the latter kind of reason-giving explanations can explain action. In a word, Davidson seems to think that the latter kind of reason-explanation fails to fully explain the action without a supplementary causal story that cites the agent's primary reasons and that 'only the causal account can answer the question what makes one of two competing primary reasons for an action the one that explains it when the agent acts on only one of them' (Ludwig 2006, p. 149). If Davidson succeeds, he would in a sense complete what Anscombe embarked on in *Intention*.

Another potential ambiguity is the notion of 'cause'. Properly understood, disagreement between Davidson and the Rylean-Wittgensteinian-Anscombean tradition is not so much whether reason-giving explanations are in a sense causal; but whether reason-giving explanations of the sort that directly cites an agent's reasons for Φ-ing (e.g. (3) above) are analyzable in terms of a story that cites the event's efficient causes (e.g. its primary reason). See G.F. Schueler (2003, p. 18).

one may need to invoke constitutive elements that are friendly to the project of Rationalism. Let us first see what the problem is (§6) before addressing its relevance for Rationalism (§7).

§6 The SEARCH for an AGENT

Agents capable of intentional actions provide arenas for many different kinds of events and not all of them are the agent's own doing. The complicated movements of my body as the result of a seizure or of being hit by a falling piano on Fifth Avenue only happen to me and I am a passive bystander (or victim) to it. Other events – say, playing Beethoven's 32nd piano sonata – are truly one's own doing. The question is: what is the actional character in the latter kind of episode that distinguishes it from the former? The Standard Story locates the actional character in the nature of some assumed acting for relation – or a relation that grounds the fact that some reason, *R*, explains one's Φ-ing – and it starts out from the idea that a sort of responsiveness on part of the agent is required. To explain that we need more material than (good) reasons and agents: we also need to say something about the agent's relation to the reasons that explain her action. We need to account for this acting for or because of relation.

The Standard (causal) Story of (human) action suggests that an answer to *Davidson's Challenge* is a metaphysical analysis that accounts for *A's* Φ-ing for some reason *R* by displaying the causal-motivational role of *A's* psychological states (e.g. beliefs and desires).[66] Some models emphasize the prior causal history of the event that is one's action; others invoke the notion of guidance and higher-order attitudes. There are also differences between what kinds of attitudes that different versions of the Standard Story appeal to – e.g. some allow the state of knowledge whereas others think that what knowledge adds to belief is not psychologically relevant.[67] The shared idea, however, is to cite an agent's attitudes – e.g. her desire (or, better, a *pro attitude*, such as wanting, holding dear, or thinking dutiful, beneficial, obligatory, or agreeable) for an end and her belief (or a *propositional attitude*, such as knowing, perceiving, noticing, remembering) that the action is a means for realizing the end – and put the attitudes into a context of a *rationalizing*[68] causal-motivational explanation.

[66] This view is often referred to as 'the standard causal story of human action', 'the standard causal story', or simply 'the standard story' and 'the standard model'. The latter two are preferable since their names do not preclude whether we talk about actions that pertain all and only to humans, or accounts that explains actions in terms of their prior causal history as opposed to guidance. It is also known in the literature as the Belief-Desire Model, or the Humean Theory of action (see Smith 1987). One of its main proponents is, of course, Davidson and the Standard Model is, at least partly, a result of his seminal contributions to action theory (Davidson 1963; 1967; 1971; 1973; 1978).

[67] For support of the latter claim, see Stephen Stich (1978, p. 574). For defence of the role of knowledge in action-explanations, see Timothy Williamson (2000, chs. 2 and 3), John Gibbons (2001) and Hornsby (2008).

[68] A rationalizing explanation in philosophy is a kind of explanation that bears some relationship to rationality and rational form. This is different from the connotation rationalizations have in contemporary psychology, where it refers to a sort of defence mechanism. Here a subject rationalizes her behaviour to come up with logical or socially desirable motives in order to appear more rational (Atkinson et. al. 1996, p. 501). An experiment involving posthypnotic suggestion may illustrate: a hypnotist instructed a subject under hypnosis to open the window if the hypnotist took off her glasses. The subject was also told to not remember the instruction. After waking up, the subject did open

In its crudest form, the model explains acting for a reason as what is caused by a relevant belief-desire pair when they rationalize the action so that 'from the agent's point of view there was, when he acted, something to be said for the action' (Davidson 1963, p. 9). The problem with that suggestion is, as Davidson later noted, that 'not any causal relation will do' (1978, p. 87). We also need to disarm cases of wayward causation and deviant causal chains (or, as Velleman puts it, cases of uncharacteristic causation (2000b, p. 7)). Variants of Davidson's well-known *Climber Case* show that it is possible for belief-desire pairs to cause the kind of behaviour that is rationalized by the content of one's attitudes in a way that is coincidental. You may, for instance, want to impress your audience and know that the only way to do that is by crying. These thoughts may be so frustrating that you actually cry thereby impressing the audience. Did you cry for a reason? Well, no. But then it is also possible for attitudes to cause behaviour it rationalizes without implying that its agent acted for a reason.[69]

Some philosophers take this to indicate that we must try to further qualify the causal-motivational explanation. Davidson, on the other hand, is sceptical about the prospect of ruling out causal deviance to specify 'the right way' attitudes cause the events that are actions; rather, he suggests we settle for a primitive notion of causation at this point.[70] However, I think a model is able to pull this off if it requires that the causal relation be a characteristic outcome of one's attitudes.[71] To act for a reason the behaviour must stand in a certain relationship to the content of one's belief by being conducive to the outcome desired. When you cried as the result of having those thoughts it was mere accident that made you act this way. Your motive, although it exercised some of its causal powers, did not exercise its characteristic powers in causing the behaviour. The motivation therefore caused the behaviour making it an action of yours – in contrast with behaviour resulting from spasms or the intake of a drug – but it is not acting for a reason since the attitudes failed to motivationally guide you. When someone intentionally Φ-es she must therefore not only want to Φ and know a way to realize her wants, but also allow this belief-desire pair to guide the behaviour towards

---

the window; however, he clearly experienced some initial hesitation to act on this impulse and seemed to carry the instruction through only by seeking a reason for his behaviour, as was apparent when he said things like, 'Isn't it a little stuffy in here?' (Hilgard, 1965)

[69] See Davidson (1973, pp. 79–80; 1976a, pp. 264–5). My version of it is discussed by Velleman (2000b, p. 7).

[70] Davidson concludes his *Climber Case* by saying that 'the causal chain must follow the right sort of route' (1973, p. 78); later he expresses pessimism about characterizing what that it is 'in non-circular terms' (1978, p. 87). We are thus left with a primitive notion of the right kind of causation on Davidson's programme. This may be a sign of sound theoretical modesty: not everything can be analysed further and causation is a good candidate for a primitive notion. Arpaly suggest this as a reply to what she calls 'the problem of content efficacy' (2006, pp. 46–7). However, given what I argue bellow – namely, that we need to account for the notion taking as one's reason – we see that full theoretical modesty at this stage of our inquiry would simply make us blind to an illuminating condition for acting intentionally, one that we can get a hand at without suggesting non-circular analysis of causation to fully solve the problem of deviant causal chains.

[71] See Velleman (2000b, p. 7). The characteristic causal powers of the attitude is determined by the kind of attitude it is (e.g. belief, desire, knowledge, valuing, etcetera) and what its object is (i.e. the content of the attitude).

its object (Setiya 2007b, p. 32).[72] What we need is purposively guided behaviour where the relevant psychological states cause and guide the behaviour in a characteristic manner. Hence, we get Velleman's characterization of the *Standard Model*:

> *We want something to happen, and we believe that some behavior of ours would constitute or produce or at least promote its happening. These two attitudes jointly cause the relevant behavior, and in doing so they manifest the causal powers that are partly constitutive of their being, respectively, a desire and a belief. Because these attitudes also justify the behavior that they cause, that behavior eventuates not only* from causes *but* for reasons.[73]

Suppose I want to know Velleman's account of rational agency; I believe that reading his *Practical Reflection* is a way to know his account. These attitudes cause me to pick up *Practical Reflection* and do some reading thereby manifesting their characteristic causal powers as a desire for ... and a belief that *p*. Since they also justify picking up *Practical Reflection* – at least, in the 'somewhat anaemic' sense required for rationalization (Davidson 1963, p. 9) – I count as having read *Practical Reflection* intentionally and for a reason.

As I will take it, one operates with a Standard Model even if one adds or suggests different states of an agent to operate in the role of Davidson's primary reasons. Perhaps one believes, as Davidson came to believe, that a *sui generis* pro attitude intention (viz. an all-out judgement or verdict that this action is best, preferable or favourable) must materialize from *prima facie* judgements of desirability and preferability provided by the agent's ordinary desires and wants (viz. that Φ-ing is preferable in so far as...) and that one acts for a reason only if such an intention brings the behaviour about in a rationalizing non-deviant manner.[74] Perhaps one operates with second-order volitions and conditions of guidance, as Frankfurt proposes (1971; 1976; 1978; 1982; 1987). All these fall under the scope of the Standard Model insofar as they explain and qualify why certain events are actions by pointing to the fact that the occurrence of the event is either occasioned by, or guided by psychological states of the agent in an account that emphasizes the causal-motivational role of these mental states.

[72] Davidson (1973, pp. 78–9) distinguished deviance in the causation of basic actions from deviance in non-basic actions, but this complication does not make any significant difference to our discussion.
[73] Velleman (2000b, p. 5).
[74] See Davidson (1978) and Ludwig (2006, p. 156). As McDowell points out, the content of Davidson's all-out judgements is unconditioned ('Φ-ing is better than its alternatives, *period*') as opposed to the *prima facie* judgements of preferability expressed by desires and wants (e.g. 'Φ-ing is better than alternatives in so far as…') (McDowell 2010, p. 418). Notice that on Davidson's account, an *all things considered* (*ATC*) judgement is not an agent's all-out (*AO*) judgement (Davidson 1970a, pp. 33–40; 1978, pp. 98–102; 1976b, p. 285). An extra step is needed to make the choice of one's *ATC* one's *AO*. It is thus possible, though not without implying irrationality, to *ATC* in favour of Φ-ing over ψ-ing and yet *AO* and thereby choose to ψ. In which case, its agent acts against her better judgement. *AO*-judgements are thus crucial components of Davidson's account of *akrasia* (weakness of will).

The problems are not solved, however, even if we accept Velleman's solution to causal deviance, add *sui generis* decisions, intentions and so on. In fact, there is a sense in which, as Velleman says, 'the discussion of "deviant" causal chains has diverted attention from simpler counterexamples' since an agent can 'fail to participate in his behavior even when it results from his motives in the normal way' (1992b, p. 126). Consider a case from Sigmund Freud's *Psychopathology of Everyday Life*:

> FREUD'S INKSTAND: *My inkstand is made of a flat piece of marble which is hollowed out for the reception of the glass inkwell; the inkwell has a marble cover with a knob of the same stone. A circle of bronze statuettes with small terracotta figures is set behind this inkstand. I seated myself at the desk to write; I made a remarkably awkward outward movement with the hand holding the pen-holder, and so swept the cover of the inkstand, which already lay on the desk, to the floor. It is not difficult to find the explanation. Some hours before, my sister had been in the room to look at some of my new acquisitions. She found them very pretty, and then remarked: 'Now the desk really looks very well, only the inkstand doesn't match. You must get a prettier one.' I accompanied my sister out and did not return for several hours. But then, as it seems, I performed the execution of the condemned inkstand. Did I perhaps conclude from my sister's words that she intended to present me with a prettier inkstand on the next festive occasion, and did I shatter the unsightly old one in order to force her to carry out her signified intention? If that be so, then my swinging motion was only apparently awkward; in reality, it was most skilful and designed, as it seemingly understood how to avoid all the valuable objects located near it.*[75]

Freud discusses this case under the heading 'erroneously carried-out actions': that description is suitable. If the answers to Freud's questions are 'yes' – which is plausible – what he describes is an event where rationalizing beliefs and desires jointly cause his breaking of the inkstand and guide it purposively. His motives steer his hand so skilfully that only the inkstand breaks leaving the other objects on his desk intact. Freud is therefore rightly puzzled by such remarkable clumsiness in combination with precision of execution. The outcome seems to be no accident, as it were. On consideration it can look as if breaking the inkstand was as aim-directed as if it had been intentionally carried out. Yet, this behaviour came about in such a way that it just cannot be a case of intentional action (at least not under that description). His motivational states were too convoluted for that. Too much is going on behind the agent's back. In a word, we seem to have a case of merely motivated action.[76]

[75] Freud (1904, p. 85).
[76] That the *Standard Model* fails on this case is defended by both Velleman (2000b, p. 2) and Setiya (2007b, p. 33). Setiya (2003, p. 351) mentions a case from D.E. Milligan (1974, p. 191-92) in the same breath as he mentions the case of Freud's inkstand (Setiya does not

The problem for the Standard Model is that it qualifies such cases as acting for a reason when it should not. On the Standard Model all it takes for an action to be done for reasons is if it is characteristically caused by relevant and rationalizing psychological states. Freud's breaking the inkstand qualifies on all fronts. Of course, this event clearly is something the agent does rather than merely undergoes and the subject is in one sense making this happen; however, there is something missing here. What is missing is the presence of intentional agency: Freud fails to participate in the unfolding of this event as its agent. He does not regulate his behaviour in the manner that distinguishes intentional actions from merely motivated actions. The Standard Model is thus at fault for not making this distinction.

A suggested fix for the Standard Model is to invoke decisions, such as Frankfurt's higher-order volitions, to determine 'what the person really wants by making the desire on which he decided fully his own' where 'the function of the decision is to integrate the person both dynamically and statically' (1987, pp. 166–7, 170). The problem is that one might have unwitting decisions, or decisions with which the agent does not identify. Velleman, for instance, shows a case where he has a long-anticipated meeting with an old friend in order to resolve a minor difference. As they talk, however, the friend's offhand comments provoke him with the result that they part in anger. Later, Velleman understands that 'accumulated grievances had crystallized' so much that he had unwittingly formed a resolution to 'sever our friendship'. In this respect, an unwitting decision made him do it; in a word, it was, as Velleman says, 'my resentment speaking, not I' (1992b, pp. 126–7). Decisions just like rationalizing belief-desire pairs seem susceptible to internality and to externality. In the same fashion Setiya argues that motivation by belief may '*not* count as acting for a reason, even though they do involve the intention to act' (2007b, pp. 34–5). A law student may, for instance, be convinced, due to pressure from her parents, that all she wants is to become a lawyer and that she is perfectly cut-out for such a career. As a result of her conviction, she may fail to take notice of all the evidence indicating otherwise, such as the fact that she spends much less time working than her peers, that she often feels tired and lethargic, or that she never gets good grades. Suppose she quits and is 'moved to do so, unconsciously, by beliefs that correspond to these facts' (Setiya 2007b, p. 35). She may then find her decision both capricious and hard to explain; nevertheless, a decision was made and moved her to quit.

mention this case when he discusses Freud's case with the same conclusion for the standard view later in his (2007b, p. 33)). In the Milligan case an alcoholic succumbs to his compulsive desire for a drink and is argued to *not* act intentionally or for reasons even though he satisfies the standard view's conditions for being an intentional action. I think there is a marked difference between these cases. The compulsive agent may experience severe stress and anxiety when he is under the pressure of his compulsive desires – something we may describe as irresistible – but he does decide to cave in to the desire. The alcoholic is coerced and much like the family father who decides to remain a loyal prison guard in the concentration camp when they threaten his family; but that does not make his action less intentional. The fact that the compulsive agent decides to act on the compulsion tells against disbanding them from the realm of intentional agency.

This story is not about someone quitting law school in order to find a more appropriate career. 'We must not be misled,' as Setiya says, 'by the use of "reason" to mean no more than "cause"' (2007b, p. 35). The 'because' in this case is the 'because' of psychological explanation and not of reason-giving explanation. It is psychologically on a par with Freud breaking his inkstand and Velleman deciding to break off his friendship in involving motivation that fails to be fully the agent's own. They resemble how Isabel Archer, the heroine of Henry James' *The Portrait of a Lady*, characterized her husband when she says that 'he himself had not taken account of his intention' (James 1907-9, p. 436). Agents may be moved to act by a primary reason, decision, or intention without thereby act for a reason. As Velleman points out, these problems run deep on the Standard Model:

> *For let there be mental states and events in abundance – motives, reasons, intentions, plans – and let them be connected, both to one another and to external behavior, by robust causal relations; still, the question will remain how the existence and relations of these items can amount to a person's causing something rather than merely to something's happening to him, albeit something mental.*[77]

To paraphrase Bratman: it seems that concerning *any* mental act or occurrence, even an intention, one can raise the question of whether or not the agent is estranged from it.[78] And these possibilities indicate that the Standard Model misses out on something significant.

At this point one might be tempted to run in to defend the Standard Model by making use of the Anscombean observation that there may be many descriptions depicting the same *descriptum* (1957, pp. 37–47).[79] Perhaps one could account for Freud's case by saying that it is an action because it is intentional under a different description, such as the description Freud is stretching for his pen. One could then say that the latter is his intentional action whereas his breaking of the inkstand is one of his actions by being another description of this event (or descriptum) and then suggest similar stories for the other cases. However, that move is to completely misunderstand the dialectic at this point. The point about merely motivated actions is that the Standard Model qualifies them as intentional even under the description of merely motivated action when it should not. Simply adding that they qualify as intentional

[77] Velleman (1992b, p. 131).
[78] Bratman's words are: 'Concerning any mental act or occurrence, even a decision, one can raise the question of whether or not the agent identifies with it' (1996, p. 193).
[79] Anscombe here defends the so-called *Identity Thesis*, or the claim that true descriptions of me at an instant, such as me raising my arm and me signaling for a cab, describe the same event (or action). The thesis is also something that Davidson subscribed to (1963; 1967; 1971, p. 59). In contrast, Goldman defends fine-grained individuation of acts where many of the different action-descriptions that Anscombe-Davidson related as portraying the same underlying action is treated by Goldman as being different acts when there are certain asymmetries between the descriptions (1970, chapter 1). On Goldman's view, the response I entertain here is thus ruled out.

actions under another description as well will not remove this problem. The Standard Model still fails to make a distinction between acting for reasons and merely motivated behaviour. To see this compare these cases with Davidson's well-known case of the prowler: an agent flips the switch in order to illuminate the room but by that he also unknowingly alerts a prowler of his presence. Here the problem is to account for how the agent may be said to intentionally illuminate the room without thereby attributing the agent with intentionally alerting the prowler. The absence of a relevant belief-desire pair causing the event under the description alerting the prowler can explain this. However, we cannot make this move in the above cases; after all, Freud is saddled with the relevant belief-desire pair and it characteristically causes the event under the description breaking the inkstand. The same goes for Velleman and for Setiya's law student. They thus count as acting for reasons on the Standard Model and show that the Standard Model, while it seems to adequately capture merely motivated action, has failed to characterize what it is to be an action done for reasons.

Perhaps one might add something to the causal-motivational influence of one's desires, beliefs, decisions and intentions to separate two levels of explanation that apply to purposive actions. Then on the first level an agent's behaviour is explained as merely psychologically motivated by mental states of the agent, whereas on the other level the agent is acting for reasons. These would then represent two ways whereby we can do things that both are explainable by reference to our psychological states but where only the latter is the 'because' signalling intentional activity. What we might need, as Velleman puts it, is to find 'an agent at work amid the workings of the mind' (1992b, p. 131). According to Bratman, Setiya, Velleman, Korsgaard and others the crucial component that is lacking for merely motivated action to rise to the status of full-blown agency and acting for reasons is thus if the agent takes (or treats) the considerations that guide her action as her reasons.[80] In a word, *taking as one's reasons* is the locus of agency. The failure of the Standard Model is thus that it fails to take account for what ought to be the primary of object of inquiry. Merely motivated actions show that an answer to the *Primary Question* – if one is to be found – will require an account of what it is to take (or treat) something as one's reason for Φ-ing.

§7 TAKING as ONE's REASON and the GUISE of the GOOD

It is at this stage in the dialectics that philosophers of action are prone to invoke rationality and reason. Incidentally, it is also at this stage that the ethical rationalist may get

[80] See Bratman (1996; 2001, p. 91), Velleman (2000b, *Introduction*; 1992b) and Setiya (2007b, p. 35).

her hopes up. The relevant 'because' needed for intentional agency is perhaps one that 'shows rationality in operation,' as McDowell says (2008a, p. 257). Indeed, Davidson himself is inclined to entertain the idea that '[f]or a desire and belief to explain an action in the right way, they must cause it in the right way, perhaps through a chain or process of reasoning that meets standards of rationality' (1974, p. 232). So perhaps the Standard Model is at fault for failing to fully display this rational engagement? Intentional actions are, after all, truly our own and there is a tradition in philosophy for thinking that a person is more intimately bound up with his own rationality: as Aristotle says, 'intellect more than anything else *is* man' (*NE* 1178a6–7). Action theory may thus need to display that 'the agent was in some way thinking or heeding what he was doing,' as Ryle puts it, so that 'the class of actions done from motives coincides with the class of actions describable as intelligent.' (1949, pp. 110–1) For the rationalist this may be good news since an account of what it is to take something as one's reason may require what I shall call a *normative interpretation* of intentional agency where an agent may need to, at least implicitly, endorse certain normative standards for action and practical thinking in order to take something as her reason and act intentionally.

The reason why an agent's rational engagement seems well suited for ringing in the group of actions that are more our own than merely motivated actions is that rational engagement may be the kind of process that may ensure that reasons have the kind of 'regulatory influence' on one's behaviour, as Velleman puts it, and the presence of which 'can make the difference between an accidental collision and a shove' (2009, p. 36). The question is what role a person's rational faculty plays when she engages rationally in this manner so as to 'rise to the status of [full-blooded] action' (2009, p. 36). What must one do to let oneself be guided by the regulatory influence of one's reasons for acting? One thing is certain, and that is that one's attitudes need not embody conscious practical deliberation; so let me start with this disclaimer. When we act intentionally there is clearly something we do to steer our actions in virtue of which it makes sense, as Ryle says, to describe our performances as 'correctly or efficiently or successfully', 'intelligent', 'careful or skilful', and so on (1949, p. 28). However, there is no need to see this rationality in operation as conscious deliberation entertained in reflective moments. 'Intentions, like beliefs, are not always and necessarily the outcome of a process of thought or of a datable act of decision,' as Stuart Hampshire says; rather, 'they may like beliefs, effortlessly form themselves in my mind without conscious and controlled deliberation.' (1960, p. 101) So if it is true that practical thought is required – that is, if there must be some process of thought by which we regulate our actions and which

allows us to 'regulate one's actions' and be more than just 'well-regulated clocks keep[ing] good time' (Ryle 1949, p. 28) – the relevant notion of practical thought should be broad enough to include everything from conscious deliberation to mundane motivation resulting from non-conscious 'balancing of reasons' and 'forming and revising of intentions and desires' (Setiya 2007b, pp. 7–8). Consequently, we should not require that an agent must be conscious in the process of taking something as her reason for acting.

What would accommodate the problem is thus if one were able to identify a possibly non-conscious process or attitude by which an agent takes some *R* as her reason for Φ-ing when she is Φ-ing intentionally, a process that is not operative when someone Φ-es as a result of mere motivation. Now, as Velleman pointed out, the problems we encountered above require us to move beyond the Standard Model to find an 'agent at work amid the workings of the mind'. Mental states and events 'in abundance' and their robust causal connections both to each other and to external behaviour will not be enough; we also need to locate the agent's contribution to her behaviour. Only then can we avoid the 'fundamental problem in the philosophy of action' and the problem of 'finding a place for agents in the explanatory order of the world' (1992b, pp. 127, 131). According to Velleman, however, this conclusion still allows us to work under the assumption that adding and reversing bits of psychological machinery may enable us to identify, as he puts it, 'events and states to play the role of the agent' and thus are 'functionally identical to the agent' (1992b, p. 137).[81] Velleman thus insists that his objection to the Standard Model is 'not that it mentions mental occurrences in the agent instead of the agent himself'; rather, the point is that 'the occurrences it mentions in the agent are no more than occurrences in him' and that 'their involvement in an action does not add up to the agent's being involved' (1992b, p. 125). What we need, according to Velleman, is thus a story of events and states that plays the causal or functional role that ordinary parlance attributes to an agent as her contribution. Velleman is thus still working under the assumption that the *Primary Question* really is our primary question.

In order to traverse the problem of missing agents and mere motivation, Velleman notices that we need an attitude that is not among the 'mental states or events whose behavioural influence might come up for review in practical thought at any level'. He concludes that the attitude must rather 'be played by whatever directs that process'; so he suggest that we identify the attitude (or complex of attitudes) that plays the functional role of

[81] Velleman's account provides a paradigmatic example of this strategy (1992b; 2000b, pp. 7–24). Also see Bratman (2001). Some philosophers take the problems of the Standard Model to dispute the whole event-based model. I have some sympathy with this, but it would be premature to say more about it now. I will simply go on as if a fix to the Standard Model is possible. By so doing, I hope to show that some accounts fail even on their own terms, as it were. Meanwhile, see Hornsby (2004a; 2008; *forthcoming*) and Lucy O'Brien (2007, ch. 8).

the agent with the attitude that 'embodies the concerns of practical thought per se' (1992b, p. 139). This may sound complicated; however, what Velleman has in mind is that we invoke some version or another of a *constitutive aim theory* for intentional actions (or a *CAT*, for short) and then identify the role of an agent with the attitude that embodies an agent's concern for this *constitutive aim* (or *CA*). A *CAT* defines intentional actions (or actions done for reasons) as behaviour geared towards or guided by some *CA* or another of actions. Candidate *CAs* for intentional agency are the satisfaction of desires, acting for the sake of the good, reflective endorsement of oneself and what one does, acting in a way that will promote self-understanding, and so on.[82] The goal identified by the specific *CAT* need not be something explicit that the agent entertains consciously or with a high degree of awareness. Rather, the concern that animates practical thought is presumably a sub-intentional desire for acting in accordance with reasons and so something that operates within us – whether we know it or not – as a constitutive element in the process of acting for reason. The *CA* thus constitutively regulates our actions whether we like it or not.[83] What *CAT* theorists claim is thus that one acts intentionally to the extent that one's behaviour aims at achieving the constitutive aim of intentional actions. What Velleman suggests as a solution to the fundamental problem of action is thus that we functionally identify the role of an agent with the cognitive mechanism or attitude that embodies one's concern for the *CA* of action. The attitude that animates practical thought and organizes our thoughts and behaviour in accordance with some candidate *CA* is thus incidentally also the attitude that plays the role of an agent.

The clue to an account of what adds up agent-involvement – viz. to the functional role 'definitive of an agent' (Velleman 1992, p. 138) – is thus a certain normative or evaluative stance on part of the agent that is embodied in one's practical thinking and drives agents to steer their behaviour in accordance with the *CA* of intentional agency. The regulatory influence of one's reasons on such a *CAT* model is thus accounted for as a sort of normative regulation where, to put it roughly, the agent adjusts her behaviour in accordance with the influence of the better reason where reasons for acting are defined in terms of the *CA*. In short, the solution Velleman offers to the problem is that an agent takes some *R* as her reasons for Φ-ing only when she takes *R* to be a good or normative reason for doing what she does. What goes wrong in the problematic cases for the Standard Model can thus be diagnosed as

[82] For accounts identifying desire-satisfaction as the *CA* of actions, see Williams (1980), Gauthier (1986), Smith (1994), and Dreier (1997); accounts emphasizing 'the good', see Raz (1999; 2010); self-knowledge as a *CA* is explored by David Velleman (2000b; 2006; 2007; 2009); and for a reflective endorsement account, see Korsgaard (1996; 2008a; 2009).
[83] See Setiya (2007b, p. 85). This is also emphasized by Velleman (2000b, p. 21).

its agent's failure to take the considerations by which he or she is moved to act as good reasons for acting; as Velleman concludes his discussion of *Freud's Inkstand*:

> *A reason for acting is something that warrants or justifies behaviour. In order to serve as the basis for a subject's behaviour, it must justify that behaviour to the subject – that is, in his eyes – and it must thereby engage some rational disposition of his to do what's justified, to behave in accordance with justifications.*[84]

In short, what Velleman invokes is the idea that one's reasons for acting must be 'motivating in virtue of their normativity,' as Korsgaard puts it,[85] and that taking some *R* as one's reason for Φ-ing is to take *R* as a good or normative reason that favours or justifies Φ-ing.

By bringing in this normative stance as an essential element of taking as one's reason, Velleman commits to a version of what is known as the *Guise of the Good* doctrine, or the claim that when agents act intentionally or for a reason they act *sub specie bonie* and under 'the guise of the good'. Velleman does this in order to identify the agent's role with the mechanism by which agents take something to be a normative or good reason for Φing; however, by so doing he ties his solution to a long-standing tradition in Western philosophy.[86] The roots of this traditional doctrine can be traced back to Ancient Greek philosophy, as when Socrates in Plato's *Republic* claimed that '[e]very soul pursues the good and does its utmost for its sake' (*R* 505e). A similar doctrine shows up in Aristotle's moral psychology when he says that 'every action and choice [*prohairesis*] is thought to aim at some good' (*NE* 1094a1), or that 'it is always the object of desire which produces movement, [and] this is either the good or the apparent good' (*OS* 433a27–29). It was later buttressed by Scholastic philosophers as the doctrine that willed or chosen actions are always done *sub specie bonie* – or, as Raz writes, *sub ratione bonie* (2010, p. 135n1) – since, as Aquinas says, in action '[r]eason comes before the will and directs its activity' (*ST* 2.1.13a.1; 1a.2ae.8.1).[87] Now these claims reveal that there are different versions of the doctrine. A controversial reading of the Socratic/Platonic doctrine says that it is only the Good itself that attracts and motivates human agents. A more psychologically realistic reading of the doctrine, however, says that human beings are guided by and steer their actions toward the apparent good and that they do what they take to be good actions, either in themselves or as good means for realizing something they conceive as a good outcome. In what follows it is the latter version that is of

[84] Velleman (2000b, p. 9).
[85] Korsgaard thinks that reasons must be 'normatively motivating', as opposed to merely motivating behaviour (2008a, p. 209).
[86] Stocker simply refers to this as 'the philosophical view' (1979, p. 740).
[87] I will refer to the Novantiqua 2008 edition of Thomas Aquinas' *Summa Theologica*.

interest given that our focus is on the state of mind in which an agent must be for his or her behaviour to rise, as it were, to the status of intentional action. The hypothesis should therefore be roughly cashed out as a requirement on the psychology of the agent when she does something that can be categorized as intentional (or done for reasons):

> *The Guise of the Good*: If *A* is Φ-ing for a reason (or intentionally), *A* ***must*** see some good in Φ-ing on this occasion.

Different ways of cashing out this doctrine will depend on: (1) what one takes to be a good reason for doing something (viz. what the *CA* or standard for action is); (2) what one takes to be the source of the requirement; (3) how one accounts for the nature of the attitude whereby one takes something as one's reason; or (4) which class of actions are governed by the doctrine. We have already briefly touched on (4): the interesting versions of the doctrine ranges over the class of action where an agent acts for a reason. If this doctrine holds for this class of actions, agents must take their reasons to be good reasons for Φ-ing. As for (1), the doctrine is stated above using 'good' as a multiform notion denoting whatever standard of goodness one takes to be constitutive for intentional agency. Earlier I mentioned candidate *CA*s: desire-satisfaction, the good, reflective endorsement, self-understanding and so on. Korsgaard, for instance, thinks one must act in accordance with one's practical identity, which is an evaluative self-conception or 'a description under which you find your life to be worth living and your actions to be worth undertaking' (1996, p. 101). Raz takes good reason to be considerations that speak in favour of the good-making features of the action (1999).[88] When one settles for one of these *CA*s one can define normative reasons for acting as the considerations that speak in favour of an action by justifying the action as the way to implement the *CA*. If one opts for desire-satisfaction, there is reason to Φ if Φ-ing is a way to satisfy one's desires in this situation. 'The guise of desire' would then be the doctrine that an agent, *A*, must take her reason *R* for Φ-ing to justify that Φ-ing is a way for *A* to satisfy her desires. As the *Guise of the Good* is stated, it is meant to subsume these versions.

The source of the requirement (2) is a more complicated matter as it depends on the nature of acting for a reason. Clearly, if the doctrine is true for all actions whose agent acts for a reason – or intentionally – it must pertain to something that is essential for intentional agency. The fact that one takes one's action to be good must therefore be a constitutive

[88] Raz endorses a version of the doctrine when he explicitly requires an agent to see her actions as good actions and that the agent must be aware of his reasons as good reasons for acting (1999, pp. 23, 23n5 and 37; 2010, p. 134).

element of acting for a reason. Where one wants to locate it then depends on what one takes to be constitutive elements of acting intentionally. That, however, requires a view about what the constitutive elements are. Thus, dealing with point (2) brings us in touch with point (3). If we accept the Standard Model, for instance, a desire (or a pro-attitude) is a necessary element in acting for a reason. It could then be argued that agents acting for a reason must see some good in their action because there is a necessary relation between having a desire and seeing some good in what one desires. That is, one may endorse an *evaluative conception* of desire. Perhaps one thinks, as Anscombe does, that there is a conceptual connection comparable to connection between 'judgement' and 'truth' and thus that it is '*required* for our concept of 'wanting' (...) that a man should see what he wants under the aspect of some good' (1957, pp. 75–6). The evaluative conception of desire may then support the *Guise of the Good* as a doctrine of intentional actions. For if a desire is necessary in order to act for a reason and having a desire entails seeing some good in what one is doing, it follows that the agent must regard her action as having some value, be good, or desirable (in the sense of warranting the desire).[89] Similar considerations may follow from thinking that acting for a reason requires an intention that necessarily involves the judgement that one's action is good.[90]

Another way to go about it is to follow Velleman and argue that taking as a reason is a constitutive element of acting intentionally and then analyze this notion in terms of an evaluative conception of taking as one's reason. As Velleman argued, the problem of accounting for agent-involvement and the functional role 'definitive of an agent' in action requires a notion of taking as one's reason in terms of rational guidance (1992a, p. 103; 1992b, p. 138). One takes something as one's reason for acting only when the action-justifying reason is treated in the eyes of the agent as if it 'reflect[s] well on, count[s] in favor of, recommend[s], or in some other sense justif[ies] the action' (1992a, p. 100). Only if that holds can an agent be said to regulate her behaviour in the right way so as to act for or in the light of reasons. The *Guise of the Good* thus operates as a crucial element in accounting for what it is to act for a reason by shedding light on one of its constitutive components: viz. taking as one's reason. In a word, the doctrine can earn its living in action theory by helping us to solve identify the functional role of an agent in action. Since the doctrine is primarily a

[89] See Velleman (1992a, p. 99). Jennifer Hawkins (2008) defends what she calls an evaluative conception of desire and the idea that to desire something is to see some good in what one desires. Hawkins emphasizes, however, that this notion of desire cannot be equated with the class of pro-attitudes. Since it is only the latter that is a plausible constitutive element of intentional action, Hawkins argues that the evaluative conception of desires does not give support to the *Guise of the Good* as a doctrine about intentional actions or actions done for reasons.

[90] Grice's notion of a *willing* involves such a judgement (1971, pp. 278–9). As Davidson points out, this idea makes it a problem to understand *akrasia* or weakness of will, where an agent acts against the recommendation of his own evaluative judgement (1978, pp. 98–9). Davidson took this to require a *sui generis* pro attitude: the *all-out* judgement that *this* action is the best. The all-out judgement may, by making this verdict, depart from one's evaluative judgements since what one takes to be best may not match with what one takes to be good.

requirement of the agent's practical thinking and a condition on what it is to take something as one's reason, however, the best way to phrase that hypothesis is as an evaluative condition on how the agent must relate to her reasons for acting when she is acting for a reason:

> *GG:* When some agent, *A*, is Φ-ing for a reason, *R*, she *must* take *R* to be a (*good*) reason *for* Φ-ing (at least to some extent).

A reason why *GG* is preferable is that it portrays the doctrine as a universal claim on what an agent must do to take something as one's reason for acting – viz. that one must take it to justify or favour one's action – without assuming anything about what an agent must do if it turns out that it is possible to act intentionally for strictly speaking *no* reason, just for a whim, because one felt like it and so on. Some will deny that this is possible.[91] For them it follows that one acts for a reason insofar as one acts intentionally; so *GG* is for them a way to explain the truth of the first formulation. However, inspired by some remarks in Anscombe's *Intention*, one may seriously entertain the idea that agents may act intentionally without acting for any reason at all, or at least act intentionally for no particular reason (e.g. for just a whim). In the first case, it is unclear whether proponents of the *Guise of the Good* are committed to explain these actions as somehow being enacted because their agent saw some good in it; and, if they are so committed, whether they can. After all, how can one take some action to be good or justified if there is, literally speaking, no reason for which one is doing it? In the second case, it seems as if we have a genuine counterexample to *GG*. Against the first problem defenders of the *Guise of the Good* have two options: either (1) reject the claim that one can act intentionally for literally speaking no reason; or (2) qualify one's claim as *GG*. The second problem, however, is a problem even for *GG* and thus presents its defenders with some resistance. I shall return to this in a short while. First, let me comment on the connection between this doctrine, Velleman's solution and Ethical Rationalism.

As Setiya points out, the interesting versions of *GG* and the *Guise of the Good* – viz. those that has something to offer to the problem that Velleman deals with – are not merely trading on the platitude that '[w]e tend to (…) choose what seems worthy of choice, and to act in ways we think we can justify, at least to some extent' (Setiya 2010b, p. 82). It is obvious that good practical thinking is responsive to reasons in this weak or generic sense. The doctrine is, rather, meant as a particular explanation and interpretation of such platitudes. It

[91] Raz, for instance, claims 'that one can only choose or decide [to act intentionally] *for* a reason' before he explains this as meaning one must act 'for what one takes to be a good reason for the option chosen' (1999, p. 8). As Setiya notes, Raz thereby claims two things: (1) intentional actions must be done for a reason; and (2) reasons must be seen under the *Guise of the Good* (Setiya 2007b, p. 36n25).

makes a bold claim about the nature of rational agency as such in a way that may allow us to diagnose the failure of the Standard Model. As a claim, it pertains to the essence of rational agency and intentional actions – or, rather, actions done for a reason – and must, as a result, hold in any possible worlds where someone acts for a reason or intentionally.[92] It is not sufficient if it merely turns out that we tend to act in ways we see as being in some way good; nor is it good enough that if this is a fact about human psychology. Rather, its claim is that intentional agents can no more help justifying their intentional actions than 'stomachs can help grinding up foodstuffs', to borrow a phrase of Richard Rorty's (1999, p. 38). Conversely, it follows, as Velleman says, that 'I cannot act for reasons if I don't care about doing what's justified or (as I would prefer to put it) what makes sense' (1992a, p. 121).

Its universality is what allows it to work as a qualification on what the regulatory influence of one's reasons must be on one's behaviour when they are one's reasons for so acting and which, in turn, may allow us to explain the difference between acting for a reason (or intentionally) and merely motivated actions. Now, as I mentioned earlier, the idea that an account of acting intentionally must allow us to see rationality in operation will also give some hope to an ethical rationalists. The reason is roughly that a candidate *CA* also works as a normative principle that carves out specific standards applying to actions. In virtue of being a principle that constitutes intentional agency it thus provides a standard for what it is to act well when one acts intentionally or for a reason. Again, to use Rorty's analogy: the stomach's capacity as an organ for grinding up foodstuffs constitutes it as the organ it is. However, its function as a food-grinder also provides it with a standard of adequacy that allows us to ascertain whether or not a particular stomach is working properly. What the rationalist may hope for is that we can say the same about acting intentionally or for reasons: namely, that the fact that there is a *CA* for intentional action that provides it with its own distinct standards of adequacy. In turn, one may use the standards for what it is to be a good action to derive good standards of practical reason so as to meet the *Difference Principle* by enabling us to appreciate something distinctive about the nature of actions and practical thought. Moreover, given that being guided by a concern for the *CA* is what constitutes agency this concern cannot be disowned by an agent while retaining her status as an agent. One must, after all, be appropriately sensitive to this *CA* in order to act intentionally. Insofar as one acts intentionally one thus commits to the *CA* as a standard of excellence for one's actions.

[92] For a discussion of this notion of essence, see Fine (1994; 2000, p. 543).

The relevance to our debate is thus that something like Velleman's solution to the problem of action – viz. the idea that there is a distinct *CA* pertaining to actions the rational sensitivity to which just is what it is to take something as one's reason for acting – may allow the rationalist to identify the distinct nature of practical reason and explain why agents must be committed these standards. If we know what a good action is – viz. one that satisfies the *CA* of actions – we are also in position to specify what good practical thinking should aim at when all goes well: namely to choose and enact the behaviour that optimally implements the *CA*. This may allow rationalists to meet the *Difference Principle* and derive standards of practical reason from the nature of action. Of course, the truth of the other premises in the rationalist project still need vindication; however, an evaluative conception of taking as one's reason combined with an account of what it is to act for a reason that is similar to Velleman's solution, will take them a long way by granting rationalists the truth of their claim that there is a real distinction between standards of practical reason and moral standards.

§8 AGAINST the GUISE of the GOOD

In order to reject Rationalism, Setiya embarks on a complicated argument to show that different candidate constitutive aims (CA) of action are false by arguing that the whole idea – viz. that actions have such CAs – on reflection turns out to be nothing but 'a philosophical mirage' (2007b, p. 1). According to Setiya, there is reason to flat-out reject the different versions of the *Guise of the Good* doctrine. As a consequence, there is little reason to believe there exist CAs of action that would allow rationalists to derive distinct normative standards of practical thought. The image of human agency that emerges instead is one that severely restricts the materials to which the rationalist can appeal in deriving independent standards of practical reasoning. Thus, the prospect of meeting the *Difference Principle* looks bleak. Setiya therefore takes his arguments against different versions of the *Guise of the Good* to 'leave no room for the rationalist project to succeed' since there will be no basis for a real distinction between the standards of practical reason and virtues of character once we give up that doctrine (2007b, p. 70). In this section, I shall expound Setiya's argument.

There are two main lines of argument against the *Guise of the Good* (or *GG*).[93] The first line of argument is familiar in the literature and relies on counterexamples that allegedly portray agents who – due to some peculiar kind of irrationality, perversity of mind, depression, or sheer indifference – act for reasons they do not take as good reasons for so

[93] For simplicity's sake, I shall refer to *GG* and the *Guise of the Good* collectively as the *Guise of the Good*. We should bear in mind, however, that we are primarily interested in it as an expression of the evaluative conception of taking as one's reason.

acting.[94] If we accept such cases, we will have counterexamples to the *Guise of the Good* or instances of agency where the regulatory influence of one's reason for acting is not a case of the reason being motivating in virtue of its normativity or justifying character. Since the doctrine is a principle or universal requirement and so cannot take exceptions – 'even if,' as Geach says, 'the exceptions are events that only James Joyce would put into a novel' (1957, p. 5) – the counterexample would give us a reason to reject the doctrine.

A case in mind is Gary Watson's story of the squash player who 'while suffering an ignominious defeat desires to smash his opponent in the face with the racquet' (1975, p. 210). If the squash player works to control himself, his desires will in a way be banished from consideration even though he may have them and experience the distress of not satisfying them. This is similar to how McDowell's courageous person may be in a state of mind where '[h]ere and now the risk to life and limb is not seen as *any* reason for removing himself' (1979, p. 56).[95] Likewise, the squash player will silence what could otherwise be considered as a reason to engage in the desired activity. If, however, the squash player should succumb to his desire and go running after his opponent, racquet raised ready to smash his face, we would say that he acts on these desires despite himself. He would act intentionally and treat his desire as a reason to smash his opponent's face; however, he may not be considering it is as a reason favouring this activity. One might, of course, say that not satisfying such a desire may be the source of frustration in such a hot-tempered agent so that its occurrence is a reason to get rid of it. However, this will only count as a reason for doing *something or another* to discharge one's desire. Rather, than favouring smashing one's opponent's face the desire counts in favour of concentrating on some boring event, taking a cold shower, and so on. When the squash player succumbs to his desire the point is thus that he may go on waving his racquet against his opponent while thinking all the while that this is crazy and that his desire is not something that favours this ridiculously childish and aggressive behaviour. If this is possible, we have a counterexample to the *Guise of the Good*.

Another case in mind is the agent who is doing something she finds pleasurable yet fails to recognize as worthwhile. As Myles Burnyeat has pointed out, there is some learning involved in enjoying something that may distinguish such proper enjoyment from 'merely

[94] For some classical discussions of such cases see Augustin (*Confessions* 2.4), Stocker (1979; 2004) and Gary Watson (1975; 1977; 1987; 2003). Also see Bratman (1996), Velleman (1992a), Frankfurt (1987; 1998a), Arpaly (2000; 2003, ch. 2) and Setiya (2007b, pp. 36–8).

[95] McDowell thinks that *silencing* considerations is a general feature of having a virtue since the virtuous person then understands notions such as 'benefit, harm, loss and so forth' in such a way that no sacrifice necessitated by the virtue counts as a loss. If, for instance, one is courageous, the danger to one's life and limbs is not overridden and outweighed by the gains of being courageous; they are rather silenced and will be conceived as no 'reason at all' or 'no loss at all' (McDowell 1980, p. 17). Also see Hursthouse (1999, p. 96n8). I do not take a stand on the adequacy of this as an account of what it is to be virtuous. All I need here is the idea that such silencing is possible but that agents may nevertheless succumb to their force and then act for considerations they have silenced.

taking pleasure' in something (1980, p. 76). That allows room for doing something that one merely takes pleasure in while finding these pleasures not to be worth one's while in the least since one now knows better.[96] Someone may for instance take pleasure in philosophy for the sense of power it can give and use her position to ask 'derisive questions at talks because that will humiliate the visiting speaker' and so on (Setiya 2007b, p. 37). Or suppose someone is taking pleasure in her holiday to Paris only because 'of the splendid photographs' he is taking (Burnyeat 1980, p. 76). The latter may have learned to disapprove of these petty pleasures, yet carry on due to a certain volitional inertia knowing all along that his touristing in Paris is not worthwhile in the least. Similarly, the malicious academic may behave as she always does with visiting speakers while despising herself for succumbing to such a habit. In both cases, its agent may just feel like doing something and then do it for those reasons without taking these reasons to make their actions worthwhile at all.[97]

In general, the nature of these counterexamples depends on the version of the *Guise of the Good* they are meant to dispute. If one argues against a *Guise of Desire*, one must find cases of agents where they do intentionally (or for reasons) what they find not to be desirable. Similarly, if one argues by counterexample against Velleman's *Guise of Self-Understanding*, one must find agents Φ-ing for reasons they do not think make their action understandable or sensible. The general idea is to find cases disproving the claim that agents Φ-ing for a reason *R* must take *R* to be a good reason for Φing, where the goodness of the action is defined in terms of the *CA* for actions. One succeeds if one can show that some agent, *A*, due to reckless irrationality, carelessness, or what have you, Φ-es for a reason *R* without taking *R* to be a good reason for Φ-ing. If one can act for a reason even if one's frame of mind goes beyond the borders of irrationality that, as Raz says, portrays '[e]ven irrational choice [as] choice for a reason (which the chooser believes to be at least a good *prima facie* reason)' (1999, p. 8).

As Setiya argues, these counterexamples need not portray irrational agents and other strange perversities of mind to achieve this. Since he believes that 'in general taking something as one's reason is not a matter of taking it as good' it is also open for him to treat the actions of naturally virtuous agents as support for his view given that their actions may show that '[e]ven fully rational agents, and fully virtuous ones, need not believe in the merit of the reasons for which they act.' (2007b, p. 38) Iris Murdoch once pointed out that 'an unexamined life can be virtuous' as it 'is perhaps most convincingly met with in simple

[96] Such an agent will resemble the incontinent agent who 'acts from his appetite but not from his choice,' as Aristotle says (*NE* 1111b13–15). Such an agent will not choose his action, in the special sense of choice (*prohairesis*) that Aristotle operates with; however, as David Wiggins points out, this is not yet to say that 'he acts non-intentionally in the ordinary sense of the English word *intentionally*' (1978-9, p. 253).
[97] For more cases, see Setiya (2007b, pp. 36–7). Also see Quinn (1993b). For the notion 'volitional inertia', see Wallace (2001, p. 91).

people', such as the 'inarticulate, unselfish mothers of large families' (1970, pp. 51–2).[98] Murdoch's point is that there is no need to be reflective in one's practical thinking in order to be morally virtuous since '[w]here virtue is concerned we often apprehend more than we clearly understand' (1970, p. 30). What the virtuous person must do – and this, as Aristotle said, is not easy and 'not for everyone' – is to do what is good and right, and do it 'to the right person, to the right extent, at the right time, with the right aim, and in the right way' (*NE* 1109a26-29). A good person will do the right thing when the situation calls for it; however, this capacity can be had even if the virtuous person fails to understand her actions as good, or fails to take her reasons as reasons for doing something. A virtuous person may simply be one who is properly sensitive to her practical situation and gives weight to the right considerations. That requires something of the agent: perception, an aptitude for finding opportunities to do good, motivational propensities, and so on.[99] But there might be no need to saddle her with an attitude that she takes there to be reasons for doing these things. Rather, one may be moved to act 'directly and unconditionally by beliefs whose objects are the reasons themselves,' as Setiya says, so that one's motivation is partly independent of explicit beliefs about reasons for doing something (2007b, pp. 72–2).[100] Acting without taking one's reasons to be reasons for doing something may therefore not be restricted to the motley crew.

Setiya thinks the above cases count against the *Guise of the Good* since he believes it is simply naïve to insist that the agent must see something good in what she is doing. His critical point thus goes beyond mere psychological accuracy with respect to some freak cases; rather, he aims to lay bare how a widely shared conception of acting intentionally is fundamentally wrong and ought to be discarded. However, Setiya also believes – correctly in my view – that a *top down strategy* that puts emphasis on counterexamples is inadequate to prove such a general point. For one thing, arguing with case-descriptions 'is likely to lead to stalemate' (2007b, p. 22, 37). The cases are hard to elaborate in a neutral way and it is more than likely that the proponents of the *Guise of the Good* may find ingenious ways in which to

[98] The point is not that we need to judge the naturally virtuous person 'simple', or think of them as 'inarticulate' creatures of the *hoi polloi*. After all, one can be an articulate aristocrat and yet quite unreflective in one's practical thinking. Countess Gemini in Henry James' novel, *The Portrait of a Lady*, exemplifies this when she displays a reckless carelessness summed up as the following attitude towards life: 'I don't care anything about reasons, but I know what I like.' (1907-9, p. 260) The Countess's opposite is found in the novel's heroine, Isabel Archer, who says that 'One ought to choose something very deliberately, and be faithful to that.' (1907-9, p. 267) Taking Countess Gemini as our model, we could well imagine a chattering, sophisticated, yet morally unreflective and naturally virtuous aristocrat and we need not assume, as Dostoevsky seems to do in *The Idiot* (1868-9), that goodness enjoins simplicity, inarticulateness and silence. See Setiya (2007b, p. 17).

[99] For discussion of some of the features a virtuous agent must embody, see Setiya (2007b, pp. 70–9). McDowell, in particular, discusses the connection between the virtuous agent's sensitivity and motivational propensities in (1979, §V; 1995b; 1996; 1998e, p. 47).

[100] Hursthouse argues that the courageous agent is moved immediately to act by thinking things like 'I could probably save him if I climb up there,' or 'No one else will volunteer' and that the temperate person acts on thoughts like 'You want it more than I do,' or 'The cheaper one will do the job' (1995, p. 25). As Setiya remarks (2007b, p. 73n7), Hursthouse suggests this as a way to interpret Aristotle's claim that the virtuous person decides on virtuous actions for themselves. Also, see Williams (1995, pp. 17–8).

make sense of the alleged counterexamples.[101] One thus risks ending up with a dialectical situation where intuitions simply disagree. So while Setiya finds the counterexamples compelling, he also finds the strategy unsatisfying. Another legitimate worry is that case descriptions will fail to illuminate what it is about human agency and our reflection on its nature that has moved philosophers to assume something like the *Guise of the Good*. If we accept them as counterexamples, one could still think of the *Guise of the Good* as saying something of importance about the nature of rational agency and that one needs to explain it insofar as one is explaining the nature of acting intentionally.

This will cut both ways: a top-down strategy with emphasis on individual cases will therefore not be sufficient to establish the *Guise of the Good* either. Anscombe, for instance, argued that someone could not simply want a saucer of mud without seeing some good in having such a thing (1957, pp. 70–1). Her case may (or may not) be compelling; but, as Setiya notes, '[a]part from a more theoretical investigation of rational agency, we won't know how seriously to take such appearances – that Anscombe's example is possible, or that it isn't' (Setiya 2010b, p. 83). What we need are general considerations that may settle the issue. In order to respect this requirement, Setiya develops a general argument from the *bottom up* to show that there is no need to conceive the act of taking some *R* as one's reason as a matter of taking *R* as a (good or normative) reason for doing something. In other words, he argues that we may understand the regulatory influence of one's reasons on one's behaviour in a non-normative or non-evaluative way. His bottom up argument is thus to show that an account of what it is to take something as one's reason for acting does not need to invoke a normative or evaluative conception of the regulatory influence of one's reasons.

To implement his strategy Setiya begins from what he takes to be general constraints on a theory of action and he then proceeds to 'show how these constraints rule out a normative interpretation of the recognition and adoption of reasons involved in acting on them' (2007b, p. 22). One constraint in particular is a general puzzle that Setiya thinks was left unanswered by Anscombe's discussions in *Intention*: namely, the question of 'why two marks of intentional action go together: its susceptibility to explanation in terms of reasons, and its being subject to what Anscombe has called "knowledge without observation"' (Setiya 2007b, p. 22–3). He is convinced one can account for what it is to act intentionally while explaining these marks of intentional agency 'without appeal to normative or evaluative thoughts' (2007b, p. 61). On his account – which I shall refer to as *Minimalist Cognitivism* or

[101] See for instance Sergio Tenenbaum (1999; 2007).

Minimalism – these marks of intentional agency are explained if we explain taking some *R* as one's reason for Φ-ing as having a desire-like belief that one *hereby* (viz. by having this desire-like belief) Φ-es because one believes *R*. Setiya suggests intentional actions are the range of actions that are causally sustained by such relevant desire-like belief. If these claims are true, Setiya can offer an account of what it is to 'decide upon our reasons' that does not equate taking as one's reasons with deciding 'that they are good or decisive reasons'; all we decide is 'that they will be the reasons on which we act' (2007b, p. 61). I can thus Φ for some reason, *R*, because I take *R* as my reason for acting, without taking it to be a good reason for Φ-ing. Minimalism thus allows cases of acting for reasons where its agent does not take her reason as a good reason for acting. If that view is sufficient, we must reject the *Guise of the Good* and similar doctrines. In a word, we can dismiss the normative interpretation of agency.

For Setiya the result of the bottom up strategy decides the truth of the *Guise of the Good* and the evaluative conception of taking as one's reasons. The bottom line of his argument against Rationalism is thus that a normative interpretation of intentional agency is uncalled for since we can account for what it is to act for a reason without an evaluative conception of taking as one's reason. The function of Minimalism is thus to provide a crucial lemma in this argument by discarding any conception of agency that invokes normative principles. We are supposed to see that Minimalism is a sufficient account of what it is to act for a reason and in turn see that we can reject the *Guise of the Good* and any conception of agency that relies on this doctrine, including Ethical Rationalism.

If Setiya's bottom-up argument succeeds, the fate of Rationalism will be settled, by and large. It is therefore his strategy that will be put under serious scrutiny in the chapters that follow. In particular, I will discuss and ultimately reject the sufficiency of Setiya's Minimalism and so dismiss his crucial lemma in the argument against Rationalism. I will start out (ch. 2) by discussing the phenomenon that is meant to lend Setiya motivation for his view: agents' knowledge and a puzzle from Anscombe's *Intention*. I will then discuss in (ch. 3) how Setiya takes this to motivate Minimalism before launching criticism against his view (chs. 3 and 4). I consider and reject an inferentialist account of agents' knowledge (ch. 5) before going on to argue (chs. 6 and 7) that the a way to accommodate the problems we have seen so far is by invoking something akin to Anscombe's notion of practical knowledge. It will then become clear that the nature of intentional agency may very well lend support to Rationalism.

2

Agent's Knowledge and Action Theory

> *The identity of the subject of willing with that of knowing by virtue whereof (and indeed necessarily) the word "I" includes and indicates both, is the knot of the world, and therefore inexplicable.*
>
> Arthur Schopenhauer[1]

§1 INTRODUCTION

According to the 19th century German philosopher Arthur Schopenhauer, the identity of the subject as agent with the subject as knower constitutes 'the knot of the world' and is a deep mystery. Though certainly a true identity it was, in his view, unfathomable, a relationship he could not pin down and thus declared to be the philosophical problem *par excellence*.[2] From within the framework of Kantian metaphysics and transcendental idealism, it appeared to Schopenhauer that this mysterious identity bound together the realm of *noumenon* (or the intelligible ground Kant hypothesized was the unconditioned and essentially unknowable ground of phenomenon (Kant *CPR* p. 112)) and the realm of *phenomenon* (or the world of appearance).[3] Unlike Kant, however, Schopenhauer observed that a peculiar insight of the world originates when the subject exercises her Will, becoming thereby active as an agent. Schopenhauer was convinced that this peculiar insight of the agent provided a sort of glimpse into the nature of the noumenon. Since a glimpse of the noumenon would concern what is essentially unknowable, however, Schopenhauer inferred that the agent's insight could not possibly be knowledge. Knowing would, after all, condition its object – or so an idealist thinks – and so cannot have the unconditioned as its object. By

[1] Schopenhauer (*FR* §42). I will be referring to the 1877 4th edition of Schopenhauer's *On the Fourfold Root of the Principle of Sufficient Reason*, as this work is translated and reprinted by K. Hillebrand in Schopenhauer's 1903 *Two Essays by Arthur Schopenhauer*.
[2] According to Schopenhauer, *The World as Will and Representation* is devoted to an explanation of an aspect of this complex: namely, the identity of the will and of the human body, and he tries to explain the latter as 'the objectification of the will' (*WWR* §18). My references to this work are to the 1969 edition translated by E.F.J. Payne.
[3] My references to the *Critique of Pure Reason* are to the 1998 edition of P. Guyer and A.W. Wood.

combining idealism with the idea that acting gives a glimpse of the noumenon, Schopenhauer was thus led to locate two radically different aspects of the world within the rational subject.

It might be tempting to diagnose Schopenhauer's 'knot of the world' as just a symptom of failure in Kantian metaphysics and of little interest beyond that particular project. However, the debates that have ranged in philosophy of action since Elisabeth Anscombe wrote *Intention* and Stuart Hampshire published his *Thought and Action* indicate that the relationship between the subject as knowing and the subject as active agent is a more general problem.[4] What Schopenhauer, Anscombe, Hampshire and others seem to have noticed is the peculiarity of the insight an agent has as a result of exercising her agency. *Pace* Schopenhauer, however, I shall leave idealist qualms aside and refer to this peculiar insight as *agent's knowledge*.[5] The idea is roughly this: insofar as an agent is acting intentionally – or, which I take to be more or less the same, when she acts for reasons – the agent seems to know the nature of her own action by virtue of being its agent. In particular, an agent knows what her action is by knowing a description that structures her movements and activities under an organized whole; and she knows why she is doing it by knowing her reasons for so acting.

In what follows, I shall discuss the peculiarity of agent's knowledge and I argue that its distinct nature and its pervasiveness in intentional agency offer an important explanatory constraint on or guideline for action theory. In my view, fully understanding the nature of intentional agency requires us to understand the nature and pervasiveness of agent's knowledge. Moreover, its necessity and nature will us to see intentional agency – or rather, the activities we exercise as agents with respect to the events that are our actions – as a particular way of knowing. I shall start out (§2), however, with Anscombe's claim in *Intention*. There is an intimate connection, she says, between acting intentionally and acting knowingly, and agents, insofar as they are acting intentionally or for reasons, must know in a particular way the nature of what they are doing. My aim here is to provide a preparatory

[4] As Thor Grunbaum points out, knowledge of our own intentional actions is 'an essential element' of so acting; however, 'it has proven extremely difficult to unpack what kind of knowledge an agent must possess of her own action if it is to play this essential role' (2009).

[5] By 'agent's knowledge' I shall mean that peculiar kind of knowledge, insight or epistemic access an agent has of the fact that she is Φing and why she is Φ-ing – or will Φ if the knowledge is future-oriented – that obtains insofar as she is Φ-ing intentionally. This is the knowledge that Anscombe famously referred to as *practical knowledge*, relating to a notion Aquinas speaks of as 'the cause of what it understands' (Aquinas *ST* 1a.2ae.3.5.1; Anscombe 1957, p. 87). I will discuss features of this knowledge in §4. It is important not to confuse this notion that of the practical knowledge to which one may refer in everyday parlance as the knowledge one has in virtue of which one is able to perform a certain task. The latter is, perhaps, best rendered in terms of what one knows how to do, something many take Ryle to have argued is a *sui generis* kind of knowledge that is not reducible to propositional knowledge or knowledge that (1949, ch. 2). There are connections between these notions since bits of knowing how may be required for agent's knowledge to occur. Setiya, for one, suggests agent's knowledge must be explained through its connection with knowledge how (2008b, p. 388). For the sake of clarity, however, they should be kept apart. It is also worth bearing in mind that by agent's knowledge I do not mean the total amount of knowledge an agent has when she acts for reasons. Her total knowledge may include conditional propositional knowledge, background knowledge, knowledge of her capabilities and, again, knowledge how. That body of knowledge will surely include knowledge which is known only observationally and may indeed be necessary for enabling one to act intentionally; however, if what I say below is correct agent's knowledge is a distinct epistemic contribution of acting intentionally that allows one to know without observation what one is doing and why. 'Agent's knowledge' is thus closer to what Kevin Falvey talks about as 'knowledge in intention' (2000, p. 41n2).

understanding of what these claims amount to and how the intimate connection is best characterized: namely, as a necessity-claim of the relationship between acting intentionally and acting knowingly. I shall then proceed to qualify these principles (§3) before discussing the epistemic distinctiveness of agent's knowledge (§4). Finally, I shall argue (§5) that if the relationship between agent's knowledge and acting intentionally really is a necessary relation, agent's knowledge will bear on action theory by functioning as its criterion of adequacy and rule in favour of what I shall call *Anscombean Theories* of action – or accounts according to which acting intentionally is, at least in part, an epistemic endeavour.

§2 AGENT'S KNOWLEDGE and the PUZZLE of *INTENTION*

In *Intention* Anscombe suggests that we demarcate the class of actions that are intentional in virtue of the reason-requesting sense of the question 'Why did you do it?' or 'Why are you doing it?' According to her proposal, intentional actions are 'the actions to which a certain sense of the question "Why?" is given application', the relevant sense of the question being 'that in which the answer, if positive, gives a reason for acting' (1957, p. 9). What Anscombe is driving at in these sections is that there is an intimate connection between acting intentionally and acting for a reason. Anscombe claims that insofar as an agent acts for a reason she acts intentionally. She thereby signs up in favour of '[t]he connection between reasons and intentional actions', which, as Setiya says, 'at least in *this* direction, is one that everyone accepts' (2007b, p. 24).[6] As I argued in the previous chapter, the fact that some action is intentional – and hence the fact that Anscombe's question 'Why?' applies to the action – is bound up with the state of the agent. We reasonably expect the agent to have had reasons for doing what she did intentionally and, if she had, that the fact that she was Φ-ing for reason *R* – rather than, say, *Q* or *P* – is grounded in facts about the agent before and/or during her performance. Whatever these facts turn out to be they make it true that *R* was the agent's reason for Φ-ing at t_1 and explain why the agent was doing what she was doing.

[6] One might supplement this claim with the claim that all actions that are done intentionally are done for a reason. Combine these claims and we get that actions are done intentionally if and only if they are done for a reason. Davidson defends this when he argues that acting intentionally just is acting for a reason (1963, p. 6). Anscombe seems to disagree when she leaves it open that one could act intentionally for 'no particular reason' since she says that the question 'Why?' is 'not refused application' when the answer is in the negative, no more than 'the question how much money I have in my pocket is refused application by the answer 'None'.' Perhaps 'I just felt like it', 'just thought I would', 'It was an impulse', done 'for no particular reason', or just 'an idle action' are ways to respond to the question 'Why?', so that these actions are intentional but done for *no reason* (1957, p. 25). For similar considerations, see Hursthouse's argument for the existence of *arational* actions (1991) but also Alvarez (2009). On Davidson's analysis such cases are done for no further reason, but they are still done for a certain kind of reason since there was a relevant belief-desire involved (i.e. the desire to simply Φ and the belief that Φ-ing is a way to Φ). However, one should keep in mind that Davidson talks about *primary reasons* here, which is his technical term for the relevant belief-desire pair that rationalized and caused the behaviour (1963, p. 4). Anscombe might not necessarily be read as claiming that there could be intentional actions without a primary reason. Agents who Φ because they 'felt like it' may have a primary reason, in Davidson's sense, without there being a reason for which they Φ (i.e. a consideration that prompts the Φ-ing).

If one thinks that what I defined in the previous chapter as *Davidson's Challenge* is adequate, one may think that action theory should investigate what features make it true that one acts for a reason. As Davidson puts it, we may then hope that there is 'an interesting principle at work' (1971, p. 43). Now an agent's reasons for acting will attach to her action through the operation of the agent's will – or in virtue of her practical thinking – since such processes allow her to take some *R* as her reason for Φ-ing (Setiya 2008b, p. 395).[7] To act for some reason *R* is thus a question of motivational and regulatory influence of *R* on one's behaviour. If there is a principle at work here – and I am not saying there need be – the task for action theorists will be to understand the psychological mechanisms in virtue of which an agent is appropriately guided by the reason for which she acts.

I shall now commence an investigation of a line of thinking in which this issue gets addressed by taking the phenomenon of agent's knowledge as its point of departure. The starting point is the contention that there is an intimate connection between agent's knowledge and intentional agency. As we shall see later – in chapters 3 and 4 – there are philosophers who take this starting point to be a promising way to address *Davidson's Challenge* since it allegedly motivates an account in which an agent necessarily enjoys a unified relation to her actions insofar as they are intentional and done for reasons. I agree that it is the right way to start; but, as we shall see, I disagree with what these philosophers think follows from this point of departure. I shall therefore use this chapter to discuss agent's knowledge as a starting point and what explanatory constraints and guidelines it gives to action theory. It will be instructive, however, to start out by discussing agent's knowledge as Anscombe describes it in *Intention* to provide some initial ideas of what it is and the role it plays in her framework.

Immediately after connecting acting intentionally with acting for reasons and the question 'Why?', Anscombe begins to demarcate the relevant sense of the question further by saying it is 'refused application by the answer: "I was not aware that I was doing that"' (1957, p. 9, 11). As Richard Moran points out, she does this because she 'needs a way of distinguishing the class of movements to which a special sense of the question "Why?" applies, but which doesn't itself employ the concepts of "being intentional" or "acting for a reason"' (2004, p. 43). Anscombe's problem is, as she says, that 'the question "What is the relevant sense of the question "Why?"" and "What is meant by "reason for acting"?" are one and the same' (1957, p. 9). Saying intentional actions are the class of actions to which a reason-request applies will therefore not demarcate the relevant class of actions in a non-

[7] Alternatively, similar processes may allow her to Φ intentionally on just a whim, or for no particular reason. See previous note.

circular way. The challenge is to characterize the relevant sense without presupposing one knows what intentional actions are, or which actions are liable to a request for reasons. Anscombe's project in *Intention* is to show that the special kind of knowledge an agent has of her own actions can be used to untie this conceptual knot. As she puts it:

> *What is required is to describe this class without using any notions like 'intended' or 'willed' or 'voluntary' and 'involuntary'. This can be done as follows: we first point out a particular class of things which are true of a man: namely the class of things which he* knows without observation.[8]

By means of this manoeuvre Anscombe connects non-observational knowledge (or 'awareness') of what one is doing with her investigation of what it is to act intentionally. Her claim is that intentional actions are a sub-class of things an agent can know 'without observation' and that this epistemic distinctiveness can help her to non-circularly characterize the relevant sense of the question 'Why?. This explains why she moves on to a careful investigation of what on the face of it seems to be another topic: namely agent's knowledge.

Anscombe's background for saying that the knowledge is non-observational is that an agent typically can say and know straight off what she is doing intentionally without needing to make observations of her own behaviour and her surroundings. Just consider the following:

> *Say I go over to the window and open it. Someone who hears me moving calls out: What are you doing making that noise? I reply 'Opening the window'. I have called such a statement knowledge all along; and precisely because in such a case what I say is true—I do open the window; and that means that the window is getting opened by the movements of the body out of whose mouth those words come. But I don't say the words like this: 'Let me see, what is this body bringing about? Ah yes! the opening of the window'. Or even like this 'Let me see, what are my movements bringing about? The opening of the window'.*[9]

The crucial idea here is that one does not seem to need an appeal to 'sense evidence', make observations about one's own behaviour, changes in the environment and so on in order to know what one is doing and why. 'A person can say what he's doing straight off—without observing himself', as Robert Teichman says (2008, p. 12). The fact that you are baking a cake because you want to surprise your partner seems to be enough to ground your knowledge that this is what you are doing. You need not look for clues to see whether there is a process

[8] Anscombe (1957, p. 13).
[9] Anscombe (1957, p. 51).

of baking going on where you are situated, or analyze patterns of your recent behaviour in order to see what comes next. The organisational whole that sorts and guides the range of different activities and movements that constitutes elements of your cake-baking is already ready at hand, as it were, insofar as you are the agent of this action and nothing is needed to show you that you are baking cake. In fact, agent's knowledge will often greatly exceed and go beyond such observational inputs. It might, for instance, not be obvious from observational evidence whether you are walking to the pub or the church if the shortest road to each place overlaps for quite a while, or you are taking some wrong turns, and so on.[10] Even in such cases it holds, as Velleman says, that '[y]ou don't find out what you're doing; in the normal case, you already know, because it was your idea, to begin with' (2007, p. 47).

In a word, the fact that you are doing something intentionally or for a reason is enough for you to know what you are doing. In particular, an agent's knowledge of what she is doing is in no need of support from what Anscombe refers to as 'separately describable sensations' (1957, p. 13).[11] We thus seem to be related to our own actions in a way reminiscent of how we have knowledge of our own minds. As O'Brien points out, it appears as if agent's knowledge is: (1) first person authoritative with respect to the nature of one's own actions (in a way one is not with respect to other's actions); (2) 'relatively a priori' in that it appears spontaneous, non-observational and immediate; and finally (3) relatively transparent in that these states cannot be 'beyond our ken' (2003, p. 359).

This is truly remarkable. An action – unlike the content of our mental states – at least appears to be an object of knowledge that is publicly available. 'I do what happens', is one of Anscombe's famous formulas, one she hesitates to assert in *Intention* while nevertheless endorsing the kind of thinking that gives rise to it (1957, p. 52). Agent's knowledge is thus knowledge of what happens – or empirical knowledge of the world – and so concerns something that extends beyond such features of us as are hidden from view. Unlike the contents of our own minds, it is an object that is 'out there', one which we are very good at identifying also from the third-person perspective; or, as Anscombe observes:

> *I am sitting in a chair writing, and anyone grown to the age of reason in the same world would know this as soon as he saw me, and in general it would be the first account of what I am doing; if this were something he arrived at with difficulty, and what he knew straight off were*

[10] A similar example is discussed by Falvey (2000).

[11] Anscombe stressed that the knowledge in question is without observation if it is not something one infers from 'separable' sensations whose content is other than the proposition inferred or the fact known. She illustrates the knowledge one has without observation with the knowledge one has of one's body based on *proprioception* (i.e. the sense of the relative position of neighbouring parts of the body): 'If only my leg had been bent, there would very likely just have been that fact and my knowledge of it, i.e. my capacity to describe my position straight off; no question of any appearance of the position to me, of any sensations which give me the position.' (Anscombe 1962, p. 73)

> *precisely how I was affecting the acoustic properties of the room (to me a very recondite piece of information), then communication between us would be rather severely impaired.*[12]

People found her formula 'I do what happens' paradoxical, Anscombe tells us, and she thinks the reason is a conflict between a widespread assumption – viz. 'what happens must be given by observation' – and her idea that agent's knowledge is something an agent 'knows without observation' (1957, pp. 13, 53). For now, however, let us just keep this remarkable feature of agent's knowledge in mind and temporarily bracket the epistemic distinctiveness of agent's knowledge – I will return to it below (§4) – to focus on a simpler claim: namely, that acting knowingly is an intimate feature of acting intentionally.

As Jennifer Hornsby says, it appears to be 'no accident' that agent's knowledge occurs when someone is acting intentionally (2005, p. 121). On the contrary, there seems to be a strong connection – perhaps even a necessary connection – between acting intentionally and having this peculiar kind of knowledge of what one is doing and why. Stuart Hampshire, who is an ally of Anscombe at this point, even suggests the connection is 'logical' between 'what a man knows or thinks that he is doing and what he intends to do' (1960, p. 95). The link between acting intentionally and knowing what you are doing thus appears to be a matter of necessity. In both directions, Hampshire's point goes as follows:

> *Necessity*: There is no question or possibility of the agent not knowing, since doing something with intention, or intentionally, entails knowing what one is doing.
>
> *Contrapositive*: If a man is doing something without knowing that he is doing it, then it must be true that he is not doing it intentionally. [13]

In light of this, it does not make sense to attribute an agent with intentionally annoying her partner when she is humming Beethoven's Ninth if she fails to register that so doing drives her partner crazy.[14] The agent is probably engaged in some intentional activity; but her intentional action cannot be correctly described as annoying her partner if this is a feature of her action she knows nothing about. Similarly, our agent could not be correctly described as humming Beethoven's Ninth in order to annoy her partner. She may have reasons for doing

[12] Anscombe (1957, p. 8).
[13] See Hampshire (1960, pp. 95 and 102). Anscombe makes similar claims (1957, p. 11) There is also an affinity to Aristotle in these claims since he thought that one needs to know 'particular circumstances of the action' when the action is voluntary (*NE* 1110b34). By characterizing this as a 'logical connection', Hampshire seems to think that the necessary connection is due to a conceptual relation between the concept of acting intentionally and the concept of acting knowingly. For the time being, I do not want to preclude whether the necessity-relation is conceptual, logical, metaphysical, nomological, or what have you. Later I shall defend and settle for the claim that the relation is one of metaphysical necessity.
[14] A similar example is discussed by Setiya (2007b, p. 25).

what she does – perhaps she is trying to memorize it – but given her ignorance of her partner's annoyance her reason for humming is not a reason for making a nuisance of herself. If she did it in order to annoy her partner, we would assume she knew she was so doing.

According to Anscombe and Hampshire, agent's knowledge is a necessary condition for acting intentionally. When we are Φ-ing intentionally, we must be Φ-ing knowingly. Insofar as an action-description unwittingly applies to what an agent does, this cannot be the correct description of what the agent did intentionally. Now this does not mean that they believe the agent must be consciously aware that one is Φ-ing. After all, I will still be writing a thesis even if this description should slip from my conscious mind during sleep. But even if what one does intentionally is not something one must be consciously aware of all the time, one is required to have agent's knowledge. I shall refer to this as *Anscombe's Dictum*:

> *Anscombe's Dictum:* If *A* is Φ-ing intentionally (or for reason *R*), then *A must* know in some special way that she is so Φ-ing.

As we shall see later, *Anscombe's Dictum* is a specific articulation of a broader idea that is shared by a range of views which I think we can categorize as *Anscombean*. Anscombeanism in this respect will be an umbrella term covering views conceiving of an intimate relationship between the nature of intentional agency and the necessity of bearing, in some form or another, a distinctively epistemic relationship to one's own action. The different versions of these views thus correspond with different ways of characterizing the nature of the required epistemic relationship. Anscombe and Hampshire characterized it in terms of (non-observational) knowledge. Other philosophers, such as Setiya, think knowledge is epistemically too ambitious and advice us to characterize the relation in terms of having a suitably qualified non-observational belief about one's action. I will return to this in chapter 3. For the time being let us for the sake of simplicity assume that the relationship is knowledge.

The main idea, as I have characterized it so far, is that the question of whether an agent knows she is Φ-ing is crucial for whether her Φ-ing is intentional. There are, however, two separable components to this idea of which we should bear in mind. Even though I will assume that knowledge of one's action includes them both, they are at least conceptually distinct. One concerns the nature of the kind of act one is doing, or what I will follow Goldman in calling its act-type or act-property (1970, p. 10). Knowledge of this feature of one's action corresponds to what Velleman calls *practical self-awareness*, the possession of which allows the agent to know or recognize some *prima facie* description that applies to her

activity (2007, p. 47). Part of what it is to have agent's knowledge, as *Anscombe's Dictum* talks about it, is to know whether the action one performs involves the act-type mowing one's lawn, reading a Saul Bellow novel, or, to use one of Anscombe's examples, replenishing the water supply (1957, p. 46). Following Velleman I shall call this *Practical Self-Awareness:*

> *Practical Self-Awareness* (*PSA*)*:* If *A* is Φ-ing intentionally (or for reason *R*), then necessarily *A* knows that what she does is a Φ-ing.

If Anscombeanism is on to something, *PSA* must clearly be satisfied when some agent acts intentionally. However, satisfying *PSA* is not enough to satisfy *Anscombe's Dictum*. In order to do that, the agent must also know her reasons for so performing this act-type, thereby being in possession of a sort of self-understanding of why she is doing it.[15] Again, an agent seems to be in this knowledgeable position in virtue of being the agent performing these actions. When you are baking a cake intentionally you know not only that this act-type – or some related description – applies to what you are doing; you also know whether you are doing it to surprise your partner, say, or because you want to try out one of grandma's old recipes. In a word, the following principle might be suggested as an addition to *Practical Self-Awareness*:

> *Self-Understanding* (*SU*)*:* If *A* is Φ-ing for reason *R*, then necessarily *A* knows that she is Φ-ing for reason *R*.

The idea is that when both *PSA* and *SU* hold *Anscombe's Dictum* holds – and vice versa – and that this is what it means for an agent to know her action. *PSA* requires that the agent knows what type of act she is performing. While *SU* requires that insofar as she is acting for a reason she must also know what her reason for so doing is.[16] I thus assume that knowledge of these components – the act-type and the agent's reason for performing the act-type in her situation

[15] Let me be more precise: what I claim here is that an agent acting for reasons knows what her reasons for acting are. This is not the same as the claim defended by Hyman, who argued that a fact cannot be *A*'s reason for Φ-ing unless *A* also knows that fact (1999, pp. 442–46). For all I have said, an agent's reasons for acting, *R*, may fail to be something she knows since *R* may not be true. Nor do I take a stand here on whether *A*'s reasons for acting must be facts. All I claim, is that *A*'s reasons for Φ-ing must be known by the agent as her reasons for Φ-ing. On my view, it is better to think of an agent's reasons for Φ-ing as *putative* facts. I also discussed this in chapter 1.

[16] Of course, an agent may perform the act-type of operating the water pump because she is replenishing the water supply and will presumably also be performing that act-type – i.e. replenishing the water supply – intentionally. According to *PSA*, she will then be required to know that she is doing both these things; however, *SU* will also require her to know that one of these things is her reason for doing the other. Furthermore, insofar as she has a reason for replenishing the water supply – say, because she wants to get water in her house, poison the people in the house by pumping poisonous water into it, and so on – *SU* requires that the agent also knows that this is her reason for replenishing the water supply. In this respect, there may be a chain of 'Why' questions pertaining to each thing one does intentionally that are answerable by a corresponding chain of reasons for acting. There might be, as Anscombe says, several descriptions 'each related to the next as a description of means to end' (1957, p. 46). *SU* then requires that the agent knows the items of the latter chain whereas *PSA* requires that the agent knows that she really is doing what the 'Why?' question is asking her to justify or provide a reason for. Together, both principles will then impose knowledge of a certain rational structure of how the different acts relate to one another. Thanks are owed to Jennifer Hornsby for reminding me of some of these additional complexities.

(its end) – amounts to agent's knowledge. By so doing, I take it for granted that an agent's intentional action is to perform this act, Φ-ing, for the sake of some end or reason, *R*.[17]

At this point some philosophers will perhaps object because certain actions seem to be intentional without being performed for a reason. They will probably agree that insofar as someone is acting for a reason she is acting intentionally and that when someone so acts she must know that *R* is her reason for Φ-ing. What is not so clear, they would say, is whether acting intentionally must involve acting for a reason. Are there not instances of intentional actions that are done for no reason, for the sake of the act itself, or just for the heck of it? In other words, some philosophers will point out that there is a group of intentional actions done for no reason that *SU/PSA* fail to capture. Now insofar as we allow intentional actions that are done for, literally speaking, no reason, this must be reflected in principles characterizing an agents' self-understanding. In particular, it should allow situations where an agent's self-understanding consists in knowing there is no reason for which they are Φ-ing:

> *Supplementary SU:* If *A* is Φ-ing intentionally for *no* (particular) reason, then necessarily *A* knows that she is Φ-ing for no (particular) reason.

A question is whether such actions could still be said to be done for the sake of something, or with a view to some end. I shall be brief about this since I think the most plausible rendering of the view that agents may act for no reason is that they may act for *no particular* reason. One could say that an agent in this situation may Φ for the sake of the Φ-ing itself. Someone may dance, for instance, just for the sake of so doing. If that does not work – or if it fails to cover the range of cases we perform for no reason – we could perhaps allow that intentional actions can also be acts performed for the sake of no end. The important point at this stage, however, is to give a characterization of agent's knowledge and the self-understanding an agent must have of her own action insofar as it is intentional, and I think *PSA*, *SU* and, if need be, *Supplementary SU*, work as preliminary principles.

In the next section I will show that we need to qualify these preliminary principles in light of certain alleged counterexamples to the Anscombean view. Before we get that far, however, let me end this section discussing some important features of the view. First, the relationship between agent's knowledge and the description under which an agent knows her action. Next, the progressive aspect of intentional actions, which I hope to illuminate. Finally, I clarify the extent to which agent's knowledge must be conscious.

[17] For a similar conception of action as an 'act done for some end', see Korsgaard (2009, pp. 11–2).

As Anscombe points out on numerous occasions, the event that is one's action can be captured by more than one description. We may describe an agent's action as 'sawing a plank', 'sawing oak', 'sawing one of Smith's planks', 'making a squeaky noise with the saw', or 'making a great deal of sawdust'. Now an agent may, as Anscombe points out, 'know that he is sawing a plank, but not that he is sawing an oak plank or Smith's plank' (1957, p. 12). That is a general point: agents may know that her activities fall under one or some of the true descriptions, $D_1,D_2,...,D_n$ all of which apply to her action without thereby knowing that it falls under all or most of $D_1,D_2,...,D_n$. According to Anscombe, this is because agent's knowledge is sensitive to the description under which she knows her action. To stick with our previous example: if an agent is humming Beethoven's Ninth without knowing that her action can be correctly described as annoying her partner, the latter description cannot be the description in our report of the agent's knowledge of her action. Consequently, this description is not an adequate description of what she is doing intentionally either. As Anscombe says, 'to say that a man knows he is doing X is to give a description of what he is doing under which he knows it' (1957, p. 12). In this respect, the description of some *A's* action as intentionally Φ-ing is sensitive to whether or not one may truly report *A*'s knowledge in action as a Φ-ing; and that in turn depends on whether *A* knows her action under this description or not.

Now it is important not to confuse this claim – viz. the claim that Φ-ing is an intentional action only if its agent knows that her action is a Φ-ing – with the similar-looking idea that the agent must be able to come up with a description of what she is doing as Φ-ing. Of course, I do agree with Falvey and others in thinking that

> *... a fundamental datum that under normal circumstances, an agent in the midst of acting can typically provide a correct answer to the question, 'what are you doing?'* [and 'why are you doing it?'] *straight off, without relying on observation or inference.*[18]

However, emphasizing an agent's knowledge in acting intentionally is not the same as saying that someone who honestly tells you what she is doing must thereby be assumed to give you the right account of what she is actually doing and for what reasons. Knowing what you are doing is not the same as knowing how to best describe your action in words. Agents may, after all, make honest mistakes. As Hampshire points out, '[w]hat I actually intend to do is not necessarily the same as I would honestly say that I intend to do, if I were asked' (1960, p. 95). In some cases my intention or action is not what I honestly declare it to be; but this is nothing

[18] Falvey (2000, pp. 22–3).

more than a mistake in characterizing it in words. In the absence of such an ability to characterize one's action it is still true that I know what I am doing and why in the sense that I did 'something with a clear intention to bring about a certain result' (Hampshire 1960, p. 96).

The second feature we need to notice concerns the events and processes we describe as actions: namely, the *progressive aspect* of the descriptions. When something is described in the progressive – e.g. 'I am writing a thesis' or, in the past tense, 'I was writing a thesis' – this may hold true even though the thing one is or was doing never comes to pass. This is due to the openness of the progressive aspect that allows the truth-conditions of 'I am writing a thesis' or 'I was writing a thesis' to be satisfied even though 'I wrote a thesis' will never be true. Who knows, I may be financially constrained, struck a deadly blow tomorrow, get bored, or encounter some unsolvable philosophical problem in the next few sections.[19] Accordingly, the knowledge one has of what one is doing can often be of the fact that one is/was writing a thesis without thereby implicating knowledge of the fact that one has written it (or ever will).[20] A related feature is the *broadness* of the progressive that, simply put, allows some slack in what one must be doing in order to count as doing something. I count as writing a thesis *these days* even if I spend a weekend skiing; and I count as writing a thesis *now* even if I spend a significant amount of my time staring out the window, 'pacing around [my] office tearing out [my] hair in frustration over how to put an important point' (Falvey 2000, p. 22).

There are, of course, some restrictions on both the openness and the broadness of the progressive. At least, this seems to be true for *telic* actions, or actions – such as 'I am writing this thesis' – that have an ending point built into them.[21] It is, for instance, dubious whether I can truly say 'I'm solving the philosophical miracle *par excellence*' if I am fully aware of my limits as a philosopher (or, alternatively, the limits of philosophy). On the other hand, it seems correct to allow Marcel Proust, in replying from his sickbed, to say 'I am writing *À la recherche du temps perdu*' even though he knows that it is impossible for him to finish his masterpiece. Perhaps, then, as Falvey points out, all that is required is for it to be possible to complete the Φ-ing though not necessarily by the agent to which the progressive description applies (2000, p. 42n6). Similarly, I am allowed to not write all of the time to be writing a thesis now, though it probably wouldn't count as so doing if I spent all my time watching old

[19] As Zoltan G. Szabó (2004, pp. 29–30) argues, an important difference between past tense sentences that share other syntactic and semantic features may be in their aspectual difference. The perfective aspect of 'Mary crossed the street' may be a false claim even though its progressive aspect-correlate, 'Mary was crossing the street', is false. The members of this pair relate semantically to one another in that the progressive describes a crossing in progress whereas the former describes the crossing as a completed process.

[20] Also see Thompson (2011).

[21] In contrast, *atelic* actions, such as 'I'm walking' or 'I'm eating', have no such endemic endpoints built into them and can go on forever, at least in principle, or be broken off at any point without threatening the truth of the description. It also follows from atelic descriptions that '*A* is/was Φ-ing' that '*A* Φ-ed/has Φ-ed' (e.g. if Sue is walking it follows that she walked or has walked). The latter is not true for telic descriptions, such as 'I am writing a thesis'. See Bernard Comrie (1976). My debt for these points goes to Sarah K. Paul (2009a, p. 16).

Seinfeld episodes. I do not have a way to systematically demarcate the truth-conditions for such progressives. What we should take note of, however, is that their openness and broadness will have a bearing on what kind of knowledge one has. Knowledge, just like the action itself, must be allowed some slack as to what counts as knowing what one is doing. It is thus no counterexample to the claim that I must agent-know the fact that I'm writing my thesis that I am spending much of that time not writing.

A final thing worth mentioning is that there is no need to associate agent's knowledge with being consciously aware of these facts. As Setiya points out, 'I may believe that I am connecting an electrical circuit and do so intentionally, as I turn on the light, without consciously reflecting on that fact' (2007b, p. 25). Agent's knowledge thus appears to be akin to the awareness one has of one's mother's name, rather than the demanding sense of consciousness one may have by tuning in one's attention. Although writing a thesis requires a certain level of reflection, it is not required that I always have this activity at the forefront of my mind for it to be true that I am doing this intentionally. I am, for instance, not consciously aware of it when I sleep and I may be completely absorbed by gossip during lunch-hour. Even if I am not consciously attending to my project it is nevertheless true that I am writing a thesis

Anscombeans will, of course, be asked for reasons for thinking that agent's knowledge – as we have described it so far – pertains to intentional agency as a matter of necessity. There is after all, as Setiya points out, 'a profound conflict in action theory between those who follow Anscombe in seeing a close connection between intentional action and knowledge of what one is doing and those who do not' (Setiya 2009a, p. 131; 2011). Below, I give voice to two lines of thinking that provide *prima facie* evidence in favour of the intimate connection: a qualification of the above principles in the face of certain sceptical counterarguments will render the principles more plausible (§3). I will then provide a positive argument that takes advantage of the distinctive epistemic relationship one has towards the nature of one's actions (§4). Given the provisory nature of these considerations I will not be able to fully home in the necessity-claim; however, that is as it should be. Only an account of agent's knowledge will put us in position to see whether this relation is a necessity of acting intentionally.

§3 KNOWLEDGE and NECESSITY

The aim of the following section is to qualify the idea that agent's knowledge is necessary insofar as someone is acting intentionally. I do so by facing certain lines of scepticism. There are two main strategies in the literature against the Anscombean idea that

acting intentionally necessarily involves acting knowingly. One line of argument goes by way of counterexamples. I shall address such cases by qualifying the claim to which Anscombeans should commit. The other line of argument tries to dismiss agent's knowledge by calling on experimental findings in cognitive science and social psychology that allegedly prove that an agent's perception of what she is doing and why is more or less illusory. I will dismiss such radical conclusions; but let me begin with the counterexamples.

According to Davidson, we have reason to doubt Anscombe's claim that for any Φ-ing where the Φ-ing is intentional its agent must know that she is so Φ-ing. To prove this he describes cases where allegedly an agent is Φ-ing intentionally without knowing whether she is. The case of the man making ten carbon copies is a case in mind, since he 'may be making ten carbon copies as he writes, and this may be intentional; yet he may not know that he is; all he knows is that he is trying' (Davidson 1971, p. 50). Similarly, a father may provide for the welfare of his children by adding a clause to his will in order to make sure his children are in the custody of people he believes will protect their best interest. The father may have grave doubts as to whether this really will provide for the welfare of his children, however, and remain so until his death. Nevertheless, if the children are taken good care of after his death as a result of his will, it seems right to say that he provided for their welfare and that he did so intentionally despite his failure to know that he was so doing (Davidson 1978, pp. 91–2).

Correctly in my view, Setiya takes cases like these to show that we need to qualify the Anscombean claim. Davidson concludes his own discussion by saying that it is still true that what an agent does intentionally must be 'known to him under some description' (1971, p. 50). In particular, he suggests the description of the agent 'trying to Φ' applies even in the case of the carbon copier, or the man providing for the welfare of his children. In this respect, the Anscombean principles could be re-written as the necessary connection between agent's knowledge and trying to do something. Several philosophers have been tempted by this line of argument. Hampshire, for instance, may say that this was something he had in mind all along when he says that 'there is a sense in which he unfailingly knows what he is trying to do, in contrast with an observer, simply because it is his intention and not anyone else's' (1960, p. 102). Other philosophers have accepted that agent's knowledge pertains of necessity to one's intentional actions and then – moved by analogous epistemic anxieties which Davidson takes advantage of in his examples – sought to define the nature of what one knows (viz. the action) by identifying it with a psychological kind, such as a trying or a willing, of which it is true to

say that one necessarily knows whenever one enacts it (O'Shaughnessy 2003, p. 346).[22] Yet another model is to think that there are dual components in the complex we recognize as the outer action. In particular, that there is a psychological kind of doing (viz. a trying) that is a component of an ontologically more complex doing (viz. the publicly available action). One may then claim that one must have agent's knowledge of the former but not necessarily of the latter. Keith Donnellan, for instance, suggests that 'our knowledge of our own intentional actions is complex [in] that it divides up, so to speak, into an element of 'direct awareness' (...) and other elements to which observation is relevant' (1963, p. 407).

As several philosophers have pointed out, however, these retractions are typically motivated by epistemic anxieties concerning the knowledge that is involved in intentional action and whether it involves knowledge of the world beyond one's inner self or the boundaries of one's body.[23] Although she on occasion came to obscure this feature of her view,[24] Anscombe warned us about making these retractions when she said that it is

> *... an error to try to push what is known back and back; first to bodily movements, then perhaps to the contraction of the muscles, then to the attempt to do the thing, which comes right at the beginning.*[25]

On her view, '[t]he only description that I clearly know of what I am doing may be of something that is at distance from me' since by allowing this one may in turn leave room for the object of such knowledge to be of what one is doing, or of 'what happens' (1957, pp. 52–3).[26] This is not an argument, of course, and we might be forced to make these retractions in the face of genuine worries and counterexamples. One thing that is problematic with such retractions, however, is that they leave critical parts of what we take ourselves to know in action to be a matter of perceptual or observational knowledge. We know non-observationally and spontaneously what our actions-as-tryings are; however, to see what the full action is – which depends on the effects of one's trying – observation would seem required. According to O'Brien, this does not square well with the first-person/third-person asymmetry and the

[22] O'Shaughnessy (2008a; 2008b) develops and defends the view that the act is identical with an agent's attempt or trying. Hornsby (1980) develops a similar view where the event of one's trying to do something is identified as the agent's action.
[23] According to Falvey, this is the so-called 'two-factor thesis' (2000, p. 21). O'Brien discusses a similar theory under the heading a 'Dual Component Model' (2003, pp. 366–70; 2007, pp. 171–6). Also see Setiya (2007b, p. 27n9) and McDowell (2010, p. 431).
[24] See McDowell's discussion of Anscombe's chalk-writing example (2010, pp. 429–31).
[25] Anscombe (1957, p. 53).
[26] McDowell has argued on more than one occasion that such epistemic anxieties are distinctively modern and reflect our questionable questions that, much like Descartes in his *Meditations*, can put us on an inward retreat. Of course, one hopes to regroup in some terrain where we are epistemically more at home so as to proceed from there and secure a proper base for the study of knowledge and agency; but, as McDowell adds, '[a]nyone who knows the dreary history of epistemology knows that this hope is rather faint' (1995a, p. 396). Anscombe's warning is that the study of action will be just as dreary if we start on an analogous inward retreat as a response to what McDowell calls 'a *highest common factor* conception of knowledge in intention' (2010, p. 431).

relative *apriority* of agent's knowledge (2003, p. 367). In addition, it would make an agent's knowledge of the full action no more transparent than the relative transparency that perceptual capacities leave room for (2003, p. 367). Indeed, as Adrian Haddock points out, if one insists that all the activity there is in intentional agency is the trying, one risks the odd view that one's bodily movements – which are mere effects of one's body – are just as much at a distance from the agent as the movement one produces in a water pump by handling the pump with one's hands (2005, p. 161). In other words, the view risks having the consequence 'that I stand to my body as a captain to his ship', as Watson says (1982, p. 467).[27]

We should also notice that even if we do embark on such a retraction the main puzzle will remain. It will, after all, still be a question of why an agent must have knowledge of these tryings. As Setiya says, '[t]he real puzzle is about the need for *cognition* [i.e. agent's knowledge] of one's action, in acting for reasons, however mediated it may be' (2007b, p. 27 [original emphasis]). Retraction is thus no solution to our epistemic problem. Setiya thus thinks that we should resist these epistemic anxieties, and ask instead whether we could have a more precise formulation of what one must know in Φ-ing intentionally. His suggestion is that an agent who Φ-es intentionally while failing to know that he is so Φ-ing must at least do something else intentionally with a view to be Φ-ing. That sounds more promising, I think. As Thompson has pointed out, many intentional actions are 'resoluble into a heterogeneous collection of sub-actions that are themselves clearly intentional' (2008, p. 106). Egg-breaking and egg-mixing relate to omelette-making in this way. They are sub-actions of the omelette-making and relate, as Thompson metaphorically puts it, as 'organs (...) [to] the whole' (2008, p. 106).[28] The point is that the man making ten carbon copies is not merely trying to make ten carbon copies; rather, he is performing many things that contribute to the whole and are proper parts of the intentional action of making ten carbon copies. The carbon printer is pressing his pen hard on the paper, writing plain letters, and so on. Similar things could be said of the father who adds a clause to his will: he is perhaps conferring with his lawyer, making sure the will is in accordance with the laws of his country, or he takes care to double-check his judgement of the people to who he gives custody, and so on.

[27] The image is invoked by René Descartes in his *Meditations on First Philosophy* where he argues that 'nature also teaches that I am present to my body not merely in the way a sailor is present in a ship, but that I am most tightly joined and, so to speak, commingled with it, so much so that I and the body constitute one single thing' (1641, meditation six (p. 51)). I am referring to R. Ariew and E. Watkins 1998 edition.

[28] There is an interesting question what makes something a proper part of the intentional action. When I make an omelette or ten carbon copies there is clearly a movement of molecules in my hand, and even though the movement of the molecules is part of the bodily movement that *is* a proper part of my action the molecule movement is intuitively no part of the intentional action that takes place. The parthood-relation here thus has a special sense. As Thompson points out, this intuitive sense of parthood is tied up with – or at least not independent of – 'the connection expressed in rationalizations generally' (2008, p. 107n2). If something is a part of an intentional action – say, the firing of a neuron in my left hand – but not a part of a rationalization or reason-giving explanation, it is not a proper part of the intentional action.

Sub-actions are things done intentionally with an aim to be Φ-ing. They can now be employed to patch up an adequate formulation of the necessity of agent's knowledge. We can say that one must at least know that there is something one is doing intentionally with an aim to be Φ-ing where this something is more than just trying to Φ (Setiya 2007b, p. 26). We may thus qualify *Anscombe's Dictum* and let it read as follows:

> *Agents' Knowledge* (AK)*:* If *A* is Φ-ing intentionally (or for reason *R*), then *A must* either be Φ-ing while knowing that she is so Φ-ing, or else she is so Φ-ing by virtue of doing other things – which is not merely trying to Φ – for which this knowledge-condition holds.[29]

AK can accommodate alleged counter-examples provided that the action where the knowledge of what one is doing intentionally can be disputed is not a basic (or primitive) intentional action.[30] As Setiya says, '[a] consequence (...) is that basic intentional actions – ones that are not done by doing something else intentionally – must be accompanied by [knowledge]' (2008b, p. 390). After all, if there is nothing else one is doing in order to be Φ-ing intentionally failing to know that one is so Φ-ing would be a counterexample even to *AK*. That will allow *AK* to accommodate the carbon copier case and the father providing for his children since their intentional actions – for which they lack knowledge – are not basic in this sense. If, however, one could find a case of intentional Φ-ing where (a) the Φ-ing is a basic intentional action; and (b) one could intentionally Φ without knowing that one is so Φ-ing, we would have a counterexample to *AK*. I return to such alleged counterexamples in chapter 3 below where I discuss in particular cases raised by Setiya when he argues in favour of Anscombean principles that characterize agent's knowledge in terms of suitably qualified beliefs rather than knowledge. For now, however, let us stick with *AK*.

A more recent line of scepticism takes advantage of the growing psychological literature and aims to demonstrate that – contrary to the Anscombean view – agents know very little about what they are doing and why they are doing these things. Allegedly, the claim that our agency is a rather benighted endeavour finds support in experiments such as those surveyed by Richard Nisbett and Lee Ross in *Human Inference: Strategies and Shortcomings*

[29] *AK* reflects the structure of principles that Setiya discusses which take an agent's necessary epistemic relation to one's actions to be suitably qualified beliefs. I return to this in chapter 3 below. See Setiya (2007b, pp. 26, 41; 2008b, p. 390).

[30] The notion of a *basic intentional* action is taken from Arthur Danto's definition of *basic* actions, which are actions that require no other action in order to be performed. A basic action is an action we can directly do, as it were (Danto 1963; 1965; 1973, ch. 2). Davidson operates with a similar notion when he talks about *primitive* actions as actions that 'cannot be analysed in terms of their causal relations to other acts of the agent' (Davidson 1971, p. 49). Here I am not making the claim that there exist such basic actions; rather, I rely on the idea that a counterexample would be successful only if there are suitable cases that involve basic intentional actions. It is therefore these counterexamples that depend on the existence of basic intentional actions.

of Social Judgement (1980).[31] Here, experiments reveal that people's explanations about themselves are often mistaken. Is this in conflict with the Anscombeans' insistence that agents must know the explanations for why they are doing something when they do it intentionally?

In my view, there are two reasons for not discarding the Anscombean insight despite these findings. First of all, the experiments surveyed by Nisbett and Ross focused on people's failure to explain moods, judgements, and preferences, and not so much on action and decision-making. As Velleman has pointed out, the few experiments that test reflective explanations of actions are ones in which the experimenters must 'go to unusual lengths to confuse subjects about the true causes of their behavior' (2007, p. 17n4). Even Nisbett and Ross concede that these findings 'say little about the degree of accuracy to be expected under less stage-managed conditions' (1980, p. 206). Of course, the Anscombean view is formulated as a necessity claim so even just a few problematic cases of agents acting intentionally under thorough ignorance would be enough to discard the insight. Having said that, the massive interactions of the researches may throw doubt on whether the agents they observed were acting intentionally. The subjects may be so confused by the scientific set-up that they think they are acting for reasons when they really are not. A more fundamental point is that what Nisbett and Ross are challenging with their result is the scientific adequacy of the agent's explanation of her own actions and not so much whether the subjects have commonsense knowledge of what they are doing and why when they act intentionally. The failure of agent's knowledge to meet scientific standards is not necessarily the failure of having knowledge in action. Anscombe, for one, never claimed that agents can explain why they did something in a way that meets scientific standards of experimental psychology. What she claimed was that they can explain it in a way sufficiently adequate for commonsensical explanations of one's own behaviour.[32] I therefore conclude that there is reason to be sceptical about the radical conclusions some may draw from such experimental findings and results.

§4 The WAY AGENTS KNOW their ACTIONS

In human agency not all knowledge is of the relevant kind. Sometimes an agent's knowledge in action does not signify intentional agency. I may, for instance, know that I am clacking my teeth together because I am having a dramatic seizure. I may even know that the seizure is the reason my teeth are clacking. Yet these movements are not intentional or performed for a reason. In fact, the nature of the knowledge in question and the way it is

[31] Also see Nisbett and Timothy Wilson (1977).
[32] For similar replies, see Velleman (2007, p. 17n4).

achieved may be a sure sign that the behaviour in question is not one's own doing. In particular, if the knowledge amounts to spectatorship, what one does will get mixed with a degree of externality so that even though the behaviour comes from within the agent's body – and even from within her own psychology – the agent is somewhat detached from it. As Anscombe points out, the question 'Why?' does not apply to my action if an answer to it is that 'I *observed* that I was doing that' (1957, p. 25 [original emphasis]); a need for observation is thus a sure sign that the action is not intentional. What I shall discuss in this section is therefore the idea that the nature of the knowledge an agent has in action must be of a particular kind and that insofar as it fails to be of this kind there will be an element of undergoing at the expense of actively doing in the action at hand. In a word, I shall argue that the way we know our actions is crucial for whether or not the thing known is an intentional action. My aim is to get a clearer understanding of the epistemic distinctiveness of agent's knowledge and my strategy for so doing is to start out from a certain pattern in a sound practice surrounding self-ascriptions of intentions and intentional actions.

Let us begin with Velleman's observation that recasting an agent as a spectator to her own behaviour is a way to show the agent lacks the control we expect of an intentional agent:

> *Imagine that your arm becomes temporarily paralyzed. When you wake up each morning, the first thing you do is to check whether you have regained control of your arm. What exactly are you hoping to find? Part of what you're hoping to find, no doubt, is that your arm moves. But movement by itself wouldn't be enough. Waking up to find you arm flapping around aimlessly wouldn't lead you to think that your control over it had been restored. You'd have to conclude instead that paralysis had given way to a spasm.*[33]

The lesson Velleman wants to draw from this case is that an agent may be detached from her actions not only because she fails to know what she does, but also by certain ways of knowing it. We may, for instance, know that we sneeze merely by virtue of being the one who sneezes; but that does not make our sneezing a case of agency, or something we control. What Velleman takes advantage of in the above description is thus a principle shared by many philosophers, namely that 'being a spectator is the diametrical opposite of being an agent' (2007, p. xii). And from the viewpoint of agent phenomenology this makes much sense. O'Shaughnessy, for instance, argues that '[c]ommon to all experiences of loss of agency is the sense of becoming a spectator of one's own actions' and, conversely, that '[c]ommon to all experiences of recovery is the sense of one's actions becoming so close to one (...) that it is

[33] Velleman (2007, pp. xi–xii).

impossible for one to remain any longer a spectator of them' (1963, p. 391). As soon as a knowledgeable agent is portrayed as a mere spectator to the events comprising her behaviour our impression, it will erode our belief that this is a case of genuine agency and we will begin to see the agent as being somewhat detached from the event. It is, of course, no one else but the spastic person whose arm is flapping around aimlessly. She may even exercise some control by grabbing it with her other arm, and so on. Even so, the point remains that the flapping about is a happening in the agent's life in contrast to the things she does actively and that a clue here is the way in which she knows what she is doing.

To get somewhat closer to the way in which one should know one's action insofar as it is intentional, let me begin by considering the observation that there are certain limits to how one may dispute an agent's knowledge of what she is doing without hereby challenging the action's status as intentional. As Velleman notes, 'if someone responds to "I am going to take a walk" with "How can you tell?", he has failed to recognize it as an expression of intention rather than belief' (2007, p. xiii). Similarly it is hard – if not straight-out impossible – to accuse someone acting intentionally 'of simply not knowing what he is doing, of sheer ignorance in this respect, without implying that his action is not intentional' (Hampshire 1960, p. 95). This point is strengthened by the fact that (1) below may be a felicitous utterance if it is taken as an assertion expressing one's belief whereas the same utterance will sound remarkably odd if it is meant to be taken as an expression of one's intention:

(1) I am failing my exam, I just know I am.[34]

Now both speech acts may be true and they seem to convey content in the indicative (descriptive, informatory) mood and so provide descriptions of what one is in fact doing. Expressing one's intention by saying 'I am getting myself a coffee' is in this respect quite similar to the way one would express one's belief by saying 'I am going to be sick'. In a word, they share the indicative character or form, something that is natural if both statements can be used to express 'knowledge on the part of the speaker' (Velleman 2007, p. xiv). The vehicle for voicing an intention must then be able to convey what one knows. As a consequence, the indicative form is 'not the distinctive mark of "predictions" *as opposed to* "expressions of intention",' as Anscombe points out (1957, p. 3).[35]

[34] The example is Anscombe's (1957, p. 2). Hampshire makes a similar observation when he notes '[t]o say "I know now what I intend to do" is a redundant way of saying "I know now what I shall do", and "I know what my intention is in doing this" is an impossibly redundant way of saying "I am doing this with intention or intentionally"' (1960, p. 102).

[35] This fact also counts in favour of taking the content of an intention in action to be given by sentences like 'I am Φ-ing' rather than the future-oriented sense as exemplified by 'I shall Φ', or all-out judgements such as 'I favour Φ-ing', since then the content of one's intention in

Another similarity between the above speech acts is that by adding the clause 'I just know I am' is a way to render the respective utterances – viz. expressing one's belief or one's intention – conversationally inappropriate. However, there the similarity stops. For whereas (1) as an expression of belief is inappropriate in the way that Paul Grice defines it so as to generate conversational implicatures, (1) as an expression of intention appears to convey nothing but confusion.[36] When you assert '*P*' and add 'I just know *P*' you add what appears to be superfluous information. Your interpreter will ask herself why you add this information and may reasonably entertain the hypothesis that perhaps you emphasized your knowledge in order to convey your epistemic distance to the fact. Adding this clause to a description of some state of affairs in the indicative mood will therefore most likely settle whether your interpreter takes it as conveying a belief or an intention. The distance and your need for assurance would make your interpreter think you are conveying your belief. It would probably raise some eyebrows if you then tried to correct your interpreter by saying that you were just emphasizing that you are in a position to know for a fact that you are going to fail the exam intentionally. If you really are bent on failing the exam, such assurance is absolutely superfluous. By adding it, your interpreter would thus read your assertion as not conveying your intention. The clause functions as a disclaimer of any formed intention. This difference is systematic and holds for a range of cases where the content of one's belief might just as well be used to express one's intention or report what one is doing intentionally. Adding such a claim to know is so odd that it challenges the utterance as an expression of an intention. On the other hand, adding this clause does not challenge the speech act of asserting what one believes since one then reads the indicative content in (1) as something the speaker considers as something that will happen to her rather than something she will do. Since the only difference is in the kind of speech act that the speaker has attached to the indicative content it suggests that the difference originates from the way one relates to this content as respectively

---

action is just the content of one's knowledge of what one is doing when one is executing it. This is part of the background for McDowell's argument (2010) against the *Volitional View* of intention and *Davidson's View* that the content of an intention is an all-out judgement. According to McDowell, Sellars (1966, p. 110) and Brandom (1994, pp. 256–9) defend the view that intentions in action are best understood as *volitions*, which started out as intentions for the future ('I will Φ', 'I shall Φ', etcetera) and that they mature into intentions in action when the time comes to act ('I shall Φ *now*'). One problem is, as McDowell points out, that some intentions in action are formed on the spur of the moment and thus are never future-oriented. (Compare Bratman's criticism of 'the Simple View' (1999b, p. 113)). Another problem is that 'now' only signals the *onset* of an intention in action, and thus the 'now' cannot detach futurity from Sellars' way of characterizing the content of an intention. As Falvey puts it, 'the future sense still lingers in "I shall raise my arm now"' (2000, p. 41n5). So we fail to have a characterization of the content that is one's intention in action during the extended period in which that intention works (McDowell 2010, p. 416). The problem with Davidson's view that the content of an intention in action is an all-out judgement is that it fails to recognize that practical reasoning is not reasoning towards the truth of a proposition. McDowell takes these failures to support the view that intentions in action are best expressed by saying things like 'I am Φ-ing'; or, in other words, that the content of one's intentions is the same as what one knows one is doing insofar as one Φ-es intentionally. Since that argument relies on considerations of the peculiarity of agent's knowledge and practical reasoning it will not help us to establish that peculiarity; but it is worth having in mind that these positions fit so neatly together.

[36] Conversational inappropriateness is defined as flouting certain conversational maxims governing conversation. One may utilize such inappropriateness to allow one's conversational partner to derive an implicature of what one is saying which reflects the speaker's intended meaning (viz. what *A* meant by saying *P* when so saying obviously flouts maxim *X*) (Grice 1991, chs. 1, 2, 3 and 5).

a believer and an intentional agent. That is, it suggests that believers and agents relate in entirely different ways to the true descriptions of their actions.[37]

A similar conclusion falls out when we consider that voicing our ignorance about some relevant indicative content that describes that one is or will be doing something is a way to provide what Hampshire calls 'a disclaimer, honest or dishonest, of any formed intention' (1960, pp. 102–3). Intentions cannot be in this way unknown to their agent. Suppose you know I am considering failing the exam intentionally. If I then tell you that I do not know whether I will fail the exam tomorrow, I will be conveying to you that I have not settled the matter or formed the intention yet. Voicing one's ignorance with respect to a belief, however, does not have the effect of disclaiming the fact that one still believes it. I may believe despite acknowledging my ignorance, saying 'Of course, I do not know whether I will fail my exam tomorrow, but I nevertheless believe I will'. Again, this bespeaks a different relationship between a believer and an intentional agent with respect to the relevant indicative content.

The sense that there might be a distinctive and special epistemology operating in intentional agency is strengthened when we observe that the indicative content of (1) can be challenged by asking 'How can you tell?' if it is used to express one's belief while this challenge would be odd if raised against an expression of intention. Asking someone who honestly tells you that she is taking steps to sue her former employer because she she was treated unfairly with 'How can you tell?' would be quite out of the ordinary. She would probably say 'What do you mean "How can I tell?" I just told that I'm going to and that settles it!' Challenging an utterance in this way thus seems to be at odds with recognizing the indicative sentence as expressing an intention. What one thereby challenges is its status as an expression of intent or as a report of one's intentional action. One does not challenge someone's expression of a belief by doing that; one merely challenges its status as justified.

What these conversational patterns reveal is first and foremost a practice of treating expressions of one's intentions and self-ascribing actions/intentions that differs from our practice for dealing with ordinary claims to know or believe something. In fact, expressing intentions seems to be similar to our practice with respect to self-ascriptions about one's own mind. Self-ascribing actions/intensions are marked by action-theoretic versions of what

[37] Since knowledge may be the norm of assertion, the fact that these utterances assert some *p* and then add that they know *p* might be a reason to think that there is some redundancy expressed by all of these utterances. However, the fact that expressing an intention and then adding that one knows appears more redundant – or weirder - than voicing a belief one doubts cannot be explained by violation of the norm of assertion and so an explanation is called for. Anscombeans will insist that we explain this by pointing to the kind of connection there is between knowing and intending. For discussion and defence of the knowledge norm of assertion, see Williamson (2000, ch. 11).

One could convey something non-redundant by saying 'I (now) know that I'm intentionally Φ-ing (or Φ-ing for the reason *R*)'. However, by so saying one will not convey that one has been supplied with knowledge of one's own action. Rather, one would convey that one now knows that 'such-and-such a form of words gives the correct description of what I am going to try to do, or of what I am now actually trying to do' (1960, p. 102). I take it that in the case where (1) is used to express an intention this reading is not available.

Elizabeth Fricker sums up as the *Minimal Phenomenon* (1998, p. 157). Swap Fricker's talk about self-ascribing mental states with talk about expressing intentions and self-ascribing actions/intentions, and we get the *Minimal Phenomenon of Agency*:

> *Authoritative*: Expression of intentions/self-ascriptions of actions/intentions are language-*game* authoritative in being ordinarily treated as such (i.e. as strong evidence for its truth).
>
> *Basic*: Expression of intentions/self-ascriptions of actions/intentions are language-game basic in being treated as such (i.e. they are taken as neither needing nor admitting of defence, nor can one require or demand explanations of how the agent knows).

Fricker also adds a principle speaking in favour of an experienced immediacy in self-knowledge (1998, p. 168). Since the phenomenology of agent's knowledge is similar to the phenomenology of self-knowledge also in this respect we should therefore also add:

> *Cognitively Spontaneous*: Expression of intentions/self-ascriptions of actions are psychologically non-inferred (i.e. not formed by a conscious process of inference).

This practice – and its corresponding phenomenology – is taken by many philosophers to reveal something about the nature of intentional agency and the way in which we know our actions.[38] However, as Fricker points out with respect to the *Minimal Phenomenon* (of the Mind), these principles are not directly lending their weight to a specific theory; rather, they count as explanatory constraints that describe a sound practice. It is thus something to be accounted for by a theory of the phenomenon.[39] There might therefore be more than one way to explain the essentials of the *Minimal Phenomenon of Agency*.

As Velleman points out, it is reasonable to think that both statements convey knowledge – viz. reporting a belief and reporting one's intention – and that 'the belief and the intention embody knowledge because they are true and appropriately connected to the facts that make them true' (2007, p. xiv). What our examples and sound practice may be taken to show, however, is that the appropriate connection differs depending on whether one relates to the fact as a believer and a spectator, or whether one relates to it as an active agent. Anscombeans want to take the phenomenology of spontaneousness at face value and in support of the claim that agent's knowledge – as opposed to the knowledge we have when we convey ordinary beliefs – is non-observational, immediate and non-inferential. For them the

[38] As we shall see later, the extent to which we can will be a matter of importance. See chapter 5.
[39] I owe the gist of these points with respect to self-knowledge to Frank O. Barel (2009).

answer to the question, as Hampshire puts it, 'how does the announcement "I know now what I shall do to-morrow" essentially differ from "I know now what will happen to-morrow"?' is thus that there is a difference in epistemology (1960, p. 103). In particular, they think it shows that in a case of intentional agency an agent knows the truth of the indicative content by virtue of settling the issue and deciding what to do. In intentional agency 'you already know' what you are doing, as Velleman points out, 'because doing it was your idea, to begin with' (2007, p. 47). Agent's knowledge is therefore spontaneous because it is the agent that originates what she knows (Velleman 2007, p. 4). In this respect, agent's knowledge departs sharply from other means of knowing. Disputing knowledge in such a case – or inquiring about her source of knowledge – is thus not so much disputing her epistemic stand in relation to the fact of her doing it as it disputes whether she has settled on doing it at all.

For Anscombeans our practice and phenomenology with respect to expressing intentions and self-ascribing intentional actions are thus taken to support the idea that it is our capacity to act for reason that enables an agent to know spontaneously what she is doing and why. Our knowledge in action is, as O'Shaughnessy puts it, first and foremost due to the fact that 'the will provides us with an answer as to how we effect what we do effect' (1963, p. 387). Of course, knowing what you are doing will require a background of observational knowledge; however, the last piece of knowledge – the one that takes its agent beyond her observationally based evidence – is agent's knowledge and is spontaneous. What gives rise to this knowledge is our capacity to act intentionally or for reasons. Following Setiya (2011), we may formulate this Anscombean idea as follows:

> *Spontaneous Knowledge*: If *A* has the capacity to act for reasons (or intentionally), she has the capacity to know *spontaneously* (or without observation or inference) what she is doing and why – in that her knowledge does not rest on sufficient prior evidence.

The assumption is not only that an agent knows why she does something and what she does insofar as she Φ-es for a reason, but that the way in which she knows it is different from the way in which one knows one is falling ill or that one will not succeed in passing an exam. Someone announcing such things can be asked for reasons for thinking those thoughts. The reason is that the content of our beliefs is usually independent of the act of so believing whereas this is not the case for the content of our intentions. The fact that I believe something – but not that I intend something – thus allows room for questions such as 'How do you know

what the outcome will be?'[40] Incidentally, a similar difference reveals itself in the case of *Freud's Inkstand* – as well as the other cases of merely motivated action which I discussed in chapter 1 – since it would not be strange to meet Freud and his explanation for why he destroyed his inkstand (viz. his hidden beliefs and desires) with the question 'How can you tell?' Nor would it be odd for him to emphasize that he knows it in order to convey that he is on to some discovery. He could, for instance, have said something like the following:

> (2) Wait a minute! I am destroying the inkstand so that my sister could buy me another one, I just know I am!

Just as we found in the case of ordinary beliefs, such conversational patterns speak in favour of a sort of independence with respect to the object of knowledge; in Freud's case, an independence that would not be there if he had wrecked his inkstand intentionally. The truth of the content of one's intention, after all, is meant to rely on being the content of one's intention, but that is not the case for the truth of the things one must figure out and discover. If one is voicing one's knowledge in virtue of expressing a belief-state 'things are,' as McDowell says, 'as one takes them to be independently of one's taking them to be that way' and '[o]ne's taking them to be that way is an exercise of a receptive capacity' (2010, p. 417). In the other case, however, the knowledge 'is not derived from the fact known'; rather, the fact one knows by virtue of exercising one's agency and having an intention is essentially bound up with one's will or practical thought. The reason why this kind of deduction does not seem appropriate if the action is intentional or done for a reason is thus also the reason why the question 'How do you know?' is not well put as a response or challenge to speakers expressing their intentions, or conveying what they do intentionally (Velleman 2007, p. xiii). To expect an expression of intention – or self-ascription of an action – to reveal some truth that its agent must have discovered is to dispute that what the agent knows is known by originating it. It would reveal an assumption of undue independence for the object of knowledge. In a word, one would dispute that the knowledge in question is agent's knowledge, or that it fails to be knowledge 'to which the notion of evidence is irrelevant', as Hampshire and Hart say (1958, p. 1), and so dispute that the object known is her action.[41]

Whether one conveys knowledge by expressing a belief or an intention therefore marks a difference in how one relates to the fact expressed by the indicative content. In

[40] I say 'assumed' because there are arguably cases where the truth of what one believes depends on whether one believes it; 'there is,' as William James puts it, 'a certain class of truths of whose reality belief is a factor as well as a confessor' (1892, p. 528). I discuss such cases in chapter 4. Also, see Rae Langton (2003).

[41] '[T]he topic of an intention' is, as Anscombe says, a matter 'on which an opinion is held without any foundation at all' (1957, p. 50).

particular, the reality of the object of knowledge expressed by voicing one's intention depends for its existence on the agent's will. Being appropriately connected to this fact and thereby knowing it by virtue of being its agent is therefore different than knowing it by merely believing this fact. This is revealed in the structure of these speech acts. The assertion, as Hampshire puts it, is an expression of the form 'I know what will happen at *t*'; whereas expressing an intention is of the form 'I know what I will do at *t*' (1960, p. 103). Knowing the first requires prediction and probabilifying evidence; knowing the latter requires no such thing in order to be adequately connected with the facts one thereby knows.[42] We should therefore also specify this way of knowing in the principle of *Agent's Knowledge*:

> *Agents' Knowledge* (AK)*:* If *A* is Φ-ing intentionally (or for reason *R*), then *A must* either be Φ-ing while having non-observational, non-inferential and spontaneous knowledge that she is so Φ-ing, or else she is so Φ-ing by virtue of doing other things – which is not merely trying to Φ – for which this knowledge-condition holds.

LET US TAKE STOCK: I have pursued a line of thinking that affirms the importance of agent's knowledge and that leads towards *AK*. The basic idea is that agents acting intentionally have a special way of knowing that they are acting for some reason, *R*, namely a knowledge that obtains in virtue of so acting. Insofar as we accept *AK*, its truth cries out for explanation. For Anscombeans this invites the idea that an agent's spontaneous knowledge in action bears on the nature of acting intentionally and the notion of taking as one's reason. In particular, they invoke such considerations to support the idea that taking as one's reason is – or, at least, constitutively involves – a distinctively epistemic relation. In this respect, they think that agent's knowledge is not a reality distinct from the reality of what it knows (viz. the action) (McDowell 2010, p. 432). What I will do below is to argue that the necessity of knowledge in action supports the idea that such knowledge pertains to the nature of what it knows.

§5 IMPACT on ACTION THEORY

As we have seen, Anscombe embarks on a discussion of the nature of agent's knowledge in order to throw light on the class of actions that are intentional. In doing so she implicitly acknowledges that the phenomenon of agent's knowledge has a bearing on action theory. In this section, I will try to explicate what this bearing on action theory really is. I will argue that what Anscombe does is a natural consequence of general methodological

[42] As Anscombe warns us, we should not think of agent's knowledge as 'a – queer – species of prediction' (1957, p. 5). Trying to figure out what one is doing, as O'Shaughnessy argues, will disrupt the 'awareness of creating' (1963, p. 387).

principles: namely, that insofar as we accept such a claim we also need to explain why there is a necessity-relation between acting intentionally and acting knowingly. First let me briefly show how a principle like *AK* carves out a group of actions that demands an account of its own: namely, a group of actions about which knowledge is essential.

One way to phrase Anscombeanism is that it puts requirements on the state of mind one must be in insofar as one is acting intentionally. These requirements show how the cognitive state of an agent influences and determines the nature of what the agent is doing intentionally. In particular, it becomes clear that enlightening – or, conversely, benightening – such an agent by altering what she knows about the nature of her own actions may in turn seriously alter the nature of her intentional activities. Knowing that her humming annoys her partner and yet persisting in doing it changes her action from unintentionally annoying to intentionally keeping up this nuisance. When she is in the know, as it were, her humming can no longer present itself as a 'desultory, purposeless and inactive' activity, to use Hampshire's phrase (1960, p. 93). In a sense, the activity has lost some of its innocence. Added knowledge has altered the range of actions for which she is directly responsible by virtue of altering the range of descriptions she can be said to have agent's knowledge of; thus it has altered the range of cases she can perform intentionally. The Anscombean principles thus matter with respect to how one individuates the actions that are intentional and truly one's own.

One consequence for action theory is that one discards – or, at least, qualifies – the deflationary sense of action that Goldman elaborates in *A Theory of Human Action* and which was duly criticized by Anscombe in 'Under a Description'.[43] For Goldman, it is true that we perform most actions unwittingly. To stick with our examples: if you inadvertently annoy your partner by humming Beethoven's Ninth, Goldman would say that you at least performed two actions there. You were humming Beethoven's Ninth and you were annoying your partner. Presumably, you were also doing a load of other things as well, such as warming the air with your breath, or letting your neighbour downstairs know you are home. After all, it is true, as Hampshire says, that '[a]t any moment in any man's waking and conscious life there is always a set of possible true answers to the question – "What is he doing now?"', only a sub-set of which are 'answers which the person himself would give' (1960, p. 169). Goldman's claim is that we should characterize all these true answers as actions.

A source of this deflationary notion of action is Goldman's rejection of what he calls Anscombe's and Davidson's *Identity Thesis* (IT). According to *IT*, the above cases are

[43] See Goldman (1970, ch. 2) and Anscombe (1979).

different descriptions, $D_1,D_2,...,D_n$, which describe one and the same underlying event (or action), *E*, where events are individuated as the same datable particular if and only if they occupy the same region at the same time.[44] For the defenders of *IT* it would therefore be possible to say that most act-descriptions unwittingly apply to one's actions without thereby entailing that we are often benighted about the actions themselves. Anscombe would insist that one must know at least some of the descriptions that apply to the event in order for the event to be one's action. For Goldman, however, the true descriptions, $D_1,D_2,...,D_n$, insofar as they involve different act-properties, correspond to different actions (or events), $E_1,E_2,...,E_n$. Failing to know most descriptions thus translates in Goldman's framework as failing to know most actions with the result that knowledge and agency seem to come widely apart.

Anscombe's vivid instinct as to what is real[45] tells her that something is wrong with Goldman's judgement. With this I agree up to a point. It is true that most actions are done unwittingly only if one fails to make an important distinction between kinds of actions and activities. For an Anscombean, knowing what one is doing belongs to the essence of the kind of acting Anscombe recognized as intentional. We should not run together the actions for which knowledge is essential with the things we can do unwittingly. If we have the former notion in mind, it would be correct to say not only that no actions are done unwittingly, but that such actions could not possibly be performed in a benighted condition. As a counterclaim to Goldman's claim, one could say that there are many true descriptions of someone's action that do not give you what the agent did.[46] Goldman's claim thus risks losing sight of a category of actions marked by the necessary presence of agent's knowledge.[47] Saying this, however, is not the same as taking a stand on the *Identity Thesis*, or to come down in favour of Anscombe's and Davidson's coarse-grained individuation of events. After all, we may recognize the distinctive class of essentially knowledgeable actions and yet think that different descriptions involving different act-properties correspond to different actions in the way Goldman sees it. Recognizing a class of actions as intentional and governed by a knowledge-condition is not deciding what the relationship is between intentional actions and

[44] Davidson initially defended a causal view of event individuation, where events were thought of as identical if and only if they had the same causes and effects (1969a). Later he retracts this suggestion in view of Quine's criticism that this criterion for individuating events presupposes the prior individuation of causes and effects, which are themselves events (Quine 1985). Instead, Davidson accepts Quine's suggestion that events are identical if and only if they occupy the same region at the same time (Davidson 1985).

[45] Bertrand Russell often made appeal to his 'vivid instinct as to what is real' as a prerequisite for doing metaphysics and logic, 'otherwise you will get into fantastic things' (1918, p. 223). I gather that without vivid instincts one will not get far in the metaphysics of action either.

[46] I owe this point to Jennifer Hornsby.

[47] This is not the same as saying that he has lost sight of the class of actions that are intentional. Goldman provides an account of these actions by emphasizing the structure of the *Standard Story*, which we discussed in chapter 1 above. On that story an action is intentional only if it is caused by relevant beliefs and desires in what Goldman calls 'the characteristic way' (1970, p. 62). See Goldman (1970, ch. 3).

the activities and actions that are not intentional. *IT* and the claim that acting intentionally carves out a group of actions for which agent's knowledge is necessary should be kept apart.

Nevertheless, this shows that the Anscombean principles can be employed for important classificatory purposes since they make a distinction between intentional actions and other things we may do unwittingly. One consequence is that it allows us to understand why some activities are more persistently intentional – or, conversely, unintentional – than others. As our example shows, annoying someone proved to be akin to intruding, offending, scaring off, alerting and a medley of other action-verbs signifying act-types agents may perform intentionally or unintentionally, wittingly or unwittingly. Other action verbs behave differently. As Anscombe points out, there are things we must do intentionally – and consequently cannot but do knowingly – such as greeting, signalling, lying, or marrying (1957, p. 85). Yet other things we do cannot be done intentionally, such as forgetting, misspelling, tripping and, as I will touch upon later, forming a belief. A test for our principles is thus whether we can correctly diagnose, categorize and explain these differences. We may for instance understand why some activity is essentially non-intentional as the impossibility of performing it knowingly. The Apostle Peter could not intentionally forget his promise to not reject Jesus since it is impossible to wittingly forget anything. Depending on what one takes the epistemic relation to be, such distinctions can now be explained and the degree to which a candidate epistemic relation succeeds in carving up this landscape will provide evidence for or against one's account. The Anscombean principles therefore have this result that they carve out a group of actions – viz. the intentional actions – for which agent's knowledge is a crucial element. As I shall spend the rest of this section arguing, this in turn necessitates that action theory – in accounting for these actions – must explain the necessity of agent's knowledge.

As we saw in §3 and §4, the principle of *Agent's Knowledge* (or AK) asserts a necessary connection between acting intentionally and acting knowingly (where the knowledge is non-observational, non-inferential and spontaneous). Agent's knowledge, so conceived, will provide those who accept it with a fundamental explanatory constraint on an account of intentional agency. In particular, I shall argue, it follows on general methodological principles that such a necessity-claim makes it urgent to account for the nature of intentional actions in a way that explains or derives the truth of *AK*. To sympathise with Anscombeanism thus makes it imperative for action theory to account for the connection between acting intentionally and acting knowingly. In this sense, agent's knowledge really pertains to the nature of its object by putting constraints on the metaphysics of action.

The structure of the claim made by *AK* is of the general form: all *Fs* must be *G*. Insofar as a philosopher accepts such truths, however, she will also need to explain why the alleged necessity holds between being an *F* and being or having *G*. As Simon Blackburn says, 'it is one thing to assert such necessities but quite another thing to have a theory about why we can do so' (1993, p. 145). The philosopher of action, in particular, who seeks to understand the nature of intentional agency, must thus venture to explain why it holds as matter of necessity that Φ-ing intentionally is to be Φ-ing knowingly in the way specified by *AK*. The background to this demand is a plausible methodological assumption in philosophy: namely, that 'we have to account for necessary connections by appeal to the natures of the things that they connect' (Setiya 2007b, p. 27n10). It must be explainable from an account of what it is to be an *F* why any property or relation, *G*, pertains to an x as a matter of necessity solely in virtue of the fact that the x is of the kind *F*. Alternatively, that this relation falls out of an account of being *G*. Let me refer to this as the *Explanatory Demand*:

> *Explanatory Demand*: If some x must be *G* solely in virtue of being an *F*, then, in order to fully understand the nature of the *Fs*, we must explain why all *Fs* are *G* by appealing either to the nature of being an *F*, of being *G*, or both.[48]

The basic idea is that to fully understand some kind, *F*, we must explain or derive the necessities that come along solely because one is an entity of this kind; otherwise one will not have grasped their nature. Any necessity pertaining to some x insofar as it is *F* will thus potentially say something about its nature since unless one can derive the necessity holding between being an *F* and being *G* from the nature of being *G* there must be something in the being an *F* responsible for the necessity. That constraint will thus pertain to action theory insofar as action theory is in the business of accounting for the nature of intentional actions.

If one accepts Anscombeanism about agent's knowledge – or principles such as *AK* – we must therefore try to explain why it is necessary that an agent knows spontaneously what her action is and why she is doing it since this necessity seems to hold solely in virtue of the fact that her action is intentional. Conversely, if we fail to explain this, we will be left with a brute necessity claim and we would fail to fully understand the nature of intentional actions. After all, it follows from *AK* that it is metaphysically impossible for some agent, *A*, to be Φ-ing intentionally without spontaneously knowing the nature of some of the things she is doing in virtue of which she is Φ-ing intentionally. There is thus a metaphysical limit to the extent to

[48] A similar methodological assumption is defended by Fine (1994; 2000). Also see, Cian Dorr (2004, sect. 1 and 3).

which one may exercise a benighted form of intentional agency. In line with the *Explanatory Demand*, then, this metaphysical limit postulated by the philosopher of action will saddle her with an explanatory burden. We need to know why the necessity holds.

One way to do this is by accounting for what an intentional action is partly in terms of the agent having spontaneous knowledge of what she is doing. If such knowledge pertains to the nature of an intentional action by virtue of being one of its constituent components it would, after all, be metaphysically possible to instantiate intentional agency in the absence of such knowledge. In other words, a metaphysical analysis of intentional actions partly in terms of having spontaneous knowledge is thus a way to explain the necessity of so knowing insofar as one acts intentionally. A metaphysical analysis is all the more tempting because it is an elegant way to explain a necessity-relation without coming into conflict with Hume's stricture on necessary relations between distinct entities or existences. According to Hume, there can be no metaphysically necessary connections between distinct entities since '[t]here is no object which implies the existence of any other if we consider these objects in themselves' (*T* 1.3.6). In contemporary discussions this principle is usually broadened and refined in different ways. First of all, it is usually taken as putting a methodological ban on necessary relations between not just distinct objects but also any entity that belongs to an ontological category that may be the relata of causal relations. That will include a ban on necessary relations between distinct events, facts, and so on. Second, it is meant primarily as a restriction on metaphysical necessity-relations. In this way, weaker kinds of necessities – such as nomological necessity-relations – may be tolerated.[49] Finally, one should take care to allow necessary relations between distinct entities that are *extrinsically typed* – or between entities that are not exclusively characterized by their intrinsic properties – so that the restriction only ranges over entities that are *intrinsically* typed.[50] The result is the following:

> *Hume's Dictum:* There are no metaphysically necessary connections between distinct, intrinsically typed, entities.[51]

Many philosophers agree with *Hume's Dictum*. Some take it to be immediately evident or intuitive. Others take it to express an analytic truth. Yet others treat it as an inference to the best explanation. I will not take a stand on this issue in what follows; however, I will register *Hume's Dictum* as something that would add explanatory burden to one's account insofar as

[49] Or maybe not: on some views nomological necessity is a kind of metaphysical necessity. See Sydney Shoemaker (1980; 1998).
[50] As Jessica Wilson notes, this is the contemporary correlate of Hume's asking us to consider objects 'in themselves' (Wilson 2010, p. 596).
[51] For this formulation and its qualifications in contemporary discussions see Wilson (2010, pp. 595–6).

the account is in conflict with it. It will therefore be registered in my discussion as a potential theoretical cost: other things being equal, one should prefer an account *A* over account *B* if *B* is in conflict with the dictum while *A* is compatible with it. Given *Hume's Dictum* and the *Explanatory Constraint* the strategy of pursuing a metaphysical analysis of intentional actions in terms of what it necessarily entails – viz. spontaneous knowledge of one's own action – thus recommends itself. An analysis of this sort would, after all, explain *AK* without conceiving acting intentionally and having spontaneous knowledge of so doing as distinct, intrinsically typed existences. In this respect, accepting principles such as *AK* may provide pressure towards a reductive account that can analyze and distinguish constitutive components of the relation one must take towards Φ-ing insofar as one is Φ-ing intentionally.

Setiya has criticised Anscombe for failing to address this task. According to him, the observations made by Anscombe in *Intention* leaves action theory with the task of explaining, on the one hand, why the question, 'Why did you do it?' or 'Why are you doing it?' – when it is a request for the agent's reasons for acting – necessarily applies to an action insofar as the agent is doing something intentionally. On the other hand, action theory will need to address the truth of *AK* and why agent's knowledge is necessary for acting intentionally. On his view, addressing these questions requires 'an account of *when* the question applies', which is something Setiya thinks Anscombe fails to do in *Intention* (Setiya 2007b, p. 26). Moreover, Setiya thinks that the connection between these two marks of intentional agency suggests a joint account that requires venturing into the metaphysics of intentional agency. In particular, Setiya takes it to add 'pressure towards reductive psychological accounts of acting for reasons' (2007b, p. 28). In his view, our need to explain Anscombe's insight in *Intention* thus supports a causal-reductive account of acting intentionally. Such an account will explain intentional actions as behaviour being causally sustained by a complex of carefully qualified attitudes and where at least one of these attitudes embodies the agent's knowledge in action.

If we look closer at this, it becomes clear that Setiya takes principles such as *AK* to suggest a metaphysical analysis where an agent's knowledge in action is accounted for as one of the constitutive components of the relation an agent has towards the events that are her intentional actions. According to Setiya, an agent relates to her actions in the required way when she is taking or treating some reason, *R*, as her reason for Φ-ing.[52] If Setiya is right, we must 'break down what is involved' in that relation if we are to 'see how one of its elements is, or involves, the relevant [epistemic elements]' (2007b, p. 28). We must then analyze this

[52] I discussed this notion in chapter 1.

relation and find that it contains, at least in part, spontaneous knowledge of what one is doing and why in order to derive the truth of principles such as *AK*. Once this complex is spelled out the idea is to define intentional actions as the behaviour causally sustained by an agent taking something as her reason for acting. Since taking as one's reason, according to such an analysis, must contain spontaneous knowledge, it would follow that an agent must have such knowledge of what she is doing and why insofar as she really is acting intentionally.

This strategy will only succeed if the *analysandum* (viz. acting intentionally) really is a metaphysical complex constituted by more basic building blocks, $B_1,B_2,...,B_n$, and one of these constitutive elements is an agent having spontaneous knowledge of what she is doing and why.[53] It might be 'natural', as Setiya says, to 'begin with the idea that acting for a reason can be [reductively] explained in terms of the agent's psychological states' (2007b, p. 28); but this is not the only alternative. There might be constituents of intentional actions that are not explicable in terms of the 'causal pushes and pulls', to use Bratman's expression, of an agent's psychological states (2001, p. 91). Perhaps we need invoke character traits as inherently evaluative features irreducible to a set of current.[54] Another way to go is to conceive the will – or practical reason – as an epistemic faculty and define or explain intentional actions as the favourable output of this epistemic faculty. On that model we may keep intentional actions as a non-composite, metaphysical primitive or simple relation while still being able to understand the necessity of agent's knowledge.

What we can agree on at this stage, however, is that to accept the truth of something like *AK* is to sign up to a certain conception of agency that I shall recognize as distinctively Anscombean. Such a conception suggests that the knowledge an agent has in action – whether the nature of this epistemic relation is best characterized in terms of knowledge, suitably qualified belief, and so on – pertains to the nature of what it knows (viz. the intentional action). I shall recognize any such theory – whether it be reductive or not – as an *Anscombean Theory* of action. An Anscombean Theory of action will need to address the issues that Setiya accused Anscombe of ignoring: namely, why the two marks of intentional actions – their susceptibility to reason-requests and spontaneous knowledge of the nature of the action – hold as a matter of necessity. Conversely, it follows that insofar as an account fails to be Anscombean in this sense it also fails to account for the necessity of agent's knowledge. Insofar as one endorses the necessity of agent's knowledge, it is a criterion of explanatory adequacy that one can derive the truth of something like *AK* from one's account of agency.

[53] For a similar claim, see O'Brien (2007, p. 129).
[54] See Williams (1995b) and Schueler (2003, ch. 3).

Setiya suggests that we can use this criterion of adequacy to diagnose the failure of the *Standard Model*. As you may recall from the previous chapter, the *Standard Model* accounts for intentional actions as behaviour caused in the right way by a relevant belief-desire pair. A problem with that model, or so I argued, is that it fails to rule out (or distinguish) cases of merely motivated behaviour from intentional actions and things we are doing for reasons. There are, after all, 'occasions on which our behaviour is guided purposively by desire, but on which it seems quite wrong to say that we are acting for reasons' (Setiya 2007b, p. 33). I concluded that there is something missing on the model – some illuminating condition – that makes the difference between merely motivated behaviour and intentional actions. A diagnosis of its failure is to say that it ignores the truth of something like *AK*. An agent may be moved to Φ by a belief that so Φ-ing is a way to realize some desired end and yet fail to know, in the way specified by *AK*, that she is so Φ-ing. Consequently, it is possible on the *Standard Model* to Φ for a reason without knowing that one is so Φ-ing. The problem with the *Standard Model* is thus that it has no resources to rule out cases of benighted agency since they would qualify as intentional actions on this model. We could therefore reject the *Standard Model* because it fails to derive the truth of *AK*. This demonstrates 'the force of the explanatory demand,' Setiya says, as '[i]t shows that this demand has bite' (2007b, p. 33). [55]

Similar considerations apply to models of intentional agency devised to account for the truth of principles such as the *Guise of the Good*, while ignoring the necessity of agent's knowledge. The *Guise of the Good* is, as you may recall,[56] a universal claim about how one must see one's action or reasons for acting insofar as one is acting intentionally:

> *Guise of the Good*: If *A* Φ-es for a reason (or intentionally), *A* must see some good in Φ-ing.

An account of intentional agency, may take this principle as its point of departure. After all, the *Guise of the Good* offers an alternative diagnosis for the failure of the *Standard Model*: namely, that it fails to distinguish actions whose agent is guided by a concern for good reasons from actions whose agent is not so concerned. If a theory of such guidance can be worked out, the difference between acting for reasons and being merely motivated to act

[55] Davidson's theory of intentional action can be interpreted as entailing the truth of something similar to *AK*. According to Davidson, 'someone does something for a reason' only if 'he can be characterized as: (a) having some sort of pro attitude towards the actions of a certain kind; and (b) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind' (1963, pp. 3–4). The second clause may, as Setiya points out (2007b, p. 33n21), be interpreted in two ways: (1) as requiring that the agent believes that a certain action would belong to the desired kind; or (2) as requiring that she believes that she is performing an action of the desired kind. (1) is captured by Velleman's formulation of the *Standard Model* and would clearly fail to entail the necessity of agent's knowledge. But (2) would entail a principle analogous to *AK* that replaced the necessity of knowledge with the necessity of a belief. Even so, however, we would 'need to know *why* the first interpretation is not enough,' as Setiya says. We need to know the following: 'Why does it take more to attach one's reason to one's action than motivation by belief?' (2007b, p. 33n21)

[56] I discussed this doctrine in chapter 1, §7.

might turn out to be the difference between being so guided by the normative strength of one's reasons for acting and the absence of such guidance. The standard way to provide such a story is by invoking *distinctively practical attitudes* and commitments to what one decides to take or treat as one's reason for Φ-ing. Such stories, however, may fail to account for why such processes must also involve agent's knowledge. '[I]t simply does not follow from the fact that I take p as a good reason to Φ, and thus believe that it is a good reason to Φ', Setiya points out, 'that I believe that I am actually doing Φ, or that I am doing anything at all' (2007b, p. 41). A theory may thus respect the *Guise of the Good* without doing anything to explain why actions done for reasons must be subject to spontaneous knowledge. This does not prove that it is wrong to focus on the *Guise of the Good* in action theory; but give such claims 'considerable force', as Setiya says, if one 'does nothing to solve the puzzle' that is important for action theory: namely, why something like *AK* holds (2007b, p. 41).[57]

What this shows is that one can put pressure on theories of intentional action by asking if they can account for and explain agent's knowledge. This demonstrates the force of *AK* as an explanatory criterion and how it may support Anscombean Theories of action. Of course, the demand only bites insofar as one accepts the truth of something like *AK* and the necessity of agent's knowledge. A problem is that there is no general agreement on the necessity of such knowledge in action. On the contrary, there is 'a profound conflict in action theory between those who follow Anscombe in seeing a close connection between intentional action and knowledge of what one is doing and those who do not,' as Setiya observes (2009a, p. 131). What I have argued so far is that conditional on accepting *AK* as true, the *Explanatory Constraint* and *Hume's Dictum* will work in favour of an Anscombean Theory of action. A dialectical point against philosophers who are not Anscombeans in this respect will thus come down to the issue of whether one has reason to accept something like *AK*. In this chapter I have given some *prima facie* evidence for the truth of such a principle. In the face of sceptical counterexamples we were able to replace it with a more accurate principle that nevertheless captured the gist of Anscombe's insight in *Intention* (§4). We also saw that the peculiar nature of an agent's knowledge in action could be explained by *AK* when it emphasized that the knowledge is non-observational, non-inferential and spontaneous (§3). At the very least, these data show that agent's knowledge is a pervasive and peculiar phenomenon in intentional agency and therefore something to be explained by action theory.

[57] In order to show that it is wrong, one would also need to show that it is possible to have an account of acting intentionally that explains what it is to take something as one's reason in a way that explains *AK* without invoking normative principles such as the *Guise of the Good*. I will consider Setiya's attempt at so doing in the subsequent two chapters.

§6 CONCLUSION

Above we started out with Anscombe's project in *Intention* and how she emphasized the way we know our actions as offering a way to break into and illuminate the conceptual knot that confronts us when we talk about intentions, intentional activities, and our reasons for doing something, and so on. I then sought to characterize the phenomenon of agent's knowledge and I argued in favour of a principle, *AK*, as the best to capture what Anscombe was after in *Intention* before I gave *prima facie* evidence for accepting *AK* as true. I turned from that to argue that accepting a principle like *AK* will bear on our understanding of the nature of intentional actions. In particular, I have argued that *AK* – with its emphasis on the necessity of non-observational, non-inferential, and spontaneous knowledge in action – will favour what I have defined as an Anscombean Theory of action, or a theory that may explain or derive the truth of *AK* from its characterization of intentional actions. What will follow below, concerns largely whether such a knowledge-condition really holds as a matter of necessity and how we should best characterize its nature. There is still room for doubting three things about agent's knowledge. First, one may, like Setiya, think that the necessary epistemic relation an agent must bear to her actions insofar as they are intentional is best characterized as a suitably qualified belief rather than knowledge. This will be the topic of chapters 3 and 4 below. Second, one may perhaps accept the pervasiveness of agent's knowledge and some of its features while nevertheless doubting its necessity. Third, one may doubt whether such knowledge really is spontaneous rather than inferential. The last two issues will recur in chapter 5 when we address Paul's Inferential Theory. In what follows, I seek to dismiss these doubts in order to claim, as Anscombe does in *Intention*, that the knowledge one has in action really is spontaneous knowledge and that it is intimately connected with what it is to act intentionally and thus has a bearing on its nature.

3

Cognitivism about Intention

The will provides us with an answer as to how we effect what we do effect.

Brian O'Shaughnessy[1]

§1 INTRODUCTION

As we saw in chapter 1, Setiya thinks a non-normative explanatory framework is sufficient as an account of the nature of acting intentionally and he takes its sufficiency to challenge and dispute the kind of framework ethical rationalists need in order to satisfy the *Difference Principle*. Setiya's claim is that a *Minimalist Cognitivist* is enough to characterize what it is to take something as one's reason; he thereby provides basis for an account of acting intentionally. In this chapter I aim primarily to describe the details of this alternative framework and say why Setiya thinks it gets support from the requirement we discussed in the previous chapter, namely that action theory must explain why an agent's knowledge in action is a non-contingent or necessary feature of acting intentionally or for reasons. I shall (§2) say why Setiya thinks the observations made by Anscombe and Hampshire supports the idea that intentions are a kind of belief or belief-like cognitive attitude. I shall then (§3) discuss what kind of belief this could be, before turning to initial worries pertaining to that view (§4).

§2 The PRIMACY of COGNITIVE ATTITUDES

In what follows, I understand *Cognitivism about intention* as an effort to explain and account for the nature of our intentions in action – viz. the intention *with which* an agent is Φ-ing insofar as she is Φ-ing intentionally and by virtue of which she is taking something as her reason for so Φ-ing – as, at least in part, a sort of belief or belief-like cognitive attitude.[2] I

[1] O'Shaughnessy (1963, p. 387).

[2] Cognitivism *about intention* should not be confused with other cognitivist views in ethics and action theory and which may, or may not, go together with Cognitivism about intention. In meta-ethics, the term is perhaps primarily associated with the view that normative or ethical judgements are like other indicative claims and assertions in that they express matters of fact or truth-evaluable propositions. A non-cognitivist about ethical judgement will then deny that the proper function of ethical claims is to state something truth-evaluable that could be the object of cognitive states and attitudes. According to them, moral claims lack descriptive meaning (Mackie 1977, p. 40). What one

shall also follow Velleman (1992a, p. 105) and Nishih Shah and Velleman (2005, p. 503) in assuming that cognitive attitudes are a mental state the occurrence of which represents some proposition or content *as being true*.[3] It follows that cognitivists about intention will treat the relation taking as one's reason as, at least in part, a matter of representing something as being true. In this section I shall discuss why one may think that such a view can be motivated by the explanatory requirement that I argued for in the previous chapter: namely, to explain the striking pervasiveness of non-observational knowledge in intentional action, or what I defined as the phenomenon of *agent's knowledge*.

Following Paul (2009a, pp. 3 and 11; 2009b, pp. 550–1), I shall distinguish between *Strong* and *Weak* versions of Cognitivism. The strong version identifies an agent's intention in action – or her taking something as her reason – with having a belief (or belief-like cognitive attitude) that one is Φ-ing (or, in the progressive, that one will be Φ-ing).[4] *Weak* versions hold that this *constitutively involves* such a belief or belief-like cognitive attitude, but leave room for non-cognitive or practical attitudes as constitutive elements of so acting:

---

expresses when one makes a moral evaluation – if one expresses anything at all – is not a moral fact akin to a proposition with an objective truth-value; rather, one gives voice to a mental state and the mental states one gives voice to are more like desires, emotions, or other conative attitudes (Wedgwood 2007, p. 4). Recently, Mark Kalderon has explored a *fictionalist* version of non-cognitivism, where the content of one's mental attitude is a proposition but one accepts it by bearing a non-cognitive attitude to it (2005). Many non-cognitivists combine their opposition to cognitivism with an *expressivist* view about what the function of ethical claims is. On these accounts what one expresses with a moral claim is not simply the truth-evaluable content of the proposition '*A* ought to Φ' or 'Φ-ing is the morally appropriate thing to do'; rather, they express specific mental states one has in virtue of which one counts as making the moral judgement. Expressivists claim to be able to account for what kind of state this is as well as its content without normative terms. For non-cognitivist versions of expressivism, see Allan Gibbard (1991) and Blackburn (1993). Also, see Terence Horgan and Mark Timmons (2000) for a cognitivist version of expressivism (or what they call *non-descriptivism*). For critical discussion, see Smith (1994, ch. 2) and Wedgewood (2007, ch. 2).

Another respect in which one can be a cognitivist is by opposing a Humean (or non-cognitivist) view of human motivation. In contemporary philosophy, *Humeanism* is the claim that motivation only comes about as the combination of cognitive attitudes, such as believing, perceiving, noticing, etcetera, with conative attitudes, such as desires, wants, wishes, and so on. However, they take the latter to be the dominant force and necessary for being motivated. In opposition to that the cognitivist about motivation thinks that 'belief alone is capable of motivating action' (Dancy 2000, p. 13) and thus that motivation can be a cognitive matter. The cognitive attitude may, for instance, cause a conative attitude that in turn moves the agent to act (Nagel 1970, ch. 5); or the cognitive attitude may directly bring the action about, as a *pure* cognitivist would argue (Dancy 1993, chs. 1–3; 2000, ch. 4).

(The sense in which contemporary Humeanism connects to the historical Hume is not always clear; however, it ties in with Hume's claim that 'reason alone [i.e. the understanding that judges from 'demonstration and probability'] can never be a motive to any action of the will' (*T* 2.3.3.1). According to Watson's characterization of Hume, reason is in a sense completely 'dumb on the question of what to do' (1975, p. 207). Watson contrasts Hume's view with Plato's view of practical reason in the *Republic*. Given Hume's segregation of reason and motivation, motivation becomes a given and not something regulated by the faculty of reason. As a result, reason becomes a 'slave of the passions' in practical matters on Hume's framework (*T* 2.3.3.4) where it was, and ought to be, the master on Plato's.)

A third way to be a cognitivist is by vouching for something like Velleman's view about practical reason (Bratman 1991, pp. 250–1; Wallace 2008). Velleman combines the claim that intentions are beliefs with the idea that practical reasoning is a kind of theoretical reasoning and governed by cognitive standards. His view is thus cognitivist in two respects. See Velleman (1985; 1992b; 1996; 2000b, *Introduction*; 2007; 2009). In the following, it would be advisable to keep these doctrines apart since Cognitivism about intention need not lead one to be cognitivist about practical reason. Setiya, for one, endorses the former while thinking that practical thought is radically different from theoretical reasoning (see Setiya (2003; 2004; 2007a; 2007b, pp. 19, 45–6; 2009b)). I will therefore refer to Setiya's position as *Minimal Cognitivism* (or *Minimalism*, for short) and refer the term 'Cognitivism' to only being cognitivist about intentions.

[3] Cognitive attitudes are mental state whose content is 'grasped as patterned after the world' (Velleman 1992a, p. 105). In contrast to this distinctively cognitive frame of mind are what I shall refer to as *practical* (or *conative*) attitudes, whose contents are 'grasped as a pattern for the world to follow' (Velleman 1992a, p. 105). In short, someone having a cognitive attitude towards *P* regards *P as true* whereas having a practical/conative attitude to *P* is to regard it as to *be made true* (Velleman 1992a, p. 105).

[4] The distinction between prospective (or *prior*) intention and intention *in* action – or, respectively, between 'those intentions that are formed prior to actions and those that are not' – is usually credited to John Searle (1983, pp. 84–5). Also see McDowell (2008b). The linguistic expression of prospective intentions is 'I will Φ' or 'I am going to Φ'. As Searle points out, one may Φ intentionally without a prior prospective intention to Φ just as one may have a prospective intention and not act on it (so-called 'pure intentions'). But when one Φ-es intentionally as a result of a prospective intention the close connection between the prospective intention and the intention in action must be explained. The cognitivists may hope that their story can explain both the nature of a prospective intention and an intention in action while allowing them to understand the connection between what Anscombe called the 'different senses' of intention (1957, p. 1).

Weak Cognitivism: Intending to Φ constitutively involves the belief that one Φ-es (or will Φ).

Strong Cognitivism: Intending to Φ just is having the belief that one Φ-es (or will Φ).[5]

In what follows, I will focus on the strong version of this view because it is in this form that Cognitivism has been taken by some philosophers of action – in particular, Setiya (2003; 2007b; 2009b; 2010b) – as providing a basis for a minimal framework of action theory whose sufficiency will count against evaluative or normative interpretations of the nature of intentional agency.[6] As such, Cognitivism about intention might provide Setiya with his desired lemma against the strategy in ethics that I defined earlier as Ethical Rationalism.[7]

According to cognitivists, the unique best way to account for agent's knowledge is by treating an agent's intention in action – viz. the intention *with which* she is Φ-ing intentionally or for a reason – as a belief-like cognitive attitude. When an agent is Φ-ing for a reason *R* she is taking *R* as her reason for Φ-ing. According to cognitivists, the agent does this by virtue of having a belief-like cognitive attitude towards the proposition that she is (or will be) Φ-ing for reason *R*. Let me run through the kind of thinking that may lead to Cognitivism.

Cognitivists take off from the Anscombean idea that the pervasiveness of non-observational knowledge in action is a phenomenon whose explanation requires that such knowledge pertains to the nature of acting intentionally. Anscombeans argue that the phenomenon of agent's knowledge – i.e. the pervasiveness of non-observational knowledge of *what* one is doing (self-awareness) and an understanding of *why* one is doing it (self-understanding) – may be taken as support for the claim that there are epistemic components or aspects to the relation an agent bears towards her actions insofar as these actions are intentional.[8] The cognitivist thus takes it that an Anscombean Theory of action is called for so that we may account for the non-observational nature of agent's knowledge and the fact that it is no accident that agents have this kind of knowledge insofar as they are acting intentionally. However, their strategy is to treat intentions in action, as well as prospective intentions, as a kind of belief or suitable belief-like cognitive attitude. For the Anscombean observation to

[5] Strong cognitivists involve Velleman (2000b; 2007; 2009) and Setiya (2003; 2007b; 2008b; 2009a). Grice (1971), Robert Audi (1973) and Wayne Davis (1984) have defended Weak Cognitivism. For more on this distinction, see Paul (2009a, p. 3; 2009b, p. 547n5).

[6] A reason for not focusing on Weak Cognitivism is that this position is even more vulnerable to the epistemic problem I discuss in chapter 4 below. Since the belief on that view will only be registering one's practical thinking it will not be sufficient to 'create the fact' it represents, as William James once put it (1892, p. 474), and may thus risk making intentions 'a case of wishful thinking' (Grice 1971, p. 268). See Paul (2009b, p. 552). Another problem is that it makes the belief-component epiphenomenal in the sense that it will leave the motivational work to be done by a prior desire or practical attitude. See Bratman (1999b, pp. 19–20) and Setiya (2010a). In order to give Cognitivism a fair hearing, I will focus on its most plausible version. I also argue against Grice's version of Weak Cognitivism in the appendix to chapter 5.

[7] Ethical Rationalism is a strategy that seeks to vindicate the normative or rational authority of an ethical life by virtue of arguing that such living is recommended on independent ground from the nature of practical reason. See chapter 1, §3.

[8] See chapter 2, §5.

support the cognitivist approach it must therefore be the case that a belief about what one is doing and why is 'the core phenomenon from which all instances of intentional action [and the relevant knowledge in action] are derived' (Setiya 2010a). In other words, what Anscombe and Hampshire originally claimed – viz. that agents acting intentionally must *know* what they are doing and why – must be re-phrased and qualified if it is to support the claim that there is a *belief-like cognitive* component to taking as one's reason.

According to Setiya, the Anscombe-Hampshire claim that agents must have 'knowledge without observation' must therefore be qualified (Anscombe 1957, p. 13). For a start, Setiya thinks we should bracket Anscombe's emphasis on the practical dimension of this knowledge. Anscombe was on the right track, Setiya says, when she emphasized the non-observational nature of our knowledge in action and that it is, as Velleman puts it, spontaneous knowledge and not a matter of finding out (Velleman 2007, p. 24). Like Anscombe, cognitivists think that agent's knowledge differs from ordinary knowledge in that the agent knows by bringing about the facts rather than being caused to know by the facts. It is even right, they think, to take it that one may explain this phenomenon by saying that what one is doing intentionally is 'known by being the content of [one's] intention', as Anscombe says (1957, p. 53). Cognitivists are thus, like ordinary Anscombeans, prone to accept that there is some truth to what I defined in the previous chapter as principle *AK*:

> *Agents' Knowledge* (AK)*:* If *A* is Φ-ing intentionally (or for reason *R*), then *A must* either be Φ-ing while having non-observational, non-inferential and spontaneous knowledge that she is so Φ-ing, or else she is so Φ-ing by virtue of doing other things – which is not merely trying to Φ – for which this knowledge-condition holds.

Where cognitivists like Setiya draw the line is when Anscombeans start taking *AK* to require talk about a distinctively practical way of knowing these things that is, as Aquinas said, 'the cause of what it understands' (*ST* 1a.2ae.3.5.1) and for which 'the notion of evidence is irrelevant' (Hampshire and Hart 1958, p. 1).[9] What cognitivists want is to retain a notion of knowledge of what you are doing (and why) such that, although you arrive at it non-observationally and spontaneously, it is nevertheless ordinary propositional knowledge (Velleman 2007, p. 25). Setiya thus argues that we should bracket the practical dimension of agent's knowledge in order to concentrate, as he says, 'on a simpler claim' (Setiya 2007b, p. 25); namely, the claim that Anscombe's question 'Why?' is refused 'application by the

[9] See Anscombe (1957, pp. 87–8).

answer: "I was not aware that I was doing that"' (Anscombe 1957, p. 11). The alleged simpler claim is thus that '[w]hat I do unknowingly, I cannot do intentionally' (Setiya 2007b, p. 25); in other words, that we interpret 'knowledge' in *AK* as referring to ordinary propositional knowledge.[10] For cognitivists, the truth of *AK* would thus suggest that the frame of mind by virtue of which an agent acts intentionally must embody a representation of things as they *are* arranged – rather than representing them as they are *to be* arranged – in order to be able to embody and carry the weight of non-practical knowledge.[11]

The next cognitivist move is to qualify the claim to *know*. Setiya thinks we should drop knowledge because it is too epistemically ambitious. It is not generally required to be in such an epistemically demanding state, he argues, in order to act intentionally. Rather, sometimes all one needs is a belief or belief-like cognitive attitude. Ordinary Anscombeans may be right in thinking about the will as 'a capacity to know what one is doing, or of what one is going to do,' Setiya says, but the best way to think of the will is as a faculty whose exercise 'may be impeded, yielding mere belief or partial belief instead of knowledge' (Setiya 2010a).[12] According to Setiya, a reason for thinking that a mere belief may be sufficient is that one may imagine cases where an agent is Φ-ing intentionally while being epistemically precluded from knowing that she is so Φ-ing. Imagine, for instance, that you are intentionally clenching your fist. According to Setiya, you will have to know that you are able to clench your fist in order to know that you are doing it. That is, you need to know that if you intend to clench your fist, you will in fact clench it. If you do not have this conditional knowledge, the knowledge that you are clenching your fist will be undermined; or so he argues. Suppose, however, that you have been suffering from a paralysis of your hand for a while, but that due to excessive optimism you nevertheless believe that you can clench your fist. Suppose further that your belief happens to be true because the paralysis has unbeknownst to you miraculously cured itself. Then what you believe you are doing – which is only accidently true – will fail to be knowledge. Yet insofar as you are able to clench your fist and by so doing seem to be doing something intentionally, knowledge of what you are doing does not seem to be required. All you seem to need is a relevant belief.[13] According to Setiya, we should therefore revise *AK* and take a principle like *Belief* as our primary explanatory target in action theory:

[10] I say 'alleged' because it is not clear that the claim that an agent has knowledge of what one is doing is simpler than the claim that the agent has practical knowledge of what she is doing. It depends, in part, on the nature of practical knowledge. See chapter 6.

[11] See Velleman (2000b, p. 24). For the moment, I will simply leave this as it is since the practical dimension of Anscombe's claim will be the focus of later discussion (chapter 6). I hope that the outcome of our present discussion will show that Setiya and other cognitivists are far too quick to give up Anscombe's original insight that knowledge in action is practical.

[12] For a similar point, see David Pears (1985, pp. 78–81).

[13] Setiya (2007b, p. 25; 2008b, pp. 389–90). For the record, I think this argument fails since it trades on the assumption that knowledge of what one is doing does not come about in virtue of the fact that one is doing it intentionally. See §4 below and chapter 6.

> *Belief*: If *A* is Φ-ing intentionally, then *A must* either be Φ-ing in the belief that she is Φ-ing, or else she is Φ-ing by doing other things – which is not merely trying to Φ – in the belief that she is doing that.[14]

If Setiya is right in thinking that Belief is the best way to retain and characterize the gist of what Anscombe and Hampshire were claiming – and I shall return to this question below – he can push Belief in front of him as a consideration that favours the idea that taking as one's reason is to form a sort of belief or belief-like cognitive attitude. The reasoning will go as follows: endorsing the truth of Belief will give us the explanatory of explaining why this necessary relation holds between acting intentionally – viz. by taking something as one's reason for Φ-ing – and having a belief about what one is doing. That is, Setiya may insist that 'the puzzle about *Intention* is to explain the necessary connection between acting for reasons and having a true belief about what one is doing' (2007b, p. 27). Just as in chapter 2, this will be the task of explaining a metaphysical necessity of why all *Fs* must be *G*. Given that the '*G*' for Setiya is a belief, the truth of such a claim can be taken to support the idea that taking as one's reason (viz. the *F*) is a sort of belief (viz. the *G*). If we take *Belief* seriously, it is therefore natural, Setiya argues, to think that the agent's intention in action must embody belief about what one is doing why. Setiya therefore takes the truth of *Belief* to suggest two things. First, that we accept *Cognitivism about intention* and the view that an intention – or taking as one's reason – *just is* having a belief. Second, that we account for intentional actions as behaviour causally sustained by having the intention-belief (viz. by taking something as one's reason). The two cognitivist reductions are as follows:

> *C1*: *A* is Φ-ing intentionally or for reason *R* if, and only if, if so Φ-ing is causally sustained by *A*'s intention to Φ (e.g. by *A* taking *R* as her reason for Φ-ing).

> *C2*: *A* is taking *R* as her reason for Φ-ing if, and only if, *A* forms a kind of belief (or belief-like cognitive attitude) that she is (or that she will be) so Φ-ing.[15]

[14] Setiya (2007b, pp. 26 and 41; 2008b, p. 390).

[15] The qualification here is due to the fact that Φ-ing intentionally does not always require an intention whose object is Φ-ing. In other words, what Bratman dubbed 'the Simple View' – or the view that an agent, *A*, can only intentionally Φ if *A* had the intention to Φ – is false as it stands (1999b, p. 112). As Bratman points out, the *Simple View* is a special case of a more general conception – the *Single Phenomenon View* – on which intentional action and the state of intention 'involve a certain common state' and which holds that the action's relationship to this state determines whether or not it is intentional (1999b, p. 112). The *Simple View* is the suggestion that the relationship requires an intention to *so act*. What is right about this view is that an intention to act is required to act intentionally, and that the latter must relate in *some* way to the former state. Where it goes wrong is in requiring an intention to Φ for any case of intentionally Φ-ing. The truth is that sometimes one intentionally Φ-es by relating to the agent's intention to do other things.

According to Setiya, filling out the details of something like *C1* and *C2* will provide what we to be found missing on the *Standard Model*,[16] while at the same time providing the basis for a framework that can explain why an agent must have non-observational and spontaneous cognition in action. For cognitivists who endorse *C1* and *C2*, it will indeed be true, as O'Shaughnessy says, that it is 'the will [that] provides us with an answer as to how we effect what we do effect' (1963, p. 387). They may also be in position to grant Anscombe that '[w]hat is necessarily the rare exception is for a man's performance in its more immediate descriptions not to be what he supposes' (1957, p. 87), without being vulnerable to the kind of case that Setiya launched against *AK* and the claim that knowledge must be involved.

The challenge for cognitivists is, of course, to account for the nature of the belief that *C2* mentions. Let me now turn to how cognitivists have dealt with this issue.

§3 The INTENTION-BELIEF

There are two main competing cognitivist accounts of the belief-like cognitive attitude that is supposed to fill the roll of an intention in action – viz. the *intention-belief*, for short. The first model invokes a psychological tendency in the agent to make sure that beliefs about one's anticipated, prospective doings (viz. the belief that I am Φing, will Φ, or will be Φing) will make themselves true. I shall call this a *full-blown* cognitivist account – or also Cognitivism about Practical Reason – since it assumes not only that intentions are belief-like cognitive attitudes but also that practical thought is animated, just like theoretical reasoning, by a deep-rooted concern for having true beliefs. One counts as having done things for reasons only insofar as one's behaviour is guided by this deep-rooted concern for having true beliefs about oneself. This is Velleman's model and its challenge is to cash out how a concern for true beliefs can animate practical thought and make us do what we think we will do.[17] The other cognitivist model seeks to avoid this additional cognitivist claim by adopting a *Minimal Cognitivist* account (or *Minimalism*). Minimalism drops talk about cognitivist psychological drives and constitutive aims of practical thinking, and tries to explain the motivational work of an intention-belief – so that holding it allows agents to act for reasons – by portraying it as a *desire-like* second-order belief whose content is self-referential. The latter, of course, is Setiya's model. The main difference between these views is where they choose to locate the additional complexity that comes along with Cognitivism about intention in order to account for the relevant motivational force of the intention-belief. In this section, I shall briefly

[16] See chapter 1, §6.
[17] See Bratman in his review essay on Velleman's book, *Practical Reflection* (Bratman 1991, pp. 250–1).

characterize the details of these views and explain why I think Setiya's Minimalism is preferable to Velleman's full-blown Cognitivism.

Like Setiya, Velleman does not set out to explain why agents acting intentionally must have knowledge about what they are doing (and why) but rather why '[w]hat you do intentionally' is also something you 'tend to do knowingly' (Velleman 2007, p. xxiv). To explain this, Velleman suggests a twin identification: namely, that, on the one hand, 'practical reasoning is a kind of theoretical reasoning', while on the other 'practical conclusions, or intentions, are the corresponding theoretical conclusions, or beliefs' (2007, p. 15). The governing idea is thus that an intention with which the agent acts is a belief about what she is doing (or will be doing) which is formed as the result of practical reasoning. The functional role of an intention is, of course, to guide and motivate its agent; so in order to have a full story of intentional action, Velleman must say something about how these beliefs fill the proper motivational role of intentions in our cognitive economy. What Velleman does is to invoke a sort of 'motive for self-inquiry' (2000a, p. 231) by postulating a deep-seated cognitive drive that makes sure that 'people are generally guided in their behavior by a cognitive motive toward self-understanding' (2006, p. 8). This drive will make sure that practical reasoning will aim at reaching conclusions about what to do – viz. intentions – that contribute to the agent's self-understanding by virtue of making sense of herself to herself. Such an inquirer will then quickly learn, Velleman argues, 'that he can make sense of himself by making sense to himself – that is, by doing what makes sense of him' (2009, p. 17). The aim of self-understanding can be achieved in two ways, as it were: by virtue of forming correct beliefs about oneself or by virtue of chipping in to contribute to the truth of what one believes about oneself. Given this drive for self-understanding Velleman can therefore predict that an agent who believes that he is Φ-ing – or will be Φ-ing – on the basis of reasoning about what it would make sense for someone like him to do will also tend to make this belief true. Intention-beliefs will therefore function as intentions due to this self-fulfilling tendency which, according to Velleman, runs so deep in human nature. The beliefs that are one's intentions are in other words 'self-fulfilling expectations that are motivated by a desire for their fulfilment and that represent themselves as such' (Velleman 2007, p. 109).[18]

[18] As Velleman points out, this mechanism or cognitive drive is not his own invention. Attribution theorists in social psychology invoke it to explain how stirred feelings can make us react and behave according to our interpretation of these feelings *thereby* verifying the interpretation of these feelings. For instance Donald G. Dutton and Arthur P. Aron reported that male subjects, who were approached by an attractive female interviewer on a long, wobbly footbridge over a deep canyon, showed signs of greater attraction than subjects who were approached by the same interviewer on a solid wooden bridge further upstream (1974). Their explanation was that the subject interpreted their anxiety and stirred feelings from the situation as owing to the female interviewer and thus took themselves to be more attracted to her than the subjects in the comparison group did, thereby determining their reactions and future behaviour. This drive for cognitive order and consistency may support the view that something similar is going on with respect to the self-attributions (e.g. the beliefs that one will Φ) that

For Velleman, intention-beliefs are therefore like ordinary beliefs in their *direction of fit* in that they represent their content as being true, thus displaying a from *world-to-mind* directedness.[19] On Velleman's story, the belief (or cognitive state) that *just is* one's intention to Φ moves one to Φ because of this tendency, drive, or concern in practical thought that takes the agent towards greater self-understanding. The picture of human agency is thus of the agent forming the belief that she will Φ – say, as a result of a piece of theoretical-cum-practical reasoning – and is moved to Φ by a tendency in the human psychological machinery that moves one to generate greater self-understanding. Velleman's story thus trades on the idea that an agent may achieve greater self-understanding – viz. more true beliefs about oneself – in one of two ways: either her beliefs are changed to correspond better with the object of understanding, or the object may change so as to fit one's beliefs. In this respect, it will be true to say that 'your efforts to understand your conduct are aided by your reciprocal efforts to make it intelligible. You don't just try to make sense *of* yourself; you also try to make sense *to* yourself' (2007, p. 8). As Velleman argues, this is something children learn early on:

> *During its second year, the child acquires a conception of itself as a cognizable object, a thing to be understood. And then it comes to see that understanding this particular thing is quite different from understanding any of the others. Other things must be understood as they are, however they may be. But the cognizable object consisting in the inquirer himself is more cooperative, since it is disposed to instantiate what the inquirer* already *understands: it is so disposed because it consists in the inquirer himself, with his drive towards self-understanding. The inquirer learns that he can make sense* of *himself by making sense* to *himself—that is, by dong what makes sense to him*.[20]

---

the cognitivist takes to animate human agency; that is, it may explain why one tends to do what one believes one will do (Velleman 2009, pp. 37–9). Also, see Velleman's review of the research literature (2000a).

Note that Velleman sometimes talks about the drive towards increasing self-knowledge as the object of an agent's desire (or concern) (1992b, pp. 140–1). At others he talks about it as a constitutive aim of action (2000b, p. 26), a self-fulfilling expectation (2000b, p. 24), or a drive (2009, p. 17). The preferable way to think of it is as a sub-agential aim or tendency in the agent akin to 'the sub-agential aim of knowing what we're saying, which inhibits us from speaking until we know what we're going to say' (2000b, p. 21). An advantage of this is that one avoids saddling an agent with too many concepts of the mechanisms and items involved in the process of acting.

[19] The metaphorical term 'direction of fit' was coined in contemporary philosophy by John Austin to indicate different ways in which world and *words* (or indicative sentences) may be related (Austin 1953). See Lloyd Humberstone (1992). It has later been associated with Anscombe's discussion of two radically different ways one might treat a list of shopping items: either as reporting what a shopper has been buying, or as instructions as to what should be bought. If there were a mismatch between the items on the list and the items bought, the nature of the mistake would depend on how the shopping list should be treated: as instructions or as a report. Largely inspired by Aquinas's two notions of truth – viz. as either 'the equation of the mind to the thing', or else 'the equation of the thing to the mind' (*ST* 1.2.1a.2) – Anscombe used this story to illustrate the difference between a mistake in judgment and a mistake in performance (1957, p. 56). Also see Searle (1983, pp. 7ff). I will use the term 'direction of fit' as referring to the two different ways in which attitudes can relate its propositions – or content – to the world. *Cognitive* attitudes entertain propositions grasped as patterned after the world and thus display a from world-to-mind direction of fit. *Conative* attitudes, on the other hand, grasp their propositions as a pattern for the world to follow thereby displaying a mind-to-world direction of fit. The content of a conative attitude is considered as a *faciendum* – or something to *be brought about* or *made* true – rather than as a *fact* or as something that *is* true. The latter is, of course, how the content is represented by a cognitive attitude. For a similar treatment of the notion 'direction of fit', see Velleman (1992a, p. 105).

[20] Velleman (2009, p. 17).

Although intention-beliefs are like other beliefs in representing things as they are arranged, they differ from ordinary beliefs by having what he calls a different *direction of guidance*. An attitude's direction of guidance is its causal order with respect to what they represent – viz. 'whether the attitude causes or is caused by what it represents' (200b, pp. 24–5). Since intention-beliefs bring about what they represent their direction of guidance is practical. This has consequences for the way intention-beliefs connect to the facts they are knowledge of:

> *Both the belief and the intention embody knowledge because they are true and appropriately connected to the facts* (...)*. The difference lies in the causal order of the connection. A belief amounts to knowledge if it is appropriately* caused by *facts, or evidence of facts, that make it true; an intention amounts to knowledge if it appropriately* causes *such facts*.[21]

An intention thus verifies its representation as to how things *are* arranged by bringing about this arrangement; or so Velleman thinks. So choice, he therefore concludes,

> *...has the same direction of fit as* [ordinary] *belief but the same direction of guidance as desire: it is a case of practical cognition.* (…) *The full description of this state is that it represents as true that we are going to do something; that it aims therein to represent something as really true; and that it causes the truth of what it represents.*[22]

One difference between Velleman and Setiya is that Velleman takes these psychological mechanisms as evidence for a constitutive requirement on acting for reasons: namely, that the agent must choose her action in light of some intelligible rationale for so doing. Velleman thus signs up to a doctrine akin to the *Guise of the Good*. The *Guise of the Good*, as we recall from chapter 1, is the requirement that agents must act for what they take to be *good* reasons for Φ-ing in order to be Φ-ing intentionally or for reasons. Velleman thinks something similar holds: namely, that an agent must act in a way that *makes sense* to her. This is not directly a normative or evaluative claim, however, since one may make sense to oneself by doing things that one fails to see as valuable or normative.[23] Velleman's requirement is rather that the action should make sense in folk-psychological terms since what is most 'explicable in folk-psychological terms' given one's character, attributes, and attitudes under the circumstances (2009, p. 27). Call this the *Guise of Making Sense*:

[21] Velleman (2007, p. xiv).
[22] Velleman (2000b, p. 25).
[23] For an argument against the normative or evaluative conception of human agency on these grounds, see Velleman (1992a).

> *Guise of Making Sense*: If *A* is Φ-ing for a reason (or intentionally), her Φ-ing *must* make sense from *A*'s point of view.

Velleman's reason for invoking this constitutive principle is the aforementioned tendency in human nature to increase one's self-understanding. 'Trying to understand yourself is inescapable,' as he says (2009, p. 16), and that tendency butts in to make sure that when agents act for reasons – or when they act *par excellence*, as he would put it[24] – they act for considerations in light of which the action makes sense. The hope is that his model can account for an agent's spontaneous self-awareness and self-understanding in action by transforming 'a rudimentary model of human motivation into a sophisticated model of autonomous agency merely by positing desires for reflective knowledge' (2007, pp. 241–2).

We should note, however, that Velleman harbours a distinctively rationalist ambition of deriving a theory of normative reasons from his full-blown cognitivist model of agency. In particular, he wants to account for good normative reasons as something that 'weigh in favor of an action (…) insofar as they contribute to an overall understanding of the action, given how the agent conceives of himself and his situation' (2009, p. 19). From that, he wants to argue that one can ground a theory of value where 'the considerations that make something valuable' just are 'considerations in light of which valuing it makes sense' (2009, p. 40). Velleman therefore seeks to ground value in the nature of human agency in a way that is strikingly similar to how ethical rationalists seek to ground morality in rational agency; the only difference is that Velleman's ground is supposed to be cognitivist through-and-through.

An advantage of modelling practical reasoning on principles of theoretical reasoning is that we seem to know more about what theoretical reasoning is (what its rational standards are, what it takes as a piece of reasoning to support its conclusion, and so on). A major disadvantage of this view is that it generates what Bratman has called 'the problem of promiscuity' (1991, pp. 261–2). Our deep-seated desire for self-knowledge and self-understanding will provide us with a motive to conform to our own expectation even if the belief in question is not an intention. If I believe I am going to trip, spill my coffee,

[24] Velleman has coined several phrases for the group of actions he is interested in explaining. In 'What Happens when Someone Acts' he describes them as full-blooded actions or actions *par excellence*, and argues that the standard belief-desire model fails to capture what is essential about them (1992b, p. 124). Later he makes a similar contrast between actions and *mere* activities (2000b, pp. 2, 2n3). Velleman explicitly denies in *Practical Reflection* that the class of actions he wants to capture is to be equated with Anscombe's class of intentional actions (i.e. actions to which the reason-requesting sense of the question 'Why?' applies). In his view, 'the category of intentional action is not a natural kind of behavior, because its boundaries are determined in part by norms of moral responsibility' (2007, p. xv n7). Velleman also cites the kind of trouble that cases like Davidson's Carbon Copies (see chapter 2, §3) make for Anscombe's claim that intentional actions must be subject to knowledge without observation (2007, p. xxiv n20). I have already argued that we can deal with the latter (chapter 2, §3).The former problem seems to turn on the claim that the concept of intentional action is sensitive to normative concerns that are irrelevant for action theory. I am not sure whether this is true even for the ordinary English idiom 'intentional action'; however, even if this were true, it would not follow that the way Anscombe employs this term in *Intention* is criticisable on these grounds.

mispronounce Setiya's name at my dissertation defence, and so on, Velleman's drive will provide me with a motivation for making these beliefs true; but that sounds crazy. Another disadvantage is that it is not clear that assimilating theoretical reasoning with practical reasoning can be defended. As Davidson says, 'the strongest argument against identifying [intentions] with [beliefs]' is, after all, that 'reasons for intending to do something are in general quite different from reasons for believing one will do it' (1978, p. 95). It is not clear how Velleman can account for this difference. A final disadvantage is that the postulated drive towards self-understanding and the corresponding *Guise of Making Sense* does not seem to hold for rational agency in general. In Velleman's discussion, at least, there are no arguments to show that this empirically confirmed tendency of human psychology pertains to the nature of rational agency as such. Such a psychological drive may, for all we know, reflect little more than an idiosyncrasy of our way of being agents in the world.[25] Velleman therefore seems to be in too much of a hurry backtracking to orthodoxy at this point:

> *Velleman's discussion is perplexing. He argues convincingly that someone can act for reasons without seeing anything good in what he is doing. But he insists that "I cannot act for reasons if I don't care about doing what's justified or (as I would prefer to put it) what makes sense". I don't see why not. According to Velleman, in acting from despair, "I am determined never to do a good or desirable or positive thing again." But surely I am also indifferent to whether my actions are justified, and I could not care less whether what I am doing makes sense.*[26]

This is not a knock-out consideration: however, *if* one were to have a cognitivist model of human agency that was more modest – viz. without invoking a *Guise of Making Sense* and without needing to model practical reasoning on theoretical reasoning – it would surely be preferable. With this in mind let us therefore turn to Setiya's model.[27]

Setiya's idea is to cash out the nature of the intention-belief in terms of an agent taking something as her reason for Φ-ing. He therefore begins with what he calls a 'circular account' of the attitude an agent must have to her reasons when they are *her reasons* for Φing:

> *Circular*: To take *R* as one's reason for Φing is to have the desire-like belief that one is Φing for the reason that *R*.[28]

[25] For similar arguments, see Setiya (2011).
[26] Setiya (2007b, p. 37n27).
[27] Another reason to put Velleman's model aside is that it is just as vulnerable as Setiya's model to the problem I discuss in chapter 4 below.
[28] Setiya (2007b, p. 42).

The problem with *Circular* – in addition to its obvious circularity – is that it does not respect the sense in which what we can have as our reasons is constrained by what we think. For instance, I cannot squeeze my partner's hand on the ground that it would turn her into a broccoli, as long as I know perfectly well that the world does not work that way; but on the *Circular* account, this can happen. Since believing something is a non-factive attitude – viz. its content may very well be false – *A* may act for the reason *R* (and thus believe that she is Φ-ing for reason *R*) without really being committed to the truth of *R*. After all, agents can erroneously believe that *R* is their reason for Φ-ing even though their lack of commitment to its truth precludes *R* from being their reason. Maybe their belief that Φ-ing for the sake of *R* is impossible, or that other things one takes oneself to know rule out *R*. Whatever the case it would then be possible to be wrong about what one believes is one's reason for Φ-ing. As Setiya notes, this problem goes away if the agent's attitude towards one's reason to Φ must be that *R* is a good reason for Φ-ing. A possible solution is therefore to invoke the *Guise of the Good* since that would commit the agent to the truth of *R* since good reasons for Φ-ing must be true (Setiya 2007b, p. 42n34). To avoid this, Setiya suggests the following *Simple Theory*:

> *Simple Theory*: To take *R* as one's reason for Φ-ing is to have the desire-like belief that one is Φ-ing because of the *belief* that *R* – where this is the 'because' of motivation.[29]

The idea behind *Simple Theory* is that taking something as one's reason is a matter of taking one's *belief* in some *R* to 'play a causal-motivational role in explaining one's action' (2007b, p. 43). That will allow us to get hold of two things: (1) from the truth of *Simple Theory* we can derive and explain the truth of *Belief*; (2) *Simple Theory* adds a cognitive constraint on taking something as one's reason by requiring that the agent believes that one is motivated by the belief that *R*. One cannot Φ because *R* – where *R* is one's reason for Φ-ing – without also being aware that one believes that *R*. As Setiya says:

> *Lear may still know that Cordelia loves him more than her sisters, but if he does not realize that he has that belief – he disavows it quite sincerely – its content is not a reason on which he can act. Such hidden beliefs may influence what we do, but not by figuring in what we* take *as our reasons to act.*[30]

[29] Setiya (2007b, p. 43).
[30] Setiya (2007b, p. 44).

Simple Theory can thus explain in what sense the agent must take account of her intention to Φ when it moves her to Φ intentionally. It is this taking account of one's intention that we found was missing on the Standard Model.[31] What these agents lack is, Setiya thinks, a *second-order belief* about the belief that moves them to act when they Φ for a reason. Not having formed these second-order beliefs, these agents could not possibly act for reasons.

On *Simple Theory*, the presence of the second-order belief really makes the crucial difference. It is necessary to have it in order to act intentionally (or for a reason); without it one would perhaps Φ because of some belief or other, but that would only amount to merely motivated behaviour. However, as Setiya also takes it to be *sufficient* it has the curious consequence that one could fail to have the relevant first-order beliefs and yet still act for a reason. Let me briefly illustrate: suppose I have difficulty acknowledging my lapse of faith in the existence of God. Perhaps I have been brought up a good Catholic, or raised by devoted Jewish parents whom I love and respect more than anything, so that it is personally hard for me to accept loss of religious faith; as Setiya says, 'I won't admit to myself that I do not have the relevant belief' (2007b, p. 44). Suppose this materializes as the erroneous second-order belief that I do believe in the existence of God. After all, a belief in any order, *n*, is still a non-factive attitude. I may therefore wrongly attribute myself with this first-order belief. Suppose further that the second-order belief moves me to pray on the ground that God is listening. Being wrong about my first-order beliefs will then make it the case that I am not moved to pray in the way I think I am; after all, I lack the belief the content of which is my reason for praying. However, it still seems true that I act for a reason in so praying. The result of the *Simple Theory* is thus that one can act for a reason whose truth one does not believe *if* one believes (erroneously) that one believes it.[32]

A final and somewhat tortuous complication is also needed since one may act by one's belief that *R* without thereby acting for the reason that *R*. As we recall, we can shed tears due to our belief that nothing less than crying will make our audience sympathetic to us without thereby shedding tears for that reason. To take oneself to be moved to Φ by one's belief that *R* is therefore not sufficient for taking *R* as one's reason to Φ; after all, the belief that *R* might move one as a matter of *mere* psychological motivation. The content of the second-order belief in *Simple Theory* depicts agents being motivated by a belief that *R*. But this is not

[31] See chapter 1, §6.

[32] Such cases are at best marginal. It is, for instance, not easy to imagine a situation where I have formed the erroneous second-order belief that I believe my partner will turn into a broccoli if I squeeze her hand. However, if I somehow or other get saddled with such a second-order belief, I could be moved to act by it so that the content of my misattributed first-order belief (viz. that my partner will turn into a broccoli if I squeeze her hand) is my reason for squeezing her hand. The reason why these cases are marginal is that forming second-order beliefs is liable to cognitive constraints of their own so that they cannot just fall out of the sky. Another thing is, as Setiya mentions, the fact that when one has a second-order belief it often produces the first-order belief that it is about (2007b, p. 44).

enough to guarantee that one is motivated for the reason that *R*. After all, any kind of mental act or occurrence could be susceptible to internality and externality (and thus be something that merely moves us to act in a detached, psychological and merely motivating way). Can we get around this problem without making the theory circular? Setiya believes we can by treating the content of the relevant second-order belief as self-referential:

> *[T]he attitude of taking* p *as my reason to act must present* itself *as part of what motivates my action. The content of taking-as-one's-reason is thus* self-referential: *in acting because* p*, I take* p *to be a consideration belief in which motivates me to Φ because I so take it*.[33]

If the content of the attitude is self-referential, it is possible to depict the agent as acting for a reason by depicting the agent as being motivated in part by itself.[34] That is, one may depict the agent as being motivated in part by the fact that one takes something as one's reason since the content of this *taking as one's reason* refers back to the act or event of so taking. The idea is therefore that the agent forms a (desire-like) belief that she is *hereby* Φ-ing because she believes *p* (where, again, the 'because' is the 'because' of psychological motivation). In this way, *Self-Reference* should make it possible to get around the problem of circularity:

> *Self-Reference*: To take *R* as one's reason for Φ-ing is to have the desire-like belief that one is *hereby* Φ-ing because of one's belief that *R*.

Self-Reference is an account of the special kind of belief that constitutes an agent's intention to Φ (or one's taking some *R* as one's reason to Φ) and is thus the essential claim in Setiya's *Minimalist Cognitivist* account of acting intentionally. In *Self-Reference* the 'because' is the *because* of a causal motivation, which has the result that Setiya's Minimalism reduces acting for a reason to the motivational role of an intention, which in turn reduces to the special kind of belief that is accounted for by the right-hand side of *Self-Reference*. In this way, Setiya has an account that matches the above principles *C1* and *C2*. His account is thus meant to achieve two things: (i) derive the truth of *Belief* (viz. that one always believes *what* one is doing and *why* insofar as one acts intentionally) from the fact that such an intention must always be present in acting for a reason; and (ii) account for the notion of taking as one's reason. Its proposal is thus that when the description on the right-hand side is satisfied some agent Φ-es for the reason *R*. In case we also allow for actions done for strictly speaking *no* reason – as

[33] Setiya (2007b, p. 45).
[34] As Setiya points out (2007b, p. 45n37), similar claims have been made by other philosophers, such as Gilbert Harman (1976), Searle (1983, pp. 83–90) and Velleman (2007, pp. 88–90, 94–7 and 140–1).

some Anscombeans think – one could simply supplement *Self-Reference* by saying that agents 'may decide to act on a groundless impulse, or for no reason at all' (2007b, p. 62).

IF *SELF-REFERENCE* ADEQUATELY accounts for what it is to take something as one's reason – and so provides a basis for an account of intentional agency – Setiya's model will have an advantage over Velleman's in that it can do without psychological drives, deep-seated cognitive motives, or controversial constitutive principles that risk reflecting our own idiosyncrasies rather than some essential feature of rational agency. For a devoted Minimalist, it is simply false to claim, Setiya argues, that 'it belongs to rational agency, in the abstract, to be exercised under the guise of the good, so that the need for a positive evaluation of action can be derived from the bare idea of its being intentional or being done for reasons' (2010, p. 83); and a similar objection could be raised against Velleman. All one needs in order to act for a reason is a belief as it is portrayed by *Self-Reference*. An agent may be moved to Φ because she has this desire-like belief while judging her reason to favour, justify, or make sense of her Φ-ing; however, she need not in order to be Φ-ing for a reason. Minimalism is therefore minimal enough to cope with the kinds of cases we discussed in chapter 1 where its agent fails to take one's reason as a good reason to Φ since it allows us to 'say what is involved in acting for a reason without appeal to normative or evaluative thoughts' (Setiya 2007b, p. 60).

Setiya therefore thinks that the sufficiency of Minimalism makes it hard to support any version of Ethical Rationalism. Without essentialist claims of universal validity, such as the *Guise of the Good* or *-of Making Sense*, it will be difficult – if not impossible – to make more than a nominal distinction between the 'should' of practical reason and the 'should' associated with the virtues of character. Since the project of Rationalism depends on this distinction the sufficiency of Minimalism supports being provisionally sceptic about the prospects of such a project in ethics. Armed with Minimalism, Setiya will therefore have taken a large step towards the conclusion that we should discard Rationalism. In the rest of this chapter I shall discuss some worries about this view; however, I will leave it for the next chapter to draw any definite conclusions. There I will argue that Minimalism should be rejected.

§4 WORRIES

In this section I shall bring up some worries that pertain to Setiya's view. I shall start by discussing the principle of *Belief* before moving over to a worry that pertains directly to the nature of the intention-belief, as Setiya portrays it.

An immediate worry with *Belief* is dialectical. We recall that Setiya raised a counterexample to the claim that knowledge must be involved in intentional agency in order to move from a principle such as *AK* to rather endorse *Belief*.[35] His reason for so doing was that a knowledgeable state was too epistemically ambitious and that sometimes all one needs to act intentionally is a belief that one is so doing which has less epistemic warrant than knowledge. A crucial premise in Setiya's story is that knowledge of what one is doing would require conditional knowledge of what one is able to do if one is trying to do it. It was this knowledge that was lacking in the case of the paralysed agent and failure to have this conditional knowledge was allegedly enough to undermine the agent's knowledge of what she was doing when she tried to raise her arm. A problem with this premise is that even Setiya has come to reject it. In his replies to criticism by Paul, Setiya concedes that the requirement that agents must have knowledge of their ability to Φ in this simple conditional sense is false (2009a, p. 134). On reflection, Setiya thinks that knowledge-how provides its agent with sufficient epistemic warrant for having a belief about what she is doing even if she lacks knowledge of her ability to Φ. This is not the same as saying that the relevant belief becomes knowledgeable due to the presence of such knowledge-how – nor do I think there is reason to ascribe mere practical capacities with such an epistemic role – however, it does weaken his case against the truth of *AK*. At the very least, it gives us reason to ask whether an agent may nevertheless know what she is doing in this case despite reporting otherwise. Perhaps the exercise itself gives rise to knowledge of what she is doing. She may perhaps be an agent who is instantiating a counterexample to the so-called *KK* principle, or the claim that knowledge entails knowing that one knows. According to Williamson, we have reason to reject this principle anyway since '[o]ne can know *p* without being in a position to know that one knows *p*' (2000, p. 107). This might therefore be a case in which the agent knows what she is doing but nevertheless fails to know that she knows it. Perhaps it is one of those rare cases where the nature of the agent's conscious experience fails to be decisive, but where 'the nature of [the agent's] behavior' proves decisive (O'Shaughnessy 1963, p. 388). If so, Setiya's argument would fail to establish *Belief* as the preferable description of agent's knowledge.

A way to reject *Belief* would, of course, be if one could argue that it fails to give expression to a necessary truth. Some philosophers have tried this by coming up with counterexamples to *Belief*. As Paul points out, we do attribute intentions and give reason-giving explanations because we want to capture 'the fact that an episode of behavior had a

[35] See my discussion in §2 above.

goal-directed, means-end structure and exhibited responsiveness to reasons' (2009a, p. 5). It is not clear that this will need to involve beliefs. Another way to go is by questioning Setiya's requirement that what one must do in the belief that one is doing it is that the 'it' cannot be something one is 'merely trying to do' (2007b, pp. 26 and 41). Consider the following:

> PARALYSIS*:* Sue's right arm has been paralyzed for awhile but there is some reason to believe that the paralysis might subside at any time. Suppose Sue makes an effort to raise her arm. If she succeeds, Sue will arguably have raised it intentionally. But Sue is cautious and not irrationally optimistic. Given her doubts she just cannot make herself believe that she is raising it. Yet, things turn out very well indeed. She manages to raise her arm, though without believing that she is so doing. Perhaps all she sensibly believed about what she was doing while raising her arm was that she was trying to raise her arm.[36]

Paralysis might be a counterexample to *Belief* since Sue is able to intentionally raise her arm without having a belief about doing it, or without the belief that she is doing anything else in order to raise her arm. Raising one's arm is arguably a basic intentional action in normally functioning agents. An agent recovering from paralysis is hoping to restore this ability. That is, she hopes to be able to directly raise her arm, i.e. without having to do something else, such as pulling a cord or starting a hoist, in order to raise it. Sue does therefore not believe that she is doing something else as a means of raising her arm. The case is problematic then since Sue, given her doubts, is unable to form the belief that she is raising her arm. '[P]erhaps [the agent] cannot feel it, or see it,' Setiya remarks, 'and [she is] not at all sure of [her] ability' (2008b, p. 390). If the only thing she believes is that she is trying to do it – or, alternatively, that she intends to do it – there will be something wrong with *Belief*.

Setiya has two replies to these two problems: first he argues that the basic question he wants to ask with a principle like *Belief* still stands even if we have to revise *Belief* to make room for trying to Φ or intending to Φ.[37] So even if *Paralysis* was a genuine counterexample to *Belief*, there is a way to revise it so that the ensuing principle could still be used to ask the questions that Setiya believes will support Minimalism. Setiya's second reply defends the accuracy of *Belief*. A case such as *Paralysis*, he says, is not so much a counterexample as a reminder that we are operating with an innocent simplification in philosophy of action and epistemology that ignores 'the fact that belief comes by degree' (2008b, p. 391). We are

[36] Similar stories are told by Bratman (2009, p. 515) and Setiya (2008b, p. 390).
[37] See Setiya (2007b, p. 27n9; 2009b, p. 531n1). The point with this reply is that even if one takes this case to support trying to Φ or other mental acts as basic intentional actions (or sufficient for performing an intentional action) the puzzle remains: why do we have to have beliefs about trying to Φ in order to be intentionally Φing? For an argument in favour of tryings as intentional actions, see Hugh McCann (1998b, pp. 94–109) and O.R. Jones (1983).

creatures who do not simply believe or disbelieve something; rather, we move in the range between these extremes and have more or less confidence in something, give more or less credence to a claim, and are having degrees of belief. In order to express something like *Belief* in this richer and more adequate psychological language, we could qualify it so as to accommodate partial beliefs. One could say that if one is Φ-ing intentionally (where Φ-ing is a basic intentional action), one must at least be *more confident* that one is Φ-ing than one would be if one did not Φ intentionally. Setiya therefore suggests the following:

> *Belief-or-Confidence:* If *A* Φ-es intentionally, then *A must* either believe (outright) that she is Φing or she is more confident of this than she would otherwise be, or else she Φ-es by doing other things (that is not merely trying to Φ) for which that condition holds.[38]

With *Belief-or-Confidence* the problem of *Paralysis* goes away. Even though Sue does not form the belief that she is raising her arm, she is arguably more confident of raising her arm than she would have been had she not done it intentionally. Setiya remarks that the only way of having a counterexample to this is if it was possible for agents to Φ intentionally while being quite confident that they are neither Φ-ing nor taking means to Φ. Such a lack of insight would be simply incredible, Setiya notes. Incredible or not, there are philosophers who believe there are such counterexamples. Paul argues that there might be cases where an agent loses some of her confidence because she is Φ-ing intentionally. Consider the following case:

> STRANGELOVE: Dr Strangelove suffers from Anarchic (or Alien) Hand Syndrome so his one and only hand is often acting with a mind (or will) of its own. Without the agent intending it the hand frequently and capably carries out an action, such as buttoning up the doctor's shirt. Suppose that Dr Strangelove desires to unbutton his shirt and that he knows that this will likely happen as a result of his disorder without him needing to interfere. However, suppose also that his lack of control infuriates him and that he would greatly prefer it if his hand unbuttoned his shirt as a result of his intention. He knows from experience, however, that when he intends to act with this hand the intention is sometimes effective, but much less reliable than the hand's tendency to unbutton shirts of its own accord. Further, if the intention is not effective, the hand will do nothing. In light of all this, Dr Strangelove's confidence that he will button his shirt – or do anything toward his aim – should decrease as a result of forming the intention. Nevertheless, if he succeeds in his effort, he will have buttoned up his shirt intentionally.[39]

[38] Setiya (2008b, p. 391).
[39] See Paul (2009b, p. 553). Setiya tells a similar story (2009a, p. 130). The rare disorder we are talking about here combines ownership to one's limbs with a 'loss of agency' with regard to the limb's behaviour. The limb is capable of executing often complicated behaviour – such as unbuttoning one's buttons, dressing off one's cloths, or manipulating tools – while the agent experiences a 'loss of agency' and a lack of ownership towards the resulting behaviour. The agent might often fail to know what the limb is doing and may frequently become aware of

Or consider the case of someone suffering from *Foreign Accent Syndrome*:

> FOREIGN ACCENT: Kate had a brain haemorrhage followed by a stroke, and as a result suffers from Foreign Accent Syndrome. Although monolingual and English, the disorder has the curious effect on Kate that she may often burst out in speech that has sometimes a French lilt and at other times a clear Scottish cadence. She could of course try to speak with such accents; but she is much less confident in her ability to impersonate these accents than she is of the reliability of her disorder to produce them now and then.

The problem with these cases for Setiya is that they portray behaviours an agent may try to do intentionally only at the cost of her confidence in successfully performing them without the intention. They therefore appear to be counterexamples to the claim that acting intentionally generates, if not an outright belief, so at least increased confidence in what one is doing.

In reply, Setiya calls attention to the fact that the counterfactual in *Belief-or-Confidence* must be handled carefully. The requirement is that the agent is more confident *than s/he would otherwise be*; but there are different ways one might take this claim. It is, for instance, not meant to cover cases where an agent tries to intentionally do what is usually automatic behaviour on their part. One may thereby suffer a decline in confidence:

> *Couldn't there be an action that is normally automatic but which can be done intentionally with a lower chance of success? If one is aware of this, one will be, on balance, less confident that one is performing that action when one is doing it intentionally – but still more confident than if one were not doing it intentionally and one's automatic system were shut down.*[40]

Breathing, for instance, is something one does automatically in this sense. One can of course decide to seize control of one's breathing reflexes and breathe intentionally while knowing all along that one's automatic system is far more reliable in producing steady breathing. I may thus become less confident that I am breathing steadily than if I gave up controlling this behaviour. But this is not a relevant alternative for the counterfactual. The relevant alternative to breathing intentionally is *not* breathing under the guidance of my automatic system; rather, the alternative is if one had no such intention *and* the automatic system was shut down. This is not the contrast invoked by the alleged counterexamples. Their comparison is structurally akin to cases where an omnipotent demon fills the role of the automatic system:

---

its behaviour because other people bring it to her attention. Anarchic (or Alien) Hand Syndrome has also earned the popular sobriquet Dr Strangelove syndrome, named after the eponymous scientist with a rather unruly limb in Stanley Kubrick's 1964 classic film.

[40] Setiya (2008b, p. 391n12).

> DEMON: Suppose that an omnipotent demon reveals to René that if he does not decide to meditate on the existence of God and the external world (in short: *to Φ*), the demon will use his powers to make sure that he will. René, who is by no means lacking in intelligence, then knows that if he decides to Φ there is always a chance that he might fail to do what he intends. However, given what the demon has revealed to him he knows that the demon will make sure that he Φ-es if he fails to intend it. The only chance of him failing to Φ is therefore if he intends to Φ. René is therefore more confident that he will Φ if he does not form the intention to Φ than if he does form the intention to Φ. If René still decides to Φ and he succeeds, Φing would be something he does intentionally despite less confidence than if he failed to intend it.

What the alleged counterexample capitalizes on is thus the reliable work of an agent external to one's intentional agency, work on which one can rely even in the absence of an intention to Φ. According to Setiya, however, *Belief-or-Confidence* is not in conflict with that since it claims only that an agent's must have increased confidence of what she is doing by virtue of acting intentionally. What it claims is that it is impossible for agents to not become more confident that they are Φ-ing as a result of Φ-ing intentionally, not that confidence of one's own intentional capacities will always outrun confidence in capacities that are external to one's intentional agency. So Paul's counterexamples will clearly not do.

Of course, if one accepts this revision of *Belief*, we will need to consider whether Minimalism finds support in our need to explain the truth of *Belief-or-Confidence*. Setiya believes that this can be done since *Belief-or-Confidence* still makes it a puzzle why there is a 'need for *cognition* of one's action, in acting for reasons, however mediated it may be' (2007b, p. 27n9). According to Paul, however, a reason for believing that 'this concession has a much more significant impact for our theory of intention than Setiya acknowledges' is that talk about partial beliefs and degrees of confidence rather than outright beliefs blocks the identification (or reduction) of intentions with beliefs (2009b, p. 550). That identity will risk predicting agents who only somewhat intend to Φ because they only have a partial belief. Since it does not seem that 'folk psychology admits of intentions coming in degrees' – at least an intention to Φ does not seem to be 'weaker for the fact that [its agent] is only somewhat confident that it will happen' – Paul concludes that *Belief-or-Confidence* fails to motivate Setiya's Minimalist Cognitvism (Paul 2009b, pp. 550–1).[41] Setiya agrees, of course, that '"partial intention" is not a good phrase for the phenomenon I have in mind' (2009a, pp. 130–

[41] An argument for the existence of partial intentions can be found in Richard Holton (2008). Holton's idea is not the same as the one that Paul objects to here, however. For Holton an intention is partial when the agent has competing plans for accomplishing the same end (2008, p. 41). Paul does not object to the idea that intentions come in degree with respect to the strength of one's practical commitment; rather, she objects to the idea that intentions vary with the degree of confidence in what one is doing. See Paul (2009, p. 551n13).

1); but, rather than giving up the account, he suggests that we change it so as to allow that the degree of commitment in intention does not match the degree of doxastic confidence:

> *Self-Reference**: To take *p* as one's reason for Φing is to have the desire-like belief (or confidence) that one is *hereby* Φ-ing because of one's belief (or confidence) that *P*.

I will take it that Minimalism is revisable to accommodate *Belief-or-Confidence* and thus that the truth of this principle – viz. if one accepts it – will be a point in favour of Minimalism since its way of accounting for intentional agency allows us to derive this principle. Let me now briefly argue that even the revised account has its problems.

What one may ask is if *Belief-or-Confidence* is guilty of an anthropocentric bias with its focus on cognitive attitudes. As Keith Hossack points out, it might well be that 'not every knower is a believer' (2007, p. 267). An omniscient knower will never need to make judgements since it already knows everything. Possessing knowledge is thus conceptually distinct from having such a psychology. A problem for proponents of *Belief-or-Confidence*, then, is that they would have to reject that such a knower would be able to act intentionally given its lack of belief-like cognitive attitudes.[42] *Belief-or-Confidence* requires a certain sort of attitudes that may well be central to our mental life. However, this may reflect nothing but a natural-historical necessity of human life; and, as Setiya points out on a different occasion, we must take care not to make the mistake of 'thinking that our characteristic form of agency shows us what agency essentially is' (2010, p. 104). The possibility of omniscient intentional agents without a belief psychology will then be a counterexample to *Belief-or-Confidence*.

Although I agree with Velleman in thinking that the philosophy of action should, in principle, 'set about describing possible agents, without concern for whether they are real' (2007, p. xxxi), I am afraid that the possibility of an omniscient agent may not convince everyone that *Belief-and-Confidence* is unduly anthropocentric. However, my point does gain impetus when we realize that a related problem pertains to *Self-Reference* principles. The cognitive attitude portrayed by both versions of *Self-Reference* is reflexive in being about itself and its own explanation. In acting for reasons we must 'have beliefs about the psychological explanation of our action', as Setiya puts it (2007b, p. 47). This feature of Minimalism has the consequence that the intention-belief – whether the belief is partial or outright – will involve having a belief-like cognitive attitude about another belief or belief-like state: namely, about the attitude that causal-motivationally explains one's action. Such

[42] So doing will presumably not add or generate any *new* knowledge since this Being is already assumed to know everything there is to know.

belief-states, whether partial or outright, are propositional attitudes and thus require concept-possession of the concepts that go in to make up its content. Beliefs involving concepts such as *ELEPHANT* or *CHESS* can only be had by agents possessing these concepts. Following Christopher Peacocke, we can say that there are certain necessary and sufficient possession conditions that pertain to some concept *C* and that specify 'a judgemental reaction, which an arbitrary thinker must be capable of if he is to have the concept' (viz. the satisfaction of which determines whether a subject possesses *C*) (1992, pp. 14–5). It is not necessary, of course, for the subject to have the metalinguistic concepts that we need in order to formulate the relevant possession conditions.[43] There is, after all, an important sense in which possession of a given concept is relatively cheap.[44] There is also a distinction between having or possessing a concept and fully mastering it. I presumably have the concept *QUARK* in the sense of merely possessing, whereas full-fledged mastery would require some years of training in fundamental physics. But some dispositions are required in order to possess the concept. For some subject, *S*, to possess the concept of *CONJUNCTION*, for instance, it would be required that *S* is disposed to infer *A* from accepting *A & B*. Similarly, *S* must perhaps know or be prone to infer from accepting that some *x* is an elephant that *x* is not a verb, or that *x* is an animal, and so on.

In a similar vein, the intention-belief requires possession of the concepts involved in its content.[45] Having an intention-belief will, as a consequence of this, require possession of concepts such as *BELIEF*, *REASON*, *BECAUSE* (in the psychological-motivation sense), *CAUSE*, and so on. The problem is that it is not obvious that agents who are able to act intentionally are agents who also possess these concepts. To possess the concept of *BELIEF*, for instance, seems to require the ability to recognize that beliefs can be false, that there is a reality beyond what one believes, that there is a possible distinction between appearance and reality, and so on.[46] A thinker need not be capable of giving an explicit formulation of these things to possess the concept; rather, it is more plausible that a sort of tacit knowledge gets reflected in her behaviour and dispositions of reasoning. If those dispositions give us reason to attribute the subject with a propositional attitude whose content involves *BELIEF*, we will have definite reason to attribute possession of this concept to the subject.

The problem for Setiya is that one will find intentional agent candidates whose behaviour fails to reflect the relevant dispositions. A plausible interpretation of such

[43] Williamson makes a related point with respect to the concept of knowledge (2000, pp. 40–1).
[44] See Burge (1979; 1986; 1990).
[45] It is widely acknowledge that concept possession comes in degrees; nevertheless, philosophers have come up with different categories for characterizing levels of concept possession. For instance, George Bealer (1998; 2004) distinguishes between nominal and determinate possession, while Burge (1990) separates mere grasp from having sharp or full grasp of the concept. There are also distinctions between partial or incomplete understanding and misunderstanding a concept (Burge 1979; Bealer 1998; 2004).
[46] See Peacocke (1992, pp. 163–4).

observations is that the subjects do not possess the concept *BELIEF*. Small children or low-functioning autistic subjects may, for instance, lack this concept because they do not have a fully developed theory of mind – viz. they are not able to infer the full range of mental states that explain action and behaviour – that would allow them to recognize and attribute belief-states to themselves and others. As Simon Baron-Cohen argues, 'a large number of studies have demonstrated that children with autism have difficulties in shifting their perspective to judge what someone else might think, instead simply reporting what they themselves know' (2001). Furthermore, many developmental psychologists have suggested that children do not understand the appearance-reality distinction and do not recognize that beliefs can be false until the children are at least three years old. Again, these subjects may be interpreted as lacking the concept of *BELIEF*. They may have a capacity for forming normal, first-order beliefs; however, failing to possess this concept will prevent them from forming beliefs about beliefs.[47] Consequently, they cannot form intention-beliefs. The minimalist must therefore either reject these empirical findings, or bite the bullet and say that these subjects cannot act for reasons.[48] That is a definite drawback and it counts against the Minimalist theory's reliance on higher-order cognitive attitudes.

§5 CONCLUSION

What one may take these problems to indicate, is that the cognitivist is on the wrong track when he pursues a strategy where belief-like cognitive attitudes or states play an essential role in the nature of acting intentionally. Both our problems may in fact be taken in support of the idea that knowledge is a more basic and essential state shared by agents with a mind and capable of acting intentionally or for reasons. I think that is an important truth; it is, however, a topic to which I will have to return below.[49] In the next chapter I want rather to argue that belief-like cognitive attitudes, as portrayed by *Self-Reference**, present the cognitivist with an epistemic problem: namely, the *Doxastic Venture Problem.*

[47] See Josef Perner (1991), Wellman, Cross and Watson (2001). Note that this is a weaker claim than what Davidson argued (1970b; 1980; 1982b, p. 102; 2001d) when he emphasized the connection between mastering the concept of a belief and having a belief-system (Davidson also emphasized the connection between having a belief-system and having a language). Davidson offers little support for this claim and it is not evident that this is the case (viz. the data from developmental psychologists). After all, why should not beliefs be like bad temper, which one can have without having the concept of it? Beliefs may for instance have a functional essence so that states occupying the functional role characteristic of beliefs in the cognitive system of some *S* qualify as beliefs regardless of whether *S* has a concept for them. At any rate, the data of developmental psychologists shows that the explanatory burden is on those who hold that beliefs require the concept of beliefs. For the idea that beliefs has a functional essence (without thereby being functionally defined), see Jerry Fodor (1998, p. 8). I also benefit from Eric Schwitzgebel's discussion (2010).

[48] Similar considerations apply to other candidate cognitive attitudes that the cognitivist operates with.

[49] See chapters 6 and 7.

4

The Doxastic Venture Problem

Your thoughts don't follow your hand at all; they lead it.

David Velleman[1]

§1 INTRODUCTION

Imagine you're having an evening off and are making up your mind how to spend it. You try to choose, form, act on and carry out an intention to do something. As you deliberate, you may have a range of alternatives in front of you that are all live options, such as inviting your soul mate to the opera, visiting your parents, running a half-marathon, watching Bergman's *Persona* in splendid solitude, or studying the last chapter of *Ethics and Limits of Philosophy*. Alternatively, you might be considering one of these options pondering about whether to go for it. If so, you are engaged in a process of practical thought, or thinking with a view to action. You weigh options, compare relative pros and cons, and do what is required to engage in a process whose end is to do something. However, at this particular moment your options are still open. The action is waiting for your decision to fall in order to get going.

Suppose you decide to study. Now ask yourself: on your way to willing that, do you need to have reasons for believing that you are in fact going to read Williams's book? 'Certainly not!' you might say. Of course, agents may have reasons for believing that she will Φ before intentionally Φ-ing. Perhaps one is the sort of person who leads a humdrum life of self-discipline and routine while possessing a high level of self-knowledge so that reading ethics is what one would expect given one's body of evidence. A process of deliberation may perhaps need to go on as if every option were epistemically open despite being in a favourable epistemic position with respect to the outcome of the deliberation.[2] However, nothing in

[1] Velleman (2007, p. 18).

[2] As Paul notes, there is a sense in which '[i]ntentions (...) are formed precisely when there is *not* sufficient evidence that one will perform the proposed action' and that 'if one did have sufficient evidence, the intention would be redundant' (2009b, p. 546). I think this is too strong. An agent may, after all, fail to know what her evidence is, as Williamson argues (2000, ch. 9). She may then have sufficient evidence for the proposition that she will Φ without knowing that she has it and it would be odd to deny that she could deliberate about Φ-ing. I do think human agents have some tendency to bring practical thinking to a halt if they have – and are aware of having – sufficient evidence for

principle seems to preclude agents from having sufficient prior evidence for what they will do. Whether we *must* is a different bag altogether. Even the most boring life seems to involve situations where an agent does something that goes beyond, or even against, what it would be reasonable to expect on her evidence. Agency often leads to unpredictable twists and turns. You may fail to know how you will respond in a novel situation, or the evidence may be misleading, and so on. At best, it seems to be a loose connection between what one decides to do and what one has evidence for thinking that one will do. Circumstances where one is about to decide to Φ need not be evidentially favourable for believing that one will so Φ.

In what follows, I will argue that lack of evidential or epistemic warrant with respect to the content of what one intends and what one does intentionally is a serious challenge for Cognitivism about intention. According to Cognitivism, an intention in action just is – or at least constitutively involves – an agent's belief that she is so Φ-ing (or, if it is a prior or future-directed intention, that she will be Φ-ing) (Setiya 2009a, p. 129; 2008b, p. 394).[3] As some philosophers have noted, however, a curious consequence of this view seems to be that it portrays intentions as a matter of faith, or reckless and wishful thinking.[4] After all, if an intention is such an intention-belief agents will often need to form beliefs when it is epistemically unreasonable and doxastically venturesome to believe their contents. From a naive and pre-theoretical perspective, however, intentions seem to be something agents can – indeed should – form with both ease and little regard for epistemic standards. Agents seem to be very unlike spectators in this respect, as Ryle pointed out, given that the former is originating where the latter is only contemplating (1949, p. 54). Of course, actions stand answerable to standards of their own; but whatever those standards are, having epistemic ground for thinking that one will Φ seems not to be required. The agent is free to act beyond, and even against, what she has or takes to be reasons for believing that she will do. In

what they will do. There is thus some truth to Harman's idea that there is a rational prohibition on predicting decisions one can make instead (1976, p. 455); but the link between what one has evidence for and what one can decide to do is not as straightforward as Paul thinks.

[3] I will treat the content of an intention in action as what one expresses when one states what one is doing. I will therefore follow Anscombe in taking the content of an intention to be 'of a type to be formally the description of an executed intention', such as greeting and promising (1957, p. 87). On this picture, an intention in action will have the form: 'I am Φing (or doing such-and-such)'. We may, if the intention is prior and future-directed, say that the content is 'I will Φ (or be doing such-and-such)'. One advantage of this view is that it will be easier to align the action with what happens (and knowledge of what one does with knowledge of what happens). If intentions relate to contents that describe what one does or will do in this way, it seems easier to respect Anscombe's slogan that 'I do what happens.' (1957, pp. 52–3) For further defence of this view, see McDowell (2010). An alternative is to treat the content of an intention as a verdict, or all-out judgement (e.g. 'Φing is all-out desirable') (Davidson 1978), or as *volition* (e.g. 'I shall Φ now') (Sellars 1966; Brandom 1994). A problem with these models is to explain why intentional actions, at least in core cases, give rise to agent knowledge of what one does since the content of the intention – which is a plausible vehicle for agent knowledge – does not match up with what one does.

Note that I do not think that prior intentions are required in order to act intentionally. One may directly do something intentionally while having an intention in action. For further discussion of the distinction between prior intentions and intentions in action, see Searle (1983, ch. 3). I will not take a stand here on whether a prior intention produces another state – viz. an intention in action – when its time comes, as Searle thinks, or whether the distinction are two aspects of the same state, as McDowell argues in his Howison Lectures (2008b).

[4] See Grice (1971), Rae Langton (2004), Setiya (2008b; 2009a; 2011) and Paul (2009a; 2009b).

particular, we seem to be no less rational for intentionally doing something that would go against epistemic reasonability if that something were the object of our beliefs.

The problem for cognitivists is that they will either have to accept this as an awkward consequence of their view that agents often are epistemically blameworthy when they act intentionally, or they must find means for saying that agents forming intention-beliefs are in their right in to do so. Below I shall argue that Cognitivism fails to find such means and that this gives us reason to reject their view. First, I shall introduce the doxastic venture problem before discussing the nature of doxastic venturesomeness in more detail (§2). I proceed to consider and reject the idea that there is a notion of evidence that cognitivists can turn to in order to show that the intention-beliefs – despite appearances to the contrary – have sufficient evidential support (§3). Finally, I shall discuss and dispute Setiya's suggestion that intention-beliefs are exempted from standard evidential requirements of belief-formation and thus epistemically warranted in a non-evidential way (§4 to §6). This failure of Cognitivism is no minor event: it also denies Setiya his crucial lemma in his argument against ethical rationalism. I shall briefly comment on this in the final section (§7).

§2 The DOXASTIC VENTURE PROBLEM

Agents like us are good at using our epistemic and active capacities to guide us through the buzzing confusion of changing circumstances and situations. As Williamson says, '[a]ction typically involves complex interaction with the environment; one needs continual feedback to bring it to a successful conclusion' (2000, p. 8). However, even if we are highly sensitive to our surroundings in this respect, we often decide to do things which – if we believed we were about to do them – fail to sit exactly with our body of evidence. We typically lack sufficient information for our evidence to tip us in favour of one string of action rather than others. 'Practical thought is [therefore] not,' Setiya says, 'a search for sufficient evidence from which to predict an action but for considerations on which (for all I know) I may or may not elect to act' (2008b, p. 396). On the contrary, agents choose their actions and form their intentions with little or no evidential concerns. 'I form the intention on the basis of practical thought, and the reasons for doing so are practical reasons,' as Setiya points out (2009a, p. 134). In this section, I will argue that insofar as intentions are a sort of belief – as cognitivists think – it appears as if agents acting intentionally often form intention-beliefs whose contents go beyond or against their body of evidence. I shall call this the *Doxastic Venture Problem* and I will argue that it presents a serious challenge for cognitivists.

The main idea behind the *Doxastic Venture Problem* is that agents acting intentionally will form and be guided by intentions whose contents often are doxastically venturous in that they fail to have epistemic warrant. I shall shortly try to be clearer about what these notions of warrant and doxastic venturesomeness involve; for now, however, let us just say that some content of one's intention – say that I am Φ-ing (or will be Φ-ing) – is something an agent may or may not have epistemic warrant or sufficient justifiers for believing. The claim is then that agents form intentions whose contents are conspicuously lacking in such a warrant. Now this is not the same as saying that an agent may be thoroughly ignorant when she acts, or is not sensitive to what she knows in deciding what to do. She clearly is. Our evidential situation may thus affect our choice of action. It may even determine, at least in part, the range of things one can do intentionally. To take a familiar example: as philosophers have noted, there are certain epistemic and modal constraints – which are hard to cash out with any precision – that restrict what can be the object of a genuine intention in action. Unless we are terribly disturbed, confused, or lack self-insight we are not able to set out to Φ intentionally if we think there is no way we can be Φ-ing. For instance, I cannot intend to fly to the moon and back by waving my arms, or try to turn my partner into a broccoli by squeezing her hand because I know very well that these are not ways for me to achieve these things. Also, as Annette Baier observes, our practical abilities may make up for a 'difference in the range of (...) things I can intend and the things I can take as a goal' since 'the proper objects of intending, unlike the proper objects of aiming at, seem limited to (...) things I can do' (1970, p. 649). If I do not know how to speak Spanish, I may not intend to do that.[5]

However, these kinds of influences on one's decision-making in action will not undermine an important experience of decision: namely, that we do not decide to Φ as a function of whether our evidence predicts that we will be Φ-ing. Regardless of whether psychological determinism is true or not an agent does not appear to resolve to Φ based on evidence supporting that she will Φ. Whatever it is, practical thinking is not – indeed cannot be – a search for sufficient evidence from which one may predict what one will be doing. Acting is not forecasting.[6] From a naive perspective, then, it seems clear that one may form an intention to Φ even in the face of evidence. If intentions are beliefs, however – as cognitivists claim – the acquisition and retention of such intention-beliefs seem to clash with standard rational requirements governing the acquisition and retention of ordinary beliefs. A platitude

[5] As Michael Dummett points out, I may not even know what it is to speak Spanish unless I also know this language (1991, p. 94).
[6] See O'Shaughnessy (1963) for an argument in favour of the idea that insofar as the agent is searching for evidence for whether she is Φ-ing the object of knowledge gets disrupted since the agent cannot be Φ-ing intentionally at the same time as she is observing herself.

says, for instance, that it is a requirement on rational thinkers that they respect their evidence and believe only what they have reason to or are epistemically entitled to. If such *strict evidentialism* is true – viz. the view that all normative reasons for believing something are evidential reasons and one is therefore epistemically entitled to believe all and only what one can reasonably believe on one's predictive evidence – agents forming intention-beliefs will seem liable to rational criticism.[7] This is what William K. Clifford expresses when he adamantly vetoes believing anything beyond what one has sufficient evidence for believing since it is 'wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence' (1877).[8] Clifford's absolute veto may sound harsh and excessive. However, it brings out an important feature about *normative requirements*: namely, that insofar as they apply to one's situation there is something wrong about violating them. Agents who fail to satisfy such normative requirements are, at least to some degree, rationally criticisable.[9] For evidentialists, for instance, this austere attitude is invoked by the following requirement:

> *Evidence Requirement* (ER)*:* Agents are required only to form beliefs that are not epistemically unreasonable on one's body of evidence.

If we accept *ER* as a requirement on belief-formation, it follows that agents will violate it and be blameworthy insofar as it applies to their situation and they come to believe some *p* that fails to be reasonable on their evidence. Whether a thinker is epistemically reasonable or unreasonable depends, of course, on how we specify her body of evidence and what a reasonable or unreasonable relation is. On the standard interpretation of *ER*, however, it is clear that agents forming intention-beliefs will often violate it. If both *ER* and Cognitivism about intention hold, the consequence seems to be that agents regularly violate a normative requirement that pertains to belief-formation. It thus appears as though having an intention were 'a case of licensed wishful thinking', as Grice said (1971, p. 268), since it violates the rules we usually abide when we manage and administer our beliefs.

In this respect, an agent *qua* being a venturesome intention-believer resembles the character in William James's famous *Climber* story. In *The Will to Believe*, James tells the story of an agent trapped on a cliff. His only means of escape is by making 'a terrible leap'

[7] For this notion of *strict evidentialism*, see Reisner (2007, p. 304; *forthcoming*).
[8] William James refers to what 'that delicious *enfant terrible* Clifford writes' thus setting him up as his opponent (James 1892, pp. 461–2).
[9] In general, a normative requirement, *N*, that requires X of some agent, *A*, in situations S_1, S_2,…,S_n, is either satisfied or violated by *A* in some S_x at some time, t_1, or else *N* fails to be a requirement on *A* at t_1 in S_x. There may be overriding reasons for failing to satisfy certain normative requirements but I take it that the requirement will fail to apply as a normative requirement on the agent in such-and-such cases. To the specific normative requirement there will typically be a range of qualifying clauses that specify when they do not apply to an agent under special circumstances. As we shall see below (§4), one gloss of the challenge against Cognitivism is as the task of specifying a principled qualifying clause to ordinary epistemic principles to capture and exempt the intention-beliefs.

across the abyss. Lacking experience the climber has no evidence for the proposition that he can make this leap; however, moved by hope and faith he nevertheless believes it is possible to make it, thereby nerving his feet to its accomplishment. According to James, this is a case where 'faith creates its own verification' (1896, pp. 528–9). If cognitivists are right in conceiving the nature of intention in action as a sort of belief, however, the story of the *Climber* will begin to resemble ordinary instances of human action. 'James' courageous leaper seems worlds away from Anscombe's shopper', Langton argues, 'but if the doxastic proposal [i.e. Cognitivism] is right there may be a connection between the leaper's faith and the shopper's intention' (2004, p. 245). This is because the content of a venturous intention-belief would be similarly related to the agent as the content of the climber's belief relate to him. For one thing, both agents lack evidence for the truth of these contents. The propositions are thus not something the agents can come to believe – given their epistemic situation – without violating ordinary epistemic requirements, such as the *ER*. If those requirements pertain to these agents and their beliefs, the propositions they come to believe will be *doxastically venturous* with respect to the agents and their epistemic situation:

> *Doxastically Venturous Propositions*: A proposition, *p*, is doxastically venturesome with respect to some epistemic subject, *S*, in a case α, if and only if it would be epistemically unreasonable for *S* to form the belief that *p* in α.

The other respect in which the agent (*qua* intention-believing something) resembles the climber is that they are both forming beliefs that create their 'own verification', as James put it. Believing enables the climber to leap across the chasm just as much as the intention-belief makes the agent to what she intention-believes she will do. These beliefs are thus what we may call *self-fulfilling*, or beliefs that, as James says, relate their bearer to a 'class of truths of whose reality belief is a factor as well as a confessor' (1892, p. 528):

> *Self-Fulfilling Belief*: A belief that *p* is self-fulfilling for some *S* if and only if the truth of *p* depends on whether *S* believes that *p*.[10]

The intention-belief is self-fulfilling due to its functional role as an intention since it is 'the kind of state that motivates one to do what can be done and guides it to completion' (Setiya 2008a, p. 39). Forming the intention-belief will thus produce its own truth (barring

[10] A self-fulfilling belief is a fortunate belief belonging in the wider category of *self-affecting* beliefs, the possession of which determines the truth-value of the proposition: *Self-Affecting Belief*: A belief that *p* is self-affecting for some *S* if and only if the truth-value of *p* depends on whether *S* believes that *p*. See Andrew Reisner (2007, p. 305).

circumstances where an agent fails to carry out her intention to Φ due to weakness of will, obstruction, or what have you). Despite that, however, one may be epistemically at fault for forming these beliefs if so forming violates sound epistemic principles. Lack of evidential warrant for *P* may thus warrant criticism for some agent coming to believe *P* despite the rather fortunate fact that so believing makes *P* true because the belief is self-fulfilling.

Venturesomeness is thus a relation obtaining between the psychological state representing something as true, its content, the epistemic subject and her epistemic situation. It is a property of the belief, *Bp*, when its content, *P*, is venturous with respect to the subject, *S*, and her epistemic situation. When it obtains *S* may not form *Bp* without this impinging on her rationality. Let me see if I can render this a bit clearer. Suppose we define a *case* as a possible total state of a system consisting of an agent (or subject) *S* at a time *t* paired with an external environment *E* (where *E* may, or may not, contain other subjects). A case is thus like a possible world with a distinctive subject and time and which we may specify by giving its conditions or facts. The range of properly indexed conditions with respect to the distinguished time and subject that obtain in a specific case in turn defines or individuates the specific case. Cases α and α* will thus be identical if exactly the same properly indexed conditions obtain in them.[11] From a given case, α, there will run branches or series of possible developments from α involving *S* at different times. A condition that may obtain in some α is whether *S* believes *P*. Let us abbreviate this as $B[S, p, t_1]$ (or simply *Bp*). Other important conditions are those that determine *S*'s epistemic situation. If one has specified a case by listing the conditions that obtain in α, one has also specified *S*'s epistemic situation in α. In turn, this defines the epistemic relationships that a case α bears to outcome-cases from α.[12]

Following James Pryor, I shall refer to the conditions that determine a belief's epistemic warrant relative to some case and subject as the subject's body of *justifiers* (2005, p. 182). A justifier of *P* is a justification-making condition that may obtain in a case α to justify *S* believing *P* (at least to some extent). A justifier is thus any condition that makes believing some *p* an epistemically appropriate attitude for *S* relative to a case and an epistemic situation. In this respect, justifiers are what Roderick Firth calls 'warrant-increasing properties' (1964, p. 549). An example is when *S* knows that *Q* is false and that *P* implies *Q*. *S*'s condition is then such that it would be epistemically appropriate for her to infer and come to believe *P*. *Bp*

[11] It is a 'centred world', as David Lewis would say (1979). We may coarsely individuate a condition by the cases in which it obtains (its *case-profile*). Two conditions *C* and *C** are identical if they obtain in exactly the same cases (viz. when they have identical case-profiles).
[12] For some of the above terminology, see Williamson (2000, p. 52).

is then *prima facie* justified (entitled, warranted, grounded, or what have you) and epistemically appropriate to the extent that sufficient justifiers for *P* obtain in a case:

> *Epistemic Appropriateness*: It is epistemically appropriate for *S* in some case α to believe *p* to the extent that justifiers, f_1-f_n, for believing *p* obtain in α.

Epistemic Appropriateness ties epistemic warrant in with the occurrence of conditions in a case while allowing warrant to come in degrees. We should notice that there is a distinction between the epistemically appropriateness of some *Bp* and whether *S* believes *appropriately*. The latter is a property of *S*. In order for some *S* to believe *P* appropriately, we will require more than just the occurrence of sufficient justifiers. It is often relevant whether *S* believes *P* for the right reasons or grounds, or whether *S* merely has an available justification for believing what she does.[13] In order to believe *P* appropriately as opposed to merely having an epistemically appropriate belief, you will need to believe *P* not only in a situation where this belief is appropriate but also for reasons that justify so believing. Insofar as you believe an inappropriate belief, you believe something inappropriately; but you may have an appropriate belief that you hold inappropriately.[14] I shall refer to this as *Appropriate Believing*:

> *Appropriate Believing*: *S* appropriately believes some *p* in some case α to the extent that justifiers, f_1-f_n, for believing *p* obtain in α and *S* believes *p* for those reasons.

Now imagine that α is a case in which some *S* at t_1 has a set of outright beliefs *Bq*, *Br*,..,*Bn* and so on, that *S* does not believe *P* in α and that *S*'s epistemic situation is in a certain determinate way. Let us then restrict ourselves to the possible developments from α where *S* forms *Bp* at some time t_{1+n} without being occasioned by a change in epistemic situation (viz. *S*'s situation is equivalent at t_1 and t_{1+n}). The case where *S* forms *Bp* will then be an epistemic duplicate case to α; however, I shall leave open whether *S*'s epistemic situation may change – for better or for worse – at time t_{1+n} when *Bp* is formed or because of this change. Given *Epistemic Appropriateness,* we can now define venturesomeness as a development from α where *S* is in an epistemic duplicate case with respect to α and forms *Bp* even though *Bp* is epistemically inappropriate with respect to α. That is, one is on a doxastic venture insofar as the belief one forms lacks sufficient justifiers given the kind of case one is in. Making a

[13] As Pryor mentions, there are probably more conditions, such as whether one has taken proper account of any evidence you have that tells against our undercuts your grounds for believing *P* (2005, p. 182).
[14] There is a parallel in the moral domain. Kant famously described an agent – viz. the case of the prudent grocer (*G* 4:397) – who did the morally right thing but not for the morally relevant reasons. Kant used this case to argue that you will need to both do what there are moral reasons for doing and act for those reasons in order to be morally praiseworthy.

transition into such a case impinges on *S*'s rationality with the result that *S* must be less than perfectly rational.[15] Venturesomeness is thus not necessarily measuring the correctness of the belief once we form it; rather, the locus of error lies in the transition. It is forming the belief that impinges on one's rationality. If *S* had been perfectly rational and satisfied all rational requirements applying to her, she would not have formed this belief.

Since one does not in general form intentions to Φ as a function of having evidential warrant and reasons for believing that one will Φ, one will often find oneself in situations where the intention-belief is venturesome in this sense – viz. by violating ordinary epistemic requirements – despite being self-fulfilling. In this way, the intention-believer would be similar to the hero appearing in Kierkegaardian or existentialist thinking, who, much like James's climber, is required to make a dramatic leap of faith to settle important world-view changing questions.[16] Such leaps into the void may well be the right way to characterize some of the most important choices we make in life; but humdrum instances of intentional agency cannot plausibly bear this mark of epistemic recklessness. It is a reason to doubt Cognitivism if it must invoke acts of intention-believing in many of our ordinary choices of action, which, like acts of faith in Bishop's gloss, will be 'an active venture which goes beyond – or even, perhaps, against – what can be established rationally on the basis of evidence and argument'? (2002, pp. 471–2)[17] In particular, if the content of one's intention in action does not need to stand in a suitable epistemic relationship to one's evidence, it raises a question about Cognitivism and the extent to which such an intention can be a kind of belief.

There are two aspects to this problem: on the one hand, Cognitivism seems to entail that agents sometimes choose and exercise voluntary control of some of the things they believe given that they will sometimes need to intention-believe some *p* despite knowing that *p* lacks epistemic warrant. In this way, Cognitivism comes along with *doxastic voluntarism*. Now doxastic voluntarism is not just the claim that people may form beliefs without abiding to the rules of sound epistemic reasoning, or that belief-formation is not exclusively a function of the perceived evidential merit of the proposition one comes to believe. It seems beyond

[15] Interestingly enough, this need not mean that *S* by making this transition becomes less rational than she was at the intervals before t_{1+n}. Believing *p* because of venturesomeness will necessarily preclude one from being perfectly rational; however, it may make one more rational than if one failed to believe *P*. For a similar point with respect to practical rationality, see Arpaly (2003, p. 36).

[16] See John Bishop (2002, p. 471). Also see, Sören Kierkegaard (1968, p. 180).

[17] My point does not depend on accepting what Bishop calls a *Doxastic Venture Model* of religious faith (2002; 2005; 2010). On that model faith is the outcome of a general capacity to bring about beliefs that go beyond – and perhaps even against – one's evidence in certain circumstances by a venturesome process of belief-formation. Someone believing a religious dogma due to faith thus exercises a general capacity to believe something under venture; or, to speak in traditional *Kierkegaardian* terms, she will exercise the capacity to believe something because she makes a *leap-of-faith*. James took the climber story to be an instance of this exercise, and a particularly interesting one because forming a belief by faith in this case leads to a happy outcome: namely, the truth of the belief. He thereby sought to vindicate religious faith by understanding it as venturesome – but somewhat acceptable – belief-formation. My point is to make a comparison between the belief involved in the climber story and an intention-belief; it will not matter if religious faith turns out not to be a doxastic venture.

doubt that we sometimes succumb to our non-intellectual, 'passional and volitional nature', as James says (1892, p. 459), and that our convictions are shaped by influences that make it possible to believe things that go beyond, or even against, the merits of our evidence. Voluntarism about beliefs involves more than that, however: namely, the idea that we may form such beliefs deliberately and at will. The problem with that, as Williams points out, is that it would involve being able to form a belief in *P* with little or no regard for its truth; and it is doubtful whether such a state can still be thought of as a belief, or something 'purporting to represent reality' (1970, p. 148).[18] This is not tantamount to saying that beliefs are 'not entirely out of control' (Kornblith 1982, p. 253); after all, we may confer evidence, make investigations, ask around and do all sorts of things to make sure we have the right sorts of belief. 'One does have a say in the procedures one undertakes that lead to' the formation of beliefs, as John Heil says (1983, p. 363). The point is rather that the beliefs are not under direct voluntary control. In contrast, our ability to decide and choose what to do by forming intentions seems to be more directly under our control, making it questionable whether we could form epistemically venturous intention-beliefs with the ease we form such intentions.

Under the cognitivist framework, however, it seems as if agents are criticisable for exercising their agency by forming venturesome intention-beliefs. If intentions are beliefs, they are liable to rational criticism insofar as they conflict with the principles of *Epistemic Appropriateness* and *Appropriate Believing*; unless, that is, there is a principled reason why they are exempted. This is not something that squares well with how we think of actions from a naive perspective since, as Paul notes, 'intending to go for a walk seems to involve no such defect of theoretical rationality' (2009b, p. 546). Similarly, it is hard to imagine that we could object to Anscombe's shopper fro failing to have evidence for the content of her intentions. If cognitivists have to concede these points, intention-beliefs would seem to be a very weird sort of beliefs indeed. The *Doxastic Venture Problem* is thus worrisome, as Setiya says, in that the

> *... exercise of practical reason involved in doing something intentionally or in making a decision for the future can* (...) *begin to seem not only different from but also incompatible with the proper functioning of theoretical reason.*[19]

This throws doubt on the claim that intentions can be a sort of belief. Given that, on the one hand, forming intentions does not appear to involve anything that could possibly license the

[18] Williams (1970) is, of course, the classic contemporary discussion of doxastic voluntarism. I have also benefitted from discussions in Earl Conee and Richard Feldman (2004, ch. 4), Pamela Hieronymi (2006), Setiya (2008a) and Andrew Reisner (*forthcoming*).
[19] Setiya (2008b, p. 397).

kind of criticism that Clifford gave voice to, while, on the other hand, agency does not seem to involve the psychological complexity associated with doxastic voluntarism, doxastic venturesomeness in action will put pressure on the very idea that intentions are beliefs.

Cognitivists will then have the following options: (1) face the situation and accept widespread epistemic recklessness in the heart of human agency. (2) reject that these cases are doxastically venturesome by arguing that, despite appearances to the contrary, agents are evidentially entitled to form these intention-beliefs. Or (3) argue that intention-beliefs for some reason are exempted from the rational standards that govern acquisition and retention of ordinary beliefs – viz. the *Evidence Requirement* – but argue that they are nevertheless epistemically warranted in a non-evidential sense. An immediate problem with (1) is that it does not square well with our naive perspective and also fails to remove the issue of doxastic voluntarism. I shall therefore focus on strategies (2) and (3) to see whether they can remove both challenges to Cognitivism by showing that the intention-belief, despite appearances to the contrary, really is epistemically appropriate. Until now, I have talked about lack of evidential warrant in the same breath as doxastic venturesomeness and danced around a standard reading of the *Evidence Requirement* as if it says all there is to say about epistemic appropriateness. As we shall see, a natural target for cognitivists is to dispute that claim.

§3 EVIDENTIALIST REPLIES

James took the lucky outcome of the climber story to count in favour of a role for venturesome beliefs in our cognitive economy given that such truths may depend on our coming to believe them. '[A] rule of thinking which would absolutely prevent me from acknowledging certain kinds of truth if those kinds of truth were really there,' he says, 'would be an irrational rule' (1892, p. 477). Now being self-fulfilling is not a general feature of doxastically venturesome beliefs or faith; after all, forming beliefs by leaps of faith is no guarantee for the truth of what one believes and there is little reason to think that doxastic ventures fare any better.[20] Being self-fulfilling is rather something that belongs to the peculiarity of cases akin to James's climber story. What James argued was therefore that it is

[20] The truth of this claim depends on how we carve out the class of beliefs that come about by faith. James's climber could just as well be deluding himself. However, some notions of faith define it as knowledge of specific truths revealed by God. The Reformer John Calvin, for instance, simply defined faith as 'a firm and certain knowledge of God's benevolence towards us, founded upon the truth of the freely given promise in Christ, both revealed to our minds and sealed upon our hearts through the Holy Spirit' (1559 3.2.7). If one defines faith as knowledge, it will pertain to all and only truths or facts due to the factivity-condition on knowledge (Williamson 2000, p. 21). Alvin Plantinga and other reformed epistemologists, for instance, argue that faith is knowledge because it produces beliefs by the operation of a special cognitive faculty – i.e. the *sensus divinitas* – that generates true beliefs about God. Plantinga thus takes faith to be a source of knowledge that makes it 'entirely right, rational, reasonable, and proper to believe in God'; however, since Plantinga still takes faith to be evidentially venturesome, it would be an epistemically rational belief that is 'without any evidence at all' (1983, p. 17). See Plantinga (1983; 2000), Paul Helm (1998) and William Alston (1991).

only in those fortunate cases that venturesome belief-formation is reasonable. James was thereby advocating a new epistemic rule (or principle) that allowed for venturesome beliefs under certain conditions.[21] Although I think James fails to establish a new epistemic rule, his strategy points to how a cognitivist should deal with the *Doxastic Venture Problem*. The problem, as we just saw, is an illicit transition that appears to warrant rational criticism and blame. The challenge is therefore to answer how this transition could be warranted. In other words, the cognitivist must show that forming an intention-belief is epistemically appropriate. In this section, I shall discuss whether there is room for arguing that agents acting intentionally have evidential warrant for the contents of their intentions.

A way for intention-beliefs to be epistemically appropriate is by not being evidentially venturesome, or by the belief being reasonable on the agent's body of evidence:

> *Evidentially Venturous Propositions*: A proposition, *p*, is evidentially venturesome with respect to some epistemic subject, *S*, in a case α, if and only if believing *p* in α would be unreasonable on *S*'s body of evidence.

One would need to show that agents have evidential warrant for what they do intentionally. If one is a strict evidentialist, as I have defined it (§2), this strategy suggests itself given that the only normative reasons for believing something would then be evidential reasons:

> *Strict Evidentialism*: There is an epistemic or normative reason *f* for some epistemic subject, *S*, to believe *p* if and only if fact *f* is a reason to believe *p* and *f* is evidence for *p*.[22]

For evidentialists, beliefs are venturesome insofar as they are evidentially unreasonable:

> *Doxastically Venturous Propositions*: A proposition, *p*, is doxastically venturesome with respect to some epistemic subject, *S*, in a case α, if and only if believing *p* in α would be unreasonable on *S*'s body of evidence.

Similarly, the strict evidentialist may reformulate the above appropriateness principles to define both epistemic appropriateness and appropriate believing in terms of respecting one's evidence. If strict evidentialism is correct, a cognitivist had better find grounds for thinking that there is evidence for the intention-beliefs.

[21] In addition to being self-fulfilling, James added that the belief must be a live option – viz. appear as a 'real possibility' to the agent – and the agent's choice must be forced (the option must be to either endorse *p* or reject *p* so that one cannot be indifferent) (1892, p. 458).
[22] See Reisner (2007, p. 304; *forthcoming*), Williamson (2000, p. 207) and Conee and Feldman (2004, ch. 4).

One way to do that is to insist on an *externalist* notion of warrant and evidence. Now the externalist position is perhaps best understood in contrast with its opposition, the *internalist* notion of evidence. On internalist notions, some epistemic availability-relation or another must obtain between subjects and their evidence for it to be true that the subject has evidence or justification for some belief. Now I think it is obvious that we need to impose some availability-requirement in order for talk about epistemic appropriateness and doxastic venturesomeness to make sense. Evidence is admittedly hard to define, but there seems to be wide agreement that some *e* is evidence for some hypothesis *h* only if *e* speaks in favour of *h* – and thus increases the likelihood of *h* being true[23] – but that *e* in addition must have some kind of 'creditable standing' for the epistemic subject *S*. Thus, the *e*s that belong in *S*'s body of evidence are thus plausibly determined within 'appropriate limits of epistemic accessibility' (Reisner 2007, p. 307). After all, without any availability-requirement on what is to count as justification or evidence with respect to some *S*, we risk an overly permissive form of justification and epistemic appropriateness with the result that one will have license to believe any fact regardless of whether one has or could have any epistemic contact with it or not.[24] Even the wildest leap of faith may then be epistemically appropriate if it just so happens to saddle one with a true belief. We might be warranted in believing facts about epistemically inaccessible parts of the universe, such as pertain to individual elusive objects – viz. objects that are incapable of being individually thought of but are yet knowable collectively since they are in a way traceable through their collective effects.[25] Such a notion of warrant and appropriateness is simply implausible. I therefore conclude some availability-requirement must hold between epistemic subjects and some *e* if *e* is to count as evidence for a hypothesis *h* for some *S*. Suppose we then take the following to define the notion of evidence:

Evidence: *e* is evidence for *h* for *S* if and only if *S*'s evidence includes *e* and $P(h|e) > P(h)$.[26]

Then, I think we can cash out the required availability-relation by saying that the subject must be aware – at least potentially – of the justifiers or evidence for believing *p* in order for *S* to have justification, evidence, or reasons for so believing. Following Michael Bergman, I shall take it that an internalist requires the following for some *S* to have justification for a belief:

[23] That is, the probability of *h* conditional on *e* should be higher than the unconditional probability of *h* (Williamson 2000, p. 186).
[24] See Reisner (2007, p. 307).
[25] For a brief discussion of the notion of elusive objects, see Williamson (2007, pp. 16–7).
[26] Williamson (2000, p. 187).

> *The Awareness Requirement*: *S*'s belief *Bp* is justified only if (i) there is something, *X*, that contributes to the justification of *Bp* – e.g. evidence for *Bp* or truth-indicator for *Bp* or the satisfaction of some necessary condition of *Bp*'s justification – and (ii) *S* is aware (or potentially aware) of *X*.[27]

We can now add this requirement directly to the above appropriateness principles, or else throw in an internalist account of justifiers/evidence as a supplement to the original principles:

> *Internalist Appropriateness*: It is epistemically appropriate for *S* in some case α to believe *p* to the extent that (i) justifiers (or evidence) f_1-f_n obtain in α; and (ii) *S* is aware (or potentially aware) of some or all of f_1-f_n.

> *Internalist Justifiers*: Some *f* is a justifier/evidence for the belief that *p* relative to some *S* if and only if *f* obtains in α, P(*p* | *f*) > P(*p*) and *S* is aware (or potentially aware) of *f*.

Depending on how we read the requirements of awareness – viz. as involving full conscious awareness, tacit knowledge, or that one must see its relevance for the truth of what one believes, and so forth – one can then juggle with internalist notions of different strengths.[28] Their opposition is the extreme externalist position that simply denies any availability-requirement for something to count as a justification or evidence for some belief.

To argue that intention-beliefs are not evidentially venturesome one will need to argue that the intention-belief is justified on the agent's body evidence. Now the fact that the intention-belief is or will be correct is not in general evidence for the truth of that proposition. For a fact to be a reason for believing something on this internalist evidentialist notion of justification, it must stand in an evidential relation to the contents of the belief (as defined by *Evidence*) (Reisner 2007, p. 307). In our case, the cognitivist will therefore need to argue that prior to some *S* Φ-ing intentionally enough facts $f_1,...,f_n$ stand in evidential relation to the fact that one will be Φ-ing (or is Φ-ing) in virtue of belonging to *S*'s body of evidence and increasing the likelihood of *S* Φ-ing. Only then will the intention-belief be justified on this notion. Now the internalist notion of evidential warrant will, however one defines the availability-requirement, single out a limited body of evidence – viz. within appropriate limits on our epistemic availability – and thus distinguish between facts that one can justify on one's body of evidence from facts that one cannot so justify. That distinction will also cut across the

[27] Bergman (2006, p. 9).
[28] See Bergman (2006). I profit from the discussion of Bergman's view in Frank O. Barel (2009, pp. 57–9).

range of actions that one can do. The range of all the things that I can do if I so choose will thus extend beyond the class of things I have evidence for thinking that I will do.

As Williamson says, '[e]pistemic justification aims at truth in a sense – admittedly hard to define – in which pragmatic justification does not' (2000, p. 207). Thinking with a view to action – if guided by a concern for practical and pragmatic justification – may cause the agent to settle for something that is less likely to happen given one's evidence (that is, less likely prior to forming one's intention). One may therefore choose to Φ even though there are not enough facts standing in the evidential relation to the fact that one will Φ *prior* to this choice. No matter how low the threshold of sufficient likelihood is there is always the chance that agents may choose to pursue what is an unlikely course of action on their body of evidence and succeed in intentionally so doing. It can happen, for example, if the agent fails to know what her body of evidence contains. As Williamson has argued, '[*w*]*hatever* evidence is, one is not always in a position to know what one has of it' (2000, p. 178). Rational thinkers may thus fail to know which beliefs are evidentially justified on their body of evidence and which fail to be so by virtue of failing to know of some item *e* that it belongs to one's body of evidence. I see no reason why intention-beliefs should be any different in this respect; one may surely form intention-beliefs for which one fails to have evidence simply because one fails to keep track of what one's evidence is. Since it is an intention-belief she will then form an evidentially venturesome belief. On this notion of evidentialism, an agent will thus be epistemically at fault for forming these intention-beliefs. Given that an extreme externalist notion of evidence cannot account for the notion of epistemic appropriateness and that on any other notion agents will form evidentially venturesome intention-beliefs given that pragmatic justification aims at something other than truth, I conclude that cognitivists about intention must seek epistemic justification beyond the usual strictures of evidentialism.

This brings us to the next evidentialist reply. In *Practical Reflection* Velleman considers a worry akin to the *Doxastic Venture Problem*. Velleman's question goes as follows: 'For how could you ever have grounds for expecting an action whose very occurrence would remain unlikely until you expected it?' (2007, p. 56) Even though Velleman talks about an agent's expectation and her lack of warrant for expecting that she will Φ, his problem is clearly the same as that we encountered above. An agent lacks evidence for a specific intention-belief or expectation, so there is a question of whether one could form this attitude without making oneself rationally criticisable for flouting sound epistemic principles. Velleman argues that even if one must form the expectation/intention-belief 'in the absence of

sufficient evidence for it' prior to the formation of this attitude, you may, as an agent, nevertheless be 'perfectly within your right to do so' (2007, pp. 56–7). The reason why agents are in their right to do this, Velleman goes on to say, is that you can rely 'on your tendency to do what you expect [or intention-believe]' (2007, p. 57). Velleman's idea is that this ensures, as Setiya says, that one has 'evidence that the belief would be true if one formed it' (Setiya 2009a, p. 132). The claim is therefore that there is some evidence, $e_1,e_2,...,e_n$, that will justify the intention-belief once the belief is formed. Velleman's suggestion is therefore to say that we are epistemically licensed to form such beliefs without sufficient prior evidence as long as we know that these beliefs will be supported by sufficient evidence once the belief is formed.

According to Setiya, Velleman thereby rejects the commonly accepted principle that evidence must 'precede the belief that it supports' (Setiya 2008b, p. 398). On this model a subject may have sufficient evidence, $e_1,e_2,...,e_n$, to justify some *p* despite lacking sufficient prior evidence for the truth of this claim since her evidence may include *in-coming* evidence, or evidence whose existence depends on forming the belief in question. In this respect, some beliefs really are their own verification.[29] We may then form some evidentially based belief, *Bp*, without responding to prior evidence. Rather, *Bp* may be based on and supported by the in-coming evidence, $e_1,e_2,...,e_n$, which comes about as a result of the belief-formation. As Velleman says, forming a self-fulfilling belief is not necessarily running ahead of evidence; rather, it may be 'running toward the evidence' (2007, p. 64). The intention-belief may be epistemically appropriate if the agent has – and knows that she has – in-coming evidence for its truth. It is 'a conclusion to which [s]he jumpt before the evidence is complete' (2007, p. 64). This will then allow her to form an intention-belief which is not only epistemically appropriate but also appropriately formed by virtue of running towards her evidence in this way. If something like in-coming evidence is accepted, we may thus defend Cognitivism by showing that it satisfies sound epistemic principles. In particular, it would enable a cognitivist to satisfy both the principle of *Epistemic Appropriateness* and *Appropriate Believing*.

The problem with this suggestion is, as Setiya argues, that most of us 'find the forming of beliefs without prior evidence epistemically suspect' (2008b, p. 400). Just consider

> *... the situation you would be in if you believed yourself (perhaps with good reason) to be watched over by a benevolent spirit, who sees to it that whenever you form a belief on a certain subject-matter (say, the winners of horse-races), it is true.*[30]

[29] So Velleman is not merely advocating *Conservatism*, or the doctrine that any belief, *Bp*, is *prima facie* justified merely in virtue of being held (Harman 1986; Andrew Chignell 2010). Rather, the idea is that there will be robust evidence for the belief when it is formed.
[30] Dorr (2002, pp. 99–100).

Velleman's account of the intention-belief is, as Setiya points out, 'epistemically on a par with forming a belief about the winning horse' in this scenario (2008b, p. 401). Like the intention-belief, a belief about the winning horse is something we know will be true as a result of coming to believe it. That is, we know something like the following conditionals:

> *Horse:* Of all horses, $h_1,h_2,...,h_n$ participating in the race, if I believe for any horse, h_x, that it will win the race, h_x will win.
>
> *Action:* Of all actions, $\Phi_1,\Phi_2,...,\Phi_n$ that I can perform at some time t_1, if I intention-believe that I will be Φ-ing$_x$, I will be so Φ-ing$_x$.

To begin with *Horse*: the problem is that forming a belief about a winning horse – although pragmatically justified – is epistemically speaking unreasonable. You may know that you will know what your own belief is once formed; and you may know that the benevolent spirit will work its magic and make the belief come true once formed. You therefore know that there will be evidence for thinking that Bucephalus wins conditional on you coming to believe it. Still it seems like an epistemically unreasonable thing to believe despite all the conditional knowledge. The conditional is, after all, no reason for believing that you will believe that this particular horse will win rather than, say, Malabar or Buttermilk. Same thing goes for the intention-belief: you may know that by forming the intention-belief you will know that you have formed it and that you will in fact do as it says. Yet your justification for these beliefs does not justify the belief that you will in fact be Φ-ing, rather than, say, γ-ing or ψ-ing. One's evidential situation is therefore as Reisner describes it in the following:

> *Suppose* (...) *that I am a reliable believer about my own future actions, because I know my own intentions and am good at carrying them out. Then the fact that I believe I shall do something is a reason to believe that I shall do it. So, suppose that if I were to believe that I shall start the day by swimming in the North Sea, then I would be correct. This conditional is not evidence that the well insulated sea-lions off the East Coast of Fife will have me as a companion tomorrow morning, for I certainly do not intend to put so much as a toe in the icy water.*[31]

The problem is therefore that in-coming evidence lends no support with respect to the specific belief you will need to form in the here and now. One may perhaps render this a bit clearer by considering how non-prior evidence fails to satisfy the probability requirement on evidence with respect to some particular belief. As I defined it above, evidence *e* for some

[31] Reisner (2007, pp. 308–9).

hypothesis *h* must contribute to the probability of the truth *h* (viz. the probability of the truth of *h* conditional on *e* must be higher than the truth of *h* unconditioned by *e* (or given not-*e*)). Suppose now that you are wondering whether Malabar or Bucephalus will win the horserace. They will then correspond with two hypotheses, h_M and h_B. You know that if you believe that Malabar will win, the benevolent spirit will work its magic and make Malabar win. However, you also know that if you believe that Bucephalus will win, Bucephalus will win. Even if we take into account in-coming evidence, both hypotheses h_M and h_B will thus enjoy an equally high likelihood of being true on your body of evidence. What you lack in this situation is thus something to increase the likelihood of believing one over the other that may allow you to discharge the consequent of the conditional. However, if you lack something to increase the likelihood of the content of your belief, you arguably lack evidence. The problem for Velleman is therefore that he portrays the evidential situation in forming an intention as being similar to the situation one would be in with respect to the above horse race. The horse race clearly does not give us ground for believing that some specific horse will win. And unless there is a way in which forming intention-beliefs differs from this scenario, we have no reason to think that Velleman's in-coming evidence will contribute to the intention-belief so that it comes out as epistemically appropriate on evidential grounds.

The failure of Velleman's suggestion illustrates something important: namely, that an agent's situation is one where the different options are equally probable given one's evidence. These options will therefore not lend evidential – and thus probabilifying – support to some specific intention-belief even if any of these options will obtain as the result of so forming this belief. What this suggests is therefore that a cognitivist must leave the strictures of strict evidentialism in order to show that the intention-belief is epistemically appropriate. Perhaps there is room for saying that despite lacking evidence for the content of the intention-belief the agent has non-evidential epistemic reasons for so believing. That is, maybe we can pry apart evidential and doxastic venturesomeness and grant that the intention-belief is evidentially venturesome but hold on to the idea that it is epistemically appropriate to form it. Perhaps, that is, there are non-evidential epistemic reasons for believing something in addition to the evidential epistemic reasons one operates with on strict evidentialism:

> *Evidential Reason*: Fact *f* is an evidential reason for agent, *A*, to believe *p*, <u>if and only if</u> *f* is a reason to believe *p* and *f* is evidence for *p*.

> *Non-evidential Reason*: Fact *f* is a non-evidential reason for agent, *A*, to believe *p*, if and only if *f* is a reason to believe *p* and *f* is not evidence for *p*.[32]

The epistemic subject, *S*, may thus lack evidence for the truth of *p* in some case α and nevertheless have epistemic justification (entitlement, warrant, or what have you) to form the intention-belief that *p*. As we shall see next, this is what Setiya is suggesting.

§4 KNOWING-HOW as NON-EVIDENTIAL JUSTIFICATION

Setiya formulates the challenge of the *Doxastic Venture Problem* as the task for cognitivists to find 'the epistemically relevant difference between forming an intention, with its corresponding belief, and making a leap of faith that is justified post hoc' (2008b, p. 401). Setiya agrees with Velleman that intention-beliefs must somehow be exempted from the requirement of prior evidence; however, Setiya thinks this is equivalent to thinking that intention-beliefs must be exempted from evidential requirements. The hope of cognitivism thus rests on being able to move beyond the strictures of strict evidentialism and come up with a principled reason for exempting this class of self-fulfilling beliefs by showing that despite lack of evidence they are nevertheless epistemically appropriate or justified. That is, one must be able to explain 'when a subject is justified in coming to believe that *p* without sufficient prior evidence' by virtue of answering the following question (2009a, p. 133):

> *What distinguishes forming an intention with a correlative belief from wishful thinking? If not by having evidence, in virtue of what are we epistemically justified in forming such beliefs?*[33]

Setiya's answer to this question is that in the absence of sufficient prior evidence intention-beliefs are nevertheless epistemically justified through our having or possessing relevant knowledge-how. Setiya says he wants to follow Ryle in treating knowledge-how as a distinctive kind of knowledge whose possession enables its bearer to do things well by virtue exercising this knowledge and which manifests itself in the skilful execution of certain abilities.[34] Setiya's strategy is thus to argue that despite lacking evidential reasons for the truth of the intention-belief the truth of the belief is nevertheless non-evidentially justified by an agent's possession of knowledge-how. In a word, the suggestion is that knowledge-how to Φ is a non-evidential justifier for the intention-belief that concerns one's intentional Φ-ing.

[32] Reisner (2007, p. 305).
[33] Setiya (2008b, p. 402).
[34] Setiya interprets Ryle (1949, ch. 2) as arguing in favour of a distinctive kind of knowledge – knowledge how – that is a *sui generis* knowledge-relation that departs from ordinary knowledge or knowledge-that. I shall question the adequacy of this interpretation below.

According to Setiya, knowledge-how will then grant warrant for any Φ-ing that is intentional since knowing how to Φ is a necessary constituent of Φ-ing intentionally:

> *Knowing How*: If some *A* is Φing intentionally (or for a reason *R*), *A* knows how to Φ.[35]

As Setiya points out, the truth is more complicated: one may try to defuse a bomb without knowing which wire to cut, pick the red one and by sheer luck stop the timer. Arguably, one does not know how to defuse the bomb in this situation but will nevertheless be held to have intentionally defused it. What seems to butt in at this point is knowledge of how to take relevant means. Suitably qualified the principle thus reads as follows:

> *Knowing How**: If some *A* is Φing intentionally (or for a reason *R*), *A* knows how to Φ, or else is Φing by doing other things that *A* knows how to do.[36]

As a necessary condition on intentional agency, *Knowing How** provides a constraint on action theory similar to the influence of the truth of *Belief* on an account of acting intentionally.[37] What Setiya suggests is that we connect the task of accounting for *Knowledge How** with the task of accounting for *Belief*. In particular, he argues, the role of knowing-how in intentional action is, at least in part, to justify the intention-belief. And, again according to Setiya, a systematic diagnosis of the failure of Velleman's account is precisely its failure to locate the epistemic role of an agent's knowledge of how to Φ.

Now Setiya does not suggest that knowing-how is simply identical to knowledge of one's intention or intentional actions. However, he wants to capitalize on the idea that it is no accident that knowing-how is characterized in epistemic terms. One way to explain this is if knowledge-how grants its bearer justification and knowledge. What he suggests is thus that we are epistemically entitled to form the intention-belief that one will Φ 'when and because we know how to perform the relevant acts' (2008b, p. 406). The fact that one knows how to Φ will therefore supply the agent with sufficient non-evidential reason to justify forming the intention-belief when the knowledge is exercised. To summarize, the idea is that:

> *...[k]nowing how to Φ is the state or condition that, together with knowledge of ability, provides the epistemic warrant for decision. Together, they justify the transition in which one*

[35] Setiya (2008b, p. 403). Many philosophers agree that knowledge-how and acting intentionally go together. Anscombe for one says that: '"Intentional action" always presupposes what might be called "knowing one's way about"' and that 'this knowledge is exercised in the action' (1957, p. 89). Stanley and Williamson seem to agree when they say, 'we do find it very plausible that intentional actions are employments of knowledge-how' (2001, pp. 442–3 and sect. I). Also, see Stanley (2010, pp. 11–12).
[36] Setiya (2008b, p. 404). As Setiya points out, it follows that the first principle must hold unqualified for basic actions.
[37] As you may recall, Setiya argues that the truth of *Belief* must be explained by a theory of intentional agency. See chapter 3.

> *forms the intention and belief that one is doing Φ or that one is going to Φ. More carefully, this transition is justified if and only if one's decision is an exercise of knowledge how to Φ and one has knowledge of ability, in the simple conditional sense.*[38]

The forming of intention-beliefs whereby one decides to dance a tango, speak Spanish, or write a thesis in philosophy is thus epistemically appropriate or justified – and so epistemically exempted from the *Evidence Requirement* – because one's decision to do these things exercises one's knowledge of how to do them. If you fail to know how to dance the tango, your decision to dance cannot exercise any such knowledge of how to dance. Consequently, one will not be epistemically justified to believe that one will and one would be at fault for forming such an intention-belief if one could form it at all.[39]

In reply to criticism Setiya has made some concessions and changes to the original account, while trying to stress other things that might not have appeared clear at first. For one thing, he has let go of the requirement that agents must have knowledge of their ability to Φ in the simple conditional sense (2009a, p. 134). On reflection, he thinks that knowing-how is sufficient for justifying the intention-belief even if one lacks knowledge of one's ability to Φ. A problem with that move is that it weakens his motivation for replacing Anscombe's original claim that acting intentionally must involve knowledge. As you may recall, Setiya thought there were counterexamples to this claim.[40] In particular, he thought that lack of conditional knowledge of one's ability would undermine one's knowledge that one is Φ-ing intentionally even if one still knows how to Φ and is able to Φ intentionally. From this, he concluded, all the agent really needs to Φ intentionally is some belief about what she is doing. This in turn, furnished support for the principle of *Belief* the truth of which, as we saw in the previous chapter, was meant to motivate his cognitivist account of intention in action. If Setiya now accepts that the intention-belief can be justified without conditional knowledge as long as the agent knows how to Φ, it may leave room for saying that the agent also has knowledge of what she is doing insofar as she knows how to Φ and is exercising that ability. In other words, this move weakens his dialectics against claims emphasizing the necessity of knowledge in intentional action and thus weakens his support for Cognitivism about intention.[41]

Another issue is Paul's criticism of Setiya's account of the intention-belief's epistemic justification.[42] According to Paul, the problem with invoking knowledge-how is that 'it does

[38] Setiya (2008b, p. 407).
[39] The example is Setiya's (2008b, pp. 406–7). What one can decide to do, however, is trying to learn it since you do know how to do that.
[40] See chapter 3, §2.
[41] The most plausible such claim is what I previously formulated as the principle of *Agent's Knowledge* (AK). See chapter 2, §4.
[42] Paul also launched other arguments against Setiya's cognitivist account. I have discussed them in chapters 3 and 5.

not tell the agent which action to *expect'* so that the agent must still leap to that conclusion (2009b, p. 556). Paul takes this to undermine Setiya's account since she believes it leaves the intention-belief just as epistemically reckless as any instance of self-fulfilling wishful thinking. In response, Setiya makes an important disclaimer: knowing-how is not and does not provide the agent with evidence or reasons to expect anything:

> *Knowing how to clench my fist is among the conditions in virtue of which I am epistemically justified in forming the belief that figures in my intention. Unlike evidence, however, it does not justify this transition by making it more likely that the belief is true, before it is formed. Even combined with knowledge of ability, knowing how to clench my fist is not a sound basis on which to predict that I am doing so, or that I am going to do so in the future*.[43]

Paul is wrong to think that the epistemic role of knowing-how was to raise probability and thus function like an ordinary piece of evidence. Obviously, knowing-how in a role as evidence would need to increase the probability of that which it justifies but the mere presence of knowledge-how fails to do that. However, knowing-how was never meant to fill this epistemic role. Setiya's basis idea is that the presence of knowing-how makes these beliefs epistemically appropriate and rational despite lacking sufficient evidence and not by providing the agent with additional evidence. Rather, knowledge-how contributes to the rationality of the intention-beliefs due to its non-evidential epistemic role.

This concession connects with the way Setiya's Minimal Cognitivist theory departs from what I defined earlier as *full-blown* cognitivist accounts.[44] Setiya does not believe that the thinking that leads to action and the formation of an intention to Φ aims at knowledge and truth. Practical thought is simply *not* driven by theoretical concerns so it does not have to respond to evidence or look for factors that increase the probability of one's doing something. What Setiya hopes to show is that even if practical thinking lacks this concern and, as a result, often ends up with the formation of intention-beliefs for which the agent lacks evidence, the content of the belief is nonetheless epistemically justified due to the necessary presence of sufficient non-evidential justifiers: namely, a relevant body of knowledge-how. Given Setiya's rejection of Cognitivism about practical thinking, there is thus no need for him to find justification for this belief that helps predicting its truth.

Setiya's reply still leaves the distinctive epistemic role of knowing-how in the dark. If it does not provide evidence, then what is its epistemic role in acting intentionally? How does

[43] Setiya (2009a, p. 134).
[44] See previous chapter.

knowing-how make the intention-belief epistemically appropriate if it is not by supplying the agent with evidential warrant? Setiya's answer is to offer three reasons for thinking that knowledge-how fills the required epistemic role. First, he says, 'we need something to distinguish forming an intention-belief, when it is epistemically justified, from wishful thinking or a leap of faith' and he thinks knowing-how is well located for the task (or 'in the right place at the right time') (2009a, p. 134). Second, he thinks the idea that knowing-how does epistemic work helps explaining why it is a kind of knowledge. Third, he thinks knowing-how is well suited for this task. To see how it fulfils these needs we need to know more about the nature of knowing how. Let us briefly see what Setiya says.

In his first paper on this topic, Setiya argued that the decision to Φ itself is an exercise of knowledge-how (2008b, p. 407). In response to Paul's criticism, Setiya retracts this view and proposes instead that knowing-how is a partly dispositional capacity that consists in the disposition to Φ in execution one's intention. Knowing-how is now accounted for as a capacity to execute one's intention once formed rather than a capacity to form the intention in the first place and in exercising this dispositional capacity for basic actions one's intention-belief will be non-accidentally true (Setiya 2009a, p. 136). Given that knowing-how to Φ is required for Φ-ing when Φ-ing is a basic intentional action – viz. the truth of *Knowing How** – it follows that the relevant intention-beliefs will be reliably and non-accidentally true. The idea is thus that when we equip an agent with this reliable basis, we can distinguish intention-beliefs from merely fortunate incidents of reckless wishful thinking. The dispositions that partly constitute knowing-how thus play an epistemic role in buttressing the epistemology of the intention-beliefs. Knowledge-how may thus enable Setiya to separate and categorize the intention-beliefs as epistemically appropriate, while leaving room for criticism of other self-fulfilling beliefs that lack sufficient prior justification and relevant knowledge-how.

This can be illustrated with the aforementioned horserace case: Setiya can point to the agent's lack of relevant knowledge-how as a reason for judging the self-fulfilling belief about Bucephalus winning the race as epistemically inappropriate and unjustified. For an intention-belief, however, possession of relevant knowledge-how may justify forming the belief in spite of the agent's lack of evidence for its truth. The possession of relevant knowledge-how may be the 'something' that can 'distinguish forming an intention-belief, when it is epistemically justified, from wishful thinking or a leap of faith' (Setiya 2009a, p. 134). Setiya's account thus appears to have sufficient resources for showing that the intention-belief is not 'epistemically on a par with forming a belief about the winning horse' (Setiya 2008b, p. 401).

In what follows, I shall focus on two main lines of criticism against this proposal. The first approach is to raise general concerns about Setiya's suggestion that knowing-how serves an epistemic or justifying function for cognitivist intention-beliefs. The latter approach will tie my worries specifically in with the nature of Setiya's Cognitivism and thus give us reason to dispute his view in particular. I shall spend the next section pursuing the first approach while leaving the specific criticism of Setiya's view for the subsequent section.

§5 PROBLEMS for the PROPOSAL

If Setiya's proposal works out, ascribing an agent's knowledge-how with a crucial non-evidential epistemic role in intentional agency is a way to move beyond the strictures of strict evidentialism and a principled way to provide non-evidential warrant for the intention-beliefs. However, in what follows, I will argue that Setiya's suggestion fails. First, I shall argue that a weakness of the proposal is its reliance on anti-intellectualism about knowledge-how. I will then argue that it is questionable whether knowledge-how can serve this justifying role since it is hard to see that it offers anything that could stand as a non-evidential reason for believing something. Third, I will argue that even if we were to grant knowledge-how an epistemic role in justifying the intention-belief it would not be something an agent could take as her reason for intention-believing that she is Φ-ing (or will be Φ-ing) and will leave agents vulnerable to the criticism that they fail to form the intention-belief appropriately.

According to Paul, a weakness with Setiya's dialectics is that there is no argument in his discussion supporting the claim that casting knowledge-how in this epistemic role is the unique best explanation for the truth of a principle like *Knowing How**. As Paul remarks, an alternative possibility would be to think that knowing-how is necessary for the action to occur rather than for forming the intention-belief to act (Paul 2009b, p. 555). In other words, Paul questions whether the truth of *Knowledge How** is support for the epistemic role that Setiya suggests for knowledge-how. Paul has a point: the reason why agents must know how to Φ when Φ-ing is a basic intentional act is that one needs it to be able to do anything at all. After all, knowing how to Φ is the crucial factor that grants its bearer a reliable dispositional capacity for basic actions whose execution will realize intentional actions.[45] Setiya's idea, however, is that one can only explain *Knowledge How** insofar as one also explains the sense in which knowing how is knowledge and why it 'deserves to be conceived in epistemic terms' (2008b, p. 405). It is to explain the epistemic aspect of knowledge-how that he casts it in this

[45] I am *not* saying that knowing how to Φ is enough to grant one a capacity to Φ. A master pianist tragically loosing both her hands may still know how to play Schubert's *Wanderer* piece despite having lost her ability to do so (Stanley and Williamson 2001, p. 416).

justifying role in intentional agency. In a word, Setiya does not think that support for his thesis is in being the unique best explanation of the truth of *Knowledge How**; rather, he seeks to vindicate his thesis as the unique best explanation of the epistemic aspect of this principle.

What I think we may question, however, is whether the epistemic role that Setiya puts knowledge-how into is the unique best explanation of the epistemic aspect of knowledge-how; after all, there are alternatives. One particularly interesting alternative is to explain this by accounting for knowing-how as a species of ordinary knowledge or knowledge-*that*. In contemporary discussions, this claim is often seen as being in opposition to Ryle's anti-intellectualism and his argument against what he called 'the intellectualist legend' (1949, p. 29). It is therefore usually referred to as *Intellectualism* about knowing-how:

> *The Intellectualist Thesis*: When an agent, *A*, knows how to Φ, *A* knows that ... .[46]

Following Greg Sax (2010), however, I think it is inaccurate to attribute the rejection of this thesis to Ryle. Ryle does not appear to be concerned about the nature of knowledge and the idea of, as Stanley and Williamson put it, 'a fundamental distinction between knowledge-how and knowledge-that' (2001, p. 411). After all, Ryle characterizes the legend he opposes as the idea that we should 'reassimilate knowing how to knowing that by arguing that intelligent performances involves the observance of rules, or the application of criteria' (1949, p. 29). He then explains that this idea involves a picture of what it is to do something intelligently where one must first 'consider certain appropriate propositions, or prescriptions' and then 'put into practice what those propositions or prescriptions enjoin' (1949, p. 29). The *Intellectualist Legend*, as Ryle sees it, is thus primarily a thesis about how the intellect guides and informs actions – viz. the 'exercise of intelligence in practice' (1949, p. 40) – that puts undue emphasis on intelligent deliberation. When Ryle talks about knowledge-how, it is therefore just an idiomatic term 'for the sort of prowess or expertise, whatever it ultimately is, in virtue of which one performs intelligently' (Sax 2010, p. 512). Ryle's concern is therefore mainly about skilful ability – or good old know-how – and he wants to show that it is not first to 'do a bit of theory and then to do a bit of practice' (1949, p. 512). His anti-intellectualism is therefore not an anti-intellectualist analysis of knowledge-how but an explanation of practical intelligence that goes against the *Intellectualist Legend*.

The *Intellectualist Thesis* is a different claim altogether and concerns the idea that knowledge-how is a species of – or metaphysically reducible to – ordinary knowledge-that. If

[46] See Stanley (2010, p. 2). Also see Stanley and Williamson (2001) and Paul Snowdon (2003).

it holds, however, it would explain the epistemic aspect of a principle like *Knowledge How** – after all, if Intellectualism is true knowledge-how is just knowledge-that – without needing to invoke a special epistemic role for knowledge-how with respect to the content of our intention-beliefs. The problem for Setiya is thus, on the one hand, that he offers no argument for why his account of its epistemic role should be preferable to the doctrine of *Intellectualism*. This is yet another dialectical worry that could further weaken the support for his view. On the other hand, Intellectualism would also undercut the motivation for Cognitivism about intention. The problem is that if knowledge-how is a species of knowing-that, the former will just like the latter be a relation between a thinker and a true proposition or fact rather than a complex dispositional ability to act. It follows that if one can account for knowledge-how as a species of knowledge-that, the justification of one's intention-belief will be inferred (or mediated) from justification of other propositions or facts. As Setiya says, 'on a propositional interpretation of knowing how, our model of knowledge in intention is inevitably inferential' because the belief that I am Φ-ing (or that I am going to) would then be justified by 'inference from things I previously knew' (viz. from the propositional knowledge that constitutes my knowledge-how) (2008b, p. 408). In a word, the truth of Intellectualism would turn knowledge-how into a piece of ordinary mediated or inferential evidence. The problem for Setiya is that he must categorically reject inferential theories of agent's knowledge in order to support Cognitivism.[47] So if knowledge-how is a species of knowing-that, Cognitivism ought to be rejected. Setiya must therefore reject Intellectualism and assert what we know in the literature as the *Standard View*:

> (1) Knowing-how does not consist in knowing that some proposition is true or that some fact obtains; knowing-how cannot be reduced to or equated with (any form of) knowledge-that.
> (2) Knowing how to Φ does in fact consist in being able to Φ (knowing-how ascriptions ascribe abilities or capacities to do the mentioned action).[48]

Now the nature of knowledge-how has been up for much debate despite strong philosophical consent favouring a fundamental distinction between the two kinds of knowledge. More recently, however, philosophers such as Stanley, Williamson and Snowdon have defended Intellectualist analyses of knowledge-how.[49] According to them, knowledge-how –like knowledge-why, knowledge-when, and knowledge-where along with others that

[47] I discuss the Inferential Theory and how it undermines support for Anscombean Theories – and thus Cognitivism – in chapter 5.
[48] Snowdon (2003, p. 2).
[49] See Jason Stanley and Timothy Williamson (2001) and Paul Snowdon (2003).

can be rendered as 'knowing *X* + infinite' constructions – are best rendered as knowledge-that constructions.[50] An immediate advantage for such an *Intellectualist Thesis* is that it enables a unified semantic account of knowledge-constructions. Another thing, is that it would allow us to explain why knowledge-how seems to be sensitive to Gettier cases. Stanley and Williamson therefore suggest that we treat the attribution '*A* knows how to Φ' as being satisfied when *A* knows of a way *w* that *w* is a way for her to Φ when this proposition or fact is entertained under a *practical mode of representation*. What it is to entertain something under a practical mode of presentation they explain as having a certain complex of dispositions towards the thing one entertains (Stanley 2010, pp. 2–6).[51]

Despite philosophers' 'continued reluctance to abandon [the *Standard View*]', as Stanley laments at some point (2010, p. 2), I think the *Intellectualist Thesis* warrants serious consideration. The irreducibility of knowledge-how and the alleged unbridgeable distinction between knowing-how and knowing-that is far less certain than once thought. The fact that Setiya's account – both his solution to the *Doxastic Venture Problem* and his motivation for Cognitivism – depends on the falsity of the *Intellectualist Thesis* is thus reason to be provisionally sceptical about his suggestion that knowledge-how plays this epistemic role.

Suppose for the sake of argument, however, that the *Intellectualist Thesis* – despite its elegance – is nevertheless false and that knowledge-how is a *sui generis* knowledge-relation irreducible to a relation between an agent and proposition (or fact) under a practical mode of representation. Let us also assume that knowing how to Φ is, as Setiya contends, a distinct kind of knowledge that consists in a complex disposition to be Φ-ing as an execution of an intention to Φ (2009a, p. 135). Suppose further that the complexity of this disposition will cover cases where one still knows how to Φ but has recently lost one's ability to ever again carry out the Φ-ing.[52] I shall now argue that even if we accept all this about the nature of knowledge-how, it cannot play the non-evidential epistemic role that Setiya needs since knowing how to Φ cannot be taken as a reason for believing that we are (or will be) Φ-ing.

[50] See Stanley and Williamson (2001, section II) and Stanley (2010, sections 1 and 3).

[51] The basic idea is that like any kind of propositional knowledge or knowledge-that the proposition or fact must be entertained under a mode of presentation given that whether one knows some *p* is in part determined by the way in which the subject conceives *p*. As Stanley and Williamson argue, there is just as much evidence for such a practical mode of presentation as there is for the first-person mode of presentation associated with 'I' thoughts, 'here' thoughts and 'now' thoughts (2001, pp. 428–30). If modes of presentation are semantically relevant this mode is semantically conveyed, but if not it is pragmatically conveyed that one entertains the proposition under this mode of presentation (2001, p. 430). Also, see prolonged argument in Stanley and Williamson (2001, section II).

[52] We may allow that a master pianist knows how to play Schubert's *Wanderer* piece despite having lost both hands in a tragic accident. Stanley and Williamson argue that such cases will count against the idea that knowledge-how ascribes a capacity to its bearer (2001, p. 416). I think this is wrong and that such cases can be accounted for by the complexity of the disposition to Φ that makes up one's knowing how to Φ. One could thus argue, to use Mark Johnston's terminology (1992, p. 233), that one's ability to play the *Wanderer* piece given that one knows how to play it has been *masked* by one's loss of hands. See Setiya (2009a, p. 31) and Michael Fara (2008, section 2).

Setiya's basic idea is that an agent's knowledge-how is a justification-making fact that non-evidentially justifies the content of an agent's intention-belief. What makes it fill this epistemic role is its reliability in bringing forth what the belief is about – viz. one's action. To stick with our example: when one knows how to play the *Wanderer* piece, the pianist has a disposition to play making sure that it is no accident that she is in fact playing the *Wanderer* piece when she intends to play it. In a word, her being able to make the intention-belief true is not just a matter of luck (Setiya 2009a, p. 136). For Setiya, knowledge-how is therefore not something you need to know that you have, or believe that you have. It can therefore serve to justify our intention-beliefs without its being the content of any assertive or propositional state of the agent. What determines the justification of one's intention-belief is whether this complex disposition is a property of the agent and not whether its agent knows or believes that this holds true of them; or, as Setiya puts it, '[t]he state of knowing how to Φ itself, not just the belief about that state, does epistemic work' (2009a, p. 137).

To see how this picture works, suppose that two agents, *A* and *B*, have duplicate evidential situations, α and β, such that any belief for which *A* has sufficient evidence is a belief for which *B* has sufficient evidence too and vice versa. Moreover, suppose that they also have identical sets of beliefs (or assertive states) so that for any belief in the truth of some proposition, *Bp*, such that it is true that *A* believes *P*, $B_A p$, it will also be true that $B_B p$ and *vice versa*. Suppose also that none of their beliefs includes the belief that they know how to Φ. Finally, suppose that *A* but not *B* knows how to Φ. According to Setiya, it is then true that *A* has epistemic justification for forming the intention-belief that she is hereby Φ-ing whereas this will fail to hold for *B*. In this respect, knowledge-how provides justification that is available to its subject without the subject being in a representational state towards it.

These claims would appear to be in sharp contrast with what we take to be necessary for being a justifier, *j*, in a subject's body of justifiers. A justifier would contribute to determine what it is epistemically appropriate to believe and it is often assumed that the subject must bear some relation or another to *j* to make it epistemically available to her (Pryor 2005, p. 194). Above I discussed a way to cash out this idea by virtue of invoking internalist requirements on justifiers and epistemic appropriateness. One question is whether Setiya's proposal is incompatible with the idea that motivates this kind of internalism: namely, the requirement of epistemic availability. In his view knowing how to Φ justifies the intention-belief despite its agent failing to have an assertive or representational state towards this justifying state. All it takes is to have this complex dispositional capacity that Setiya takes to

be an anti-intellectualist or non-propositional kind of knowledge. One may then ask whether knowledge-how makes anything epistemically available to its subject in the way required for something to serve as one's justifier. If the answer is no, we must ask whether that would disrupt the justifying role of knowledge-how. This brings up two challenges for Setiya: on the one hand, it appears that the agent may fail to bear an availability-relation to an item playing a justificatory role. On the other hand, it seems as if the alleged justifying item may also fail to provide any suitable object for such an availability-relation. I shall now argue that Setiya may deal with the first challenge while I think the latter proves devastating.

First, let me note that the problem is not that Setiya allows something to justify one's belief without the agent knowing that this is something that justifies her belief. Paralleling Williamson's account of evidence (2000, ch. 9), we should allow that to happen. You may not always know what your evidence is, or whether some *e* belongs in your body of evidence; similarly, we should leave room for items that justify one's beliefs without the agent knowing that this is something that belongs to her in this epistemic sense. The question is rather whether possession of knowledge-how, as Setiya conceives it, makes anything epistemically available to its subject. Having an assertive or representational state towards some *j* is one way to make it epistemically available to its subject. Some philosophers will perhaps argue that for some candidate item to be epistemically available the subject must bear an assertive or representational state – say knowledge or belief – towards some content that corresponds to the item giving one justification. In the case of an ordinary justifier, the idea seems to be that a justifier needs to involve some representational content or another that purports to say what the world is like or what it is not. In turn, this conviction has led many philosophers to claim that the only thing that can play a justifying role will have to be other beliefs (or, at least, belief-like cognitive states).[53] The premise for this conclusion is the claim that for any entity to serve as a justifier it must be the case that the candidate justifier has or involves a propositional content that its subject represents assertively (viz. as telling one how the world is or is not). This is what James Pryor formulates as the *Content Requirement*:

> *The Content Requirement*: In order for some *j* to be a justifier for some *S*, *j* must have propositional content and *S* must represent that proposition assertively.[54]

[53] Davidson, for instance, claims that 'nothing can count as a reason for holding a belief except another belief' (1986, p. 141). McDowell agrees when he takes it that '[w]e can justify beliefs we hold about how things are (...) only by appealing to what are in fact further beliefs we hold about how things are' (McDowell 1983, p. 128).
[54] Pryor (2005, p. 188).

According to Pryor, the *Content Requirement* is an offspring of Sellars' attack on the *Myth of the Given* and formulates an allegedly necessary condition on what can be a *prima facie* justifier (Pryor 2005, p. 183; Sellars 1991, pp. 164–70). If it holds, it would clearly rule out knowledge-how as a justifier in cases where the agent fails to know or believe that she knows how to Φ. This may in turn put pressure on its epistemic role. There are, however, two parts to the *Content Requirement*: (1) that the justifier must be available to the subject by virtue of the subject bearing some assertive relation to it; and (2) that the purported justifying item must be a suitable object for such a relation by virtue of being a propositional content.

Either of these claims would, of course, put pressure on the suggested justifying role of knowledge-how in cases where an agent possesses it without knowing or believing that she does since she would then only have a complex non-propositional dispositional capacity. At this stage, then, a defender of the distinct epistemic role of knowledge-how may begin to wonder whether the *Content Requirement* holds universally. Clearly, one's justifiers must be in some sense epistemically available to their subject; however, translating this into a requirement about being in a relevant representational state towards a proposition corresponding to one's justifier may just smack of question begging. Pryor for one has argued that this notion of epistemic availability would rule out the event of you having a headache as a justifier for the belief that you are feeling this way if you fail to stand in a further representational state towards the event of you having a headache (2005, pp. 194–5). Since a headache may justify one's belief that one is feeling this way, it seems as if epistemic availability is weaker than being in a representational state towards a corresponding content or proposition of one's justifier. If we grant this and can find some other means for cashing out the availability requirement, the event of you having a headache may justify the belief that you are feeling this way without thereby having to be the object of one of your representational states. The question will then be whether we could make a similar move with respect to knowledge-how; maybe its role as a justifying factor may retain, just like a headache, its justifying force without being an object of the agent's representational states.

I think not. The reason is that I think we can derive minimal requirements of epistemic availability from the thinking that has motivated philosophers to accept the *Content Requirement* and that these minimal constraints will rule out the envisioned epistemic role for knowledge-how. Let us take this systematically. The two components of the *Content Requirement* are, on the one hand, the claim that a subject must relate to her justifiers in a certain way to make them epistemically available to her. On the other, it makes a claim on the

nature of what justifies. For the sake of simplicity, let us call the first the necessity of an *epistemic availability relation* and the second the necessity of being what I shall call *an epistemic availabel*. The idea behind the first necessity is that a subject must *have* whatever justifies her belief in order for her belief to be epistemically appropriate or reasonable. A belief, *Bp*, may after all be reasonable in the sense that there are good reasons for believing it without *Bp* being epistemically appropriate for *S*. As Peter Unger points out, there is no inconsistency in saying that 'What he believes is reasonable, but he has no reason at all for believing it' (1975, pp. 203–4). Without *S* having an availability relation towards what does the justifying with respect to *Bp* this belief would be epistemically inappropriate for *S*. The other half of the claim pertains to the justifying item itself and requires that it must make available what I have just called an epistemic *availabel* so that there is something there that could be the *relatum* of the availability-relation. I shall call this the *Minimal Requirement*:

> *The Minimal Requirement*: In order for some *j* to justify some *Bp* for *S*, *j* must be an epistemic *availabel* and *S* must bear an epistemic availability relation to *j*.

Knowledge-how can obtain without its agent bearing an assertive relation to it. If assertive relations were the only means of having an availability-relation, it would thus be devastating for Setiya's proposal. However, I think Setiya can get around this by saying that the exercise-relation – that is, the fact that one is able to exercise one's knowledge-how insofar as one possesses it – may serve as an availability-relation with respect to knowledge-how. I shall therefore take it that knowledge-how satisfies this half of the *Minimal Requirement*. However, I think it fails on the other half. In order to justify, knowledge-how must make a relevant epistemic *availabel* available to *S*. What must minimally hold for something to count in favour of some belief, *Bp*, with respect to an epistemic subject, *S*, is that *S* must be in position to take this as her reason for forming *Bp*. That condition lays down an important proviso on what the item that is doing the justifying: namely, that it must be such that it is able to ground, at least in part, the relevant belief. In a word, it must be 'the sort of thing,' as Pryor points out, 'you could *take as* a reason' for believing something (2005, p. 194 [original emphasis]). For knowing how to Φ to justify the intention-belief that one is (or will be) Φ-ing it must therefore be something one could take as one's reason for so believing.

Now the event of having a headache can clearly be taken as a reason for believing that one is feeling this way. After all, you may attend to the state of yours and on that basis form the true belief that you are feeling this way. Having a headache is a fact that counts in favour

of the belief that one is feeling this way and therefore qualifies as being what Unger calls 'propositionally specific' (1975, p. 205). A headache is a justifier not because it is in fact represented by a proposition or content, but because it is the kind of epistemic object that could be so represented and thereby serve as a reason for believing that one is feeling this way. The problem for Setiya, however, is that there is nothing in the possession of the complex dispositional capacity of knowing how to Φ that gives rise to such an epistemic *availabel* – or something that is propositionally specific in this way – which one could take as a reason for having the intention-belief. After all, there is no fact or event corresponding to the possession of this capacity that favours the content of one's intention-belief. Knowledge how to Φ can therefore not serve its alleged epistemic role for *this* belief since there are no epistemic *availabel* that one may take as one's reason for intention-believing that one is (or will be) Φ-ing. The absence of a suitable epistemic object thus makes it hard to accept knowledge-how as serving such a justifying role. There is nothing there to ground the content of one's intention-belief and so nothing one could take as one's reason for this belief. I therefore conclude that there is nothing there to make this belief epistemically appropriate.

A third problem is that even if the mere possession of knowledge how to Φ gave us something to make the relevant intention-belief epistemically appropriate there will still be room for rational criticism of the agent forming an intention-belief. The reason is that given the nature of knowledge-how – in particular the fact that one may have it without knowing or believing it – there will be plenty of room for agents Φ-ing intentionally in epistemically inappropriate ways. Recall that epistemic appropriateness is not the same as believing something appropriately since one may still fail to believe some *p* for the right reasons:

> *Appropriate Believing*: *S* appropriately believes some *p* in some case α to the extent that justifiers, f_1-f_n, for believing *p* obtain in α and *S* believes *p* for those reasons.

Even if we were to grant that some *S* could take her knowing how to Φ as a reason for believing something – and, remember, I spent the last few paragraphs arguing that we cannot – it will typically not be the reason we have for intention-believing that we are (or will be) Φ-ing. In some minor cases, or course, we may perhaps Φ just because we can Φ. Dolly Parton, in her song *Jolene*, asks Jolene not to take her man 'just because you can'. Similarly, I decide to ride my bike just because I can. However, in many cases we are doing things for other reasons. If knowing how to Φ epistemically justifies these intention-beliefs, one may thus invoke the principle of *Appropriately Believing* to criticize the intention-believer for having

formed her belief for the wrong reasons since they are not 'properly arrived at' (Conee and Feldman 2004, p. 93). A problem with Setiya's suggestion is thus that even if we were to grant its hold as an account of epistemic appropriateness for the intention-belief, it nevertheless fails to remove the grounds for rationally criticizing the agent for her intention-beliefs since we may still criticize her for the way she forms it. Since that does not square well with our naive perspective it gives us reason to doubt that intentions are beliefs.

Finally, I shall briefly mention two more problems. First, I do not think that knowledge-how is sufficient to provide warrant for relevant beliefs as there are counterexamples to this claim. Second, the *Doxastic Venture Problem* remains for an important group of actions: namely, the basic actions we are able to perform because we decide to do them. To take the counterexample first. I think some belief, *Bp*, may fail to be epistemically appropriate despite being accompanied by relevant knowledge-how. Suppose for instance that some agent, *A*, unbeknownst to herself, knows how to sing Dolly Parton's *Jolene* because she had learned it under hypnosis at session with a crazy therapist. Suppose also that the therapist has hypnotized *A* to believe that she is singing *Jolene* (or will so sing) whenever she sees a police officer. Finally, let us imagine that the therapist monitors her belief-states at all times and has the power to ignite her with motivation so that she will begin to exercise her knowledge-how and thus begin to sing no sooner than she has formed this belief. While out and about later in the day, *A* encounters a police officer, believes she is singing and finds that she really is singing. In that scenario, Setiya will have to say that the belief is epistemically justified. After all, the agent believes something for which she has non-evidential justification – viz. the relevant knowledge-how – and it is her knowledge-how that makes sure that she has true beliefs about what she is doing when she sings the tune. Despite having the relevant knowledge-how, however, there is a mere contingency connecting her true belief with the event. She cannot help it – after all, she is the victim of a modern-day Svengali's scheme – but her belief nevertheless appears epistemically inappropriate and unwarranted despite being reliably true due to her knowing-how. After all, it is only true because the scientist monitors her and ignites her motivation that her belief turns out to be true. If knowledge-how really exempts intention-beliefs from the strictures of evidentialism, it ought to do so in this case too. However, I take it that this case obviously involves an unwarranted – though in a way self-fulfilling – belief and thus that relevant knowledge-how is not sufficient for getting one off the hook by warranting relevant beliefs.

Possession of knowledge-how does not appear to cover the range of cases for which the *Doxastic Venture Problem* arises. Just consider the cases where agents gain an ability to Φ – where Φ-ing is a basic intentional action – by virtue of deciding to Φ. This may strike one as odd at first. After all, there clearly are certain restrictions on what one can intend to do which are determined by what one knows how to do. Speaking Spanish, for instance, may not be something I can intend to do without knowing how to do it. As Dummett argues, without knowing how to do it I might not even know what Spanish is (1978; 1991, pp. 93–7; 1993, pp. x–xi). Similarly, we may have reason to doubt that someone can really intend to dance the tango without knowing how to do it. All she may be able to intend is learning how to do it. However, I think there are certain things one will know how to do only because one has decided to do them. Something about the process of making up one's mind may take an agent to a level where confidence and resolution contribute to expand her repertoire for basic actions and gain basic practical abilities. She may then intentionally be Φ-ing and thereby – viz. by forming the intention to Φ – enable herself to Φ. Insofar as this happens and she goes on to Φ I see no reason for criticizing her for forming the intention. If intentions are a sort of belief, however, they were formed in the absence of relevant knowledge-how and will be criticisable on ordinary epistemic principles. After all, these principles will be in play since the agent lacks the relevant knowledge-how and so has nothing to exempt the belief from the grip of these principles. We will then have a case of doxastic venture, though this time one in which knowledge-how cannot make the intention-belief epistemically appropriate because the relevant knowledge-how only came about as the result of forming this intention-belief.

This concludes my general worries about Setiya's proposal. We found that his suggestion stands no stronger than the rejection of the *Intellectualist Thesis*. Then I argued that knowing how to Φ fails to satisfy the minimal requirement there is for playing an epistemically justifying role, namely that its subject must be able to take it as her reason for believing the thing it justifies. The problem here was that no epistemic object is available by the mere fact that one knows how to Φ that could serve as the subject's reason for believing that she is (or will be) Φ-ing. In turn, this suggests that there is nothing in a mere capacity to do something that could make such a belief epistemically appropriate. A third problem was that even if we granted all this, Setiya's proposal would still leave room for rational criticism of the intention-believer given that she would more often than not intention-believe some *p* for the wrong kinds of reasons. Finally, I entertained a brief counterexample to the idea that the mere possession of knowledge-how will give epistemic warrant to the relevant beliefs and

I argued that the problem remains for a certain range of cases. I take this to count decisively against the idea that knowledge-how makes our intentions epistemically appropriate. In turn, this counts against the whole idea that intentions in action are beliefs.

One may ask why I have not pursued an evidentialist line against the idea that knowing-how is a non-evidential justifier. After all, there is a strong presumption in favour of the idea. As Williamson puts it, '[a]n epistemically justified belief (...) must be epistemically justified by something' and 'whatever justifies it is evidence' (2000, p. 208). A swift reply to Setiya would therefore be to deny the following:

> *Non-evidential Justification*: There is some fact *f* such that *f* is justifying for agent, *A*, the belief that *p* and *f* is not evidence for *p*.

The falsity of *Non-Evidential Justification* would rule out Setiya's suggestion that knowledge how to Φ non-evidentially justifies intention-beliefs. No facts could then be non-evidentially justifying anything and that rules out facts about agents knowing how to Φ. If cognitivists are right about the nature of intentions, the fact that some agent, *A*, knows how to Φ, will ensure that *A* reliably holds a true belief about one's Φ-ing. However, in the absence of evidence an evidentialist will no doubt point out that this belief cannot be epistemically appropriate since on their notion the kind of justification that is characteristically epistemic turns entirely on evidence (Conee and Feldman 2004, p. 83). What I have argued, however, has tried to avoid relying on an evidentialist notion of epistemic justification. I take this to be strength of the arguments. Given that Setiya's suggestion was to move beyond strict evidentialism and find that the intention-belief finds exemption from ordinary requirements, it was better to reject the account on its own turf, as it were, rather than raise evidentialist worries.

§6 A SPECIFIC PROBLEM for SETIYA

In the previous section, we encountered problems for the idea that knowledge-how can function in an epistemic role towards the content of one's intention-belief. One of these problems was that knowing how to Φ does not provide its subject with a suitable object that she could take as her reason for intention-believing that she is Φ-ing (or will Φ). A minimal requirement on some justifier, *j*, if *j* is to justify a belief that-*P* for some *S* is that *S* can take *j* as her reason for believing *P*. Similarly, we could evoke a *Minimal Requirement* on items playing a justifying role, a requirement which, as I argued, knowledge-how fails to satisfy. What I will argue in this section is that a similar problem pertains to the desire-like belief that

Setiya identified as an agent's intention in action. Given the content of this belief, it is impossible for some agent to take anything as her reason for believing its content. Since the desire-like belief was meant to be literally taken as a belief, it follows what Setiya takes to be the content of one's intention in action cannot be epistemically appropriate. Again, I take this to count against the whole idea that intentions are beliefs.

Setiya has been trying to come up with a condition that makes the intention-belief rational and epistemically appropriate despite the agent's lack of evidence for the truth of its content. His proposal is a general norm that allows a certain kind of belief to be formed under specific conditions. As a general schema, we may formulate this norm as follows:

> *Epistemic Norm*: Insofar as *A* is in a case α where *C* holds, it is epistemically appropriate for *A* to believe that she is hereby Φ-ing (or will be Φing) intentionally.

The *Epistemic Norm* tells us what an agent must do in order to be justified in believing something. As Pryor points out, you are only complying with it if the fact that you are in a case α such that condition *C* holds guides you or is your reason for so believing (2005, pp. 195–7). A minimal requirement on justification is thus that *S* could comply with a relevant epistemic norm and be guided by whether condition *C* holds or not. It must therefore be possible for the thinker to take the fact that she is in *C* as a reason for forming the belief.

As I argued in the previous section, knowledge-how fails such fundamental constraints with respect to the intention-belief. The nature of the specific problem for Setiya, however, is somewhat different. Setiya, as you may recall, argues that one's intention in action is a desire-like belief. Despite being desire-like, he has emphasized that this state really has the same features as the state of believing something. 'Like belief, it represents its content as being true,' he says, but in addition the desire-like intention-belief has motivational features (2007b, p. 40). This is an important point: after all, the less belief-like this state is the less suitable it will be as an explanation why agents must have beliefs about what they are doing and why, which is the explanatory constraint that Setiya has set for a theory of intentional action.[55] As a result, the *Epistemic Norm* should hold for desire-like beliefs no less than for ordinary beliefs.

The problem, however, is that the content of the desire-like beliefs will not allow an agent to have epistemic reasons for believing their content. The content of these beliefs is that one is *hereby* Φ-ing (or doing Φ) because of the belief that *p* (Setiya 2007b, p. 46). Here '*p*' is meant to act as the agent's reason for so Φ-ing. The intention-belief is thus, according to

[55] See previous chapter and the discussion of *Belief*.

Setiya, about one's Φ-ing intentionally, or for the reason *p*. Now this content also involves the notion 'hereby'. As Setiya says, its function is to depict 'me as being motivated partly by itself' (2007b, p. 45). The clause is there to make the belief self-referential and point to the formation of the desire-like belief itself as part of the motivational explanation for the intentional action. One thus believes that one is Φ-ing for a reason because one believes *p* but also because one is forming the desire-like belief that one is so Φ-ing.

The self-referential nature of this content makes it impossible for agents to treat something as their epistemic reason for forming the belief that one is hereby (viz. by forming this belief) Φ-ing for a reason. Now there are several layers to this problem. First, it seems as if beliefs formed in response to epistemic reasons are often not desire-like. If I form the belief that I will apply for job vacancies the next few months, I may do so for good epistemic reasons without thereby getting any favourable inkling towards this endeavour. Since there is no reason to expect an ordinary belief to be desire-like – even if it pertains to my own behaviour – there is no reason to expect intention-beliefs to be desire-like insofar as they are responses to one's epistemic reasons. Taking some *j* as my epistemic reason to intention-believe something may thus lead to the belief losing its motivational features.

However, even if one were able to pull this stunt off and form desire-like beliefs in response to epistemic reasons, $j_1,j_2,...,j_n$, this fact will make sure that the resulting behaviour is not intentional or an instance of acting for reason. In a word, were the agent to succeed in this it would threaten the analysis that Setiya takes the desire-like beliefs to offer: namely, an account of what it is to take something as one's reason when one is acting intentionally. An agent is not Φ-ing intentionally when her behaviour is based on sufficient epistemic justification. Intending to Φ is not – and cannot be – forecasting to Φ. On the contrary, agents form an 'intention on the basis of practical thought,' as Setiya says, and their 'reasons for doing so' – viz. for forming the intention – 'are practical reasons' (2009a, p. 134). Practical thought aims at action, whereas thinking with a view to truth and epistemic justification, as Williamson says, is a process whereby one 'aims at truth in a sense – admittedly hard to define – in which pragmatic justification does not' (2000, p. 207). The epistemic frame of mind could not be one's background for forming a desire-like belief that one hereby Φ-es while at the same time preserving this as the basis for an analysis of what it is to act intentionally or for a reason. If an agent were to form an intention-belief as the result of thinking with a view to believing something true rather than with a view to action, it would no longer appear to be true that she is hereby – viz. by forming this desire-like belief – Φ-ing

intentionally or for a reason. The desire-like belief can therefore not be formed based on taking something as one's reason for so believing while preserving the desire-like belief, so formed, as an analysis of the agent taking some *p* as her reason for Φ-ing.[56]

Somewhat paradoxically, it therefore follows that if one were to succeed in forming such a desire-like belief in this epistemic way – viz. by virtue of observing one's reasons for believing its content – it would no longer be true that one is hereby (viz. by forming this belief) Φ-ing intentionally or for a reason. Forming the desire-like belief in this epistemic manner will thus be self-affecting in a rather unfortunate way in that it would lead to its own falsification. In a word, the desire-like belief, so formed, would be self-falsifying:

> *Self-Falsifying Belief-formation*: Taking *R* as one's reason for forming the belief that *p* is a self-falsifying belief-formation for some *S* if and only if the falsity of *p* depends on whether *S* takes *R* to be her reason for believing *p*.

If the agent were to form the desire-like belief that she hereby Φ-es because she believes *p* for some epistemic reason, it would no longer be true that she hereby Φ-es for a reason. Since it is a minimal requirement that its agent can take what justifies her believing *p* as justifying, it follows that the desire-like belief can only be justified by something that would render the belief false if one took it as justifying so believing. Epistemic justification may sometimes be odd; but it cannot be so weird that following its recommendation is a sure way to believe something false. It would follow that such beliefs have no epistemic justification and will therefore be rigidly venturous or notoriously inappropriate, at least in an epistemic sense. They will be as unattainable – or just as rigidly irrational, epistemically speaking – as the thought, 'I am not thinking anything right now'. Intention-beliefs cannot function in their role as intentions if they are beliefs of this sort. I therefore take this consequence to be a decisive objection to the whole idea that intentions can be desire-like beliefs, as Setiya describes them.

A commentary is needed on what may otherwise appear as a puzzling result. In the last argument, I have *not* argued that agents cannot have evidence or justifiers for a belief that they are Φ-ing intentionally or for some reason, *R*. Nothing I have I said stands in the way of an agent taking some *j* as a reason for thinking that she is Φ-ing intentionally. Nor have I argued that taking some *j* as one's reason for believing that one Φ-es intentionally stands in the way of the practical thinking on which one's intention to Φ is based. What I have argued

[56] In a word, being able to form the desire-like belief for epistemic reasons will make it unfit for capturing the truth of what Setiya identified as the *Circular Model*: to take *p* as one's reason for Φ-ing is to have the desire-like belief that one is Φ-ing for the reason that *p* (2007b, p. 46). I discussed the relationship between this circular model and the desire-like belief in the previous chapter.

is rather that one cannot have epistemic justification for the belief that one is hereby (viz. by forming this belief) Φ-ing for reason *R*. The fact that nothing can be treated as justifying this self-referential thought without rendering it false suggests that it is not a state that needs epistemic justification. I take this to support the conclusion that an agent's attitude whereby she takes something as her reason is not a belief. In other words, this specific problem counts as a decisive reason to reject Setiya's cognitivist account of intentions.

§7 CONCLUSION

In this chapter, I have argued against the idea that intentions can be beliefs by pursuing the *Doxastic Venture Problem*, or the problem that the contents of these beliefs will often lack epistemic justification and thus be doxastically venturesome. To accommodate this problem, cognitivists must distinguish intention-beliefs from other instances of epistemically reckless thinking. I argued first that externalist notions of evidence could provide evidence for intention-beliefs only at the cost of making almost any truth epistemically appropriate. I then argued that any availability-relation pertaining to the notion of an agent's body of ordinary evidence gives rise to the *Doxastic Venture Problem*. In considering non-standard notions of justification, I argued that Velleman's idea of in-coming evidence – or evidence that will come about conditional on forming the belief in question – fails to distinguish between intention-beliefs and other cases of reckless, though self-fulfilling, belief-formation. Finally, I argued against Setiya's proposal that knowledge-how gives non-evidential justification for the content of intention-beliefs. On the one hand, I argued against the idea that knowledge-how could justify these beliefs. In particular, I questioned the consequence that Setiya must deny the *Intellectualist Thesis* about knowledge-how, before I argued that knowledge-how – anti-intellectually conceived – fails the *Minimal Requirement* by failing to provide an appropriate object that an agent may take as her reason for having or forming the intention-belief. I also objected to the proposal on the grounds that it fails to deal with the problem of intention-believing appropriately, that there are counterexamples to the idea that knowledge-how exempts evidentialist strictures on relevant beliefs, and, finally, that there is still room for doxastic venturesomeness for certain intention-beliefs. After that, I argued specifically against Setiya's notion of the intention-belief and I contended that it is not possible for an agent to have epistemic justification for these beliefs because it is impossible to take anything as one's justification for forming them without thereby rendering the beliefs false. I thus concluded that this rules out an epistemic role for knowledge-how for these beliefs; but also that the

nature of Setiya's intention-beliefs rules out finding justification at all. Together with the considerations I raised against Cognitivism about intention in the previous chapter, this gives us reason to dispute the whole idea that an intention is a belief.

I began to discuss Cognitivism about intention in the previous chapter – in particular in Setiya's Minimalist version – because such a conception of intentional agency was meant to function as a crucial lemma in Setiya's argument against ethical rationalism. As I have defined them, rationalists pursue a justificatory strategy in ethics where they seek to vindicate the normative authority of ethics and morality by showing that the standards of practical reason non-circularly justify the standards of good conduct and the virtues of character. I argued in chapter 1 that this strategy requires a certain normative interpretation of intentional agency. In particular, it depends on being able to derive, at least in outline, the standards of practical reason from the nature of action and intentional agency. Setiya's Minimalist Cognitivism was meant to block this step for the rationalist by coming up with an adequate conception of intentional agency that fails to be normatively rich enough to warrant the rationalists' assumption. The centrepiece of this Minimalist framework is the idea that an intention in action – or the attitude by which an agent takes some reason, *R*, as her reason for Φ-ing – can be accounted for as a sort of desire-like belief rather than as a practical commitment whereby one endorses Φ-ing for reason *R*. What the last two chapters show is that this idea fails. Whatever an agent does when she is acting intentionally or for a reason she is not forming a desire-like belief. Setiya's crucial lemma in his bottom-up argument against rationalism is therefore false. Of course, this is no vindication of rationalism. Nor does it remove the explanatory constraint that Setiya took to motivate a cognitivist account in the first place – viz. to explain the pervasiveness of agent's knowledge. However, it may indicate that Setiya took the right step in the wrong direction. I shall therefore continue to pursue this explanatory constraint to see whether there really is room for rationalists to be optimistic.

The collapse of Cognitivism shows that an intention in action is not a doxastic attitude about what one is doing and why. Velleman is therefore right: the agent's thoughts don't follow her actions at all; they lead them (2007, p. 18). Unlike Velleman, however, I think this requires a practical conception of thinking with a view to action. The next chapters will discuss whether such a view may allow us to account for agent's knowledge.

5

The Inferential Theory

> [*S*]*ome people* (...) *say that what one knows as intentional action is only the intention, or possibly also the bodily movement; and that the rest is known by observation to be the* result, *which was also willed in the intention. But that is a mad account: for the only sense I can give to 'willing' is that in which I might stare at something and will it to move.*
>
> G.E.M. Anscombe[1]

§1 INTRODUCTION

Agent's knowledge – or the knowledge an agent has of her own intentional actions in virtue of being their agent – is the linchpin of what I have defined as an Anscombean approach to intentional action.[2] Anscombeans take the peculiarities of agent's knowledge – first and foremost the fact that it is non-observational and appears to be immediate or non-inferential, but also the fact that it appears to be a non-coincidental result of acting intentionally – to indicate that an agent's knowledge in action is spontaneous and comes about solely in virtue of the fact that one acts intentionally. According to them, the nature of agent's knowledge will thus have a bearing on how we account for the nature of the actions one thereby knows. In chapter 2 I argued that conditional on accepting the necessity-claim – viz. the claim that agents *must* spontaneously know some of the things they are doing insofar as they are acting intentionally – this Anscombean line of thinking can be supported by general methodological principles. A theory of action will consequently need to explain why the processes, attitudes, or mental events in virtue of which an agent acts intentionally or for reasons must give rise to agent's knowledge as a matter of necessity. Anscombeans are

[1] Anscombe (1957, p. 52).

[2] The expression a 'linchpin of Anscombe's account of intentional action' is used by Teichman (2008, p. 12). I discussed the connections between agent's knowledge and Anscombean Theories in chapter 2, §5. Briefly put, an Anscombean Theory of action starts out from the claim that an agent's distinctive epistemic access to the nature of what she is doing and why she does it – whether the nature of this access is best characterized in terms of knowledge, qualified beliefs, partial beliefs, increased confidence, or what have you – is a necessary condition for acting intentionally. The necessity of such agent's knowledge must be explained by an account of intentional agency and this may in turn require epistemic elements in the intention with which an intentional action is executed. In particular, I argued, the process by which an agent takes something as her reason to act will need to be, at least in part, an epistemic relation.

therefore inclined to think that the necessity of knowledge in action is due to the fact that acting intentionally must involve having an epistemic relation to the fact that one acts for some reason *R*. In particular, agent's knowledge may require us to think of the relation that obtains between agent, reason and action when the agent takes some *R* as her reason for acting as involving, at least in part, epistemic attitudes and elements. For Anscombeans, the fact that an agent knows or has epistemic access to her own actions is therefore not independent of the nature or essence of what one thereby knows. In a word, they take it that our way of knowing our own intentional actions pertains to the nature of these actions.

Intentional actions are not alone in being something an agent can tell straight off and without observation that she is doing, however. As Anscombe says, intentional actions are 'a sub-class of the events in a man's history which are known to him *not* just because he observes them' (1957, p. 24 [original emphasis]). One can, for instance, know straight off that one has hiccupped without needing to observe clues or having knowledge of the peristaltic movements of the gut. However, there is no analogous suggestion that such knowledge pertains to the nature or essence of these processes. Our knowledge is no more part of nor necessary for the process of hiccupping to take place than one's psychology pertains to the nature of the process of digesting one's dinner.[3] Knowing straight off that one has hiccupped is rather a result of how one is epistemically situated in relation to the act of hiccupping, the nature of which is independent of one's knowledge. Could something analogous be said about agent's knowledge? Could we explain the nature and pervasiveness of agent's knowledge in intentional agency without thinking that it bears on the nature of what it knows?

This chapter will look closely at one of the most plausible non-Anscombean alternatives. The alternative is an effort to explain agent's knowledge and account for its pervasiveness in human agency while nevertheless rejecting the Anscombean conclusion that agent's knowledge pertains to the nature of its object. In particular, I shall focus on what is known as the Inferential Theory that has recently been defended by Sarah K. Paul and which explains agent's knowledge as something we know inferentially from non-observationally knowing our intentions and suitable background conditions. If such a model is sufficient, Paul argues, it will make room for having a *Distinctive Practical Attitude* (or DPA) theory of intentions in action since one may then explain agent's knowledge as being inferred from

[3] This is not the same as denying that our psychology may influence these processes. Interaction between our psychological and bodily states is a complex affair and our psychology presumably both influences and is influenced by such non-psychological bodily processes. What I am contrasting here are processes the nature of which could not be fully explained or understood without invoking the higher-order psychological states or knowledge of the agent undergoing these them – such as the mental act of deciding something, or a linguistic act of communication – and those processes where these states are not a part of what makes the process what it is.

knowledge of our distinctively practical intentions. What is common among different DPA views is, according to Paul, that they conceive intentions in action – or the attitude by which one is taking some *R* as one's reason for Φ-ing – as a matter of forming DPAs, or 'distinctively practical, conative commitments that do not constitutively involve the belief that one will do what one intends' (2009a, p. 11). If one can account for agent's knowledge from such DPA-intentions, one may thus avoid the epistemic difficulties that we saw pertaining to Cognitivism about intention – or the view that one's intention in action is or constitutively involves a peculiar sort of belief – while at the same time overcoming one of the chief obstacles that Setiya has raised against DPA-accounts of intentional agency: viz. their difficulty in accounting for the nature and pervasiveness of agent's knowledge. In a word, the Inferential Theory may give the upper-hand back to the DPA-views over Cognitivism.

That in itself is no minor result. As we recall from chapter 1, I argued that Setiya needs a non-normative theory of one's intention in action – or of the notion of taking as one's reason – in order to reject Ethical Rationalism, viz. a strategy in ethics that seeks to defend the rationality and normative authority of morality on the basis of normative interpretation of what it is to act intentionally or for a reason. I then showed (chapter 3) that he sought to motivate such an alternative account – viz. his own *Minimalist Cognitivism* about intention – as the unique best way to account for the phenomenon of agent's knowledge, as he interprets this phenomenon. With Paul's model on the table, however, it would no longer be clear whether agent's knowledge cold still be used in support of his view. In fact, an Inferential Theory will tend to stand in the way of what I have defined as Anscombean approaches to intentional agency. Setiya, in particular, would thus lack an argument against Ethical Rationalism. I have already argued (chapter 4) that there are epistemic difficulties with Setiya's Cognitivism. I thus agree with Paul that the problems with these views point in the direction of views emphasizing the distinctively practical nature of acting intentionally; however, I disagree with her when she takes the failure of Cognitivism as a reason to abandon Anscombeanism. In order to defend the motivation there is for Anscombeanism, however, I will need to stop Paul's Inferential Theory from neutralizing it. In other words, I will at the very least need to level the score between her inferential account of agent's knowledge and an Anscombean account. As a result of this chapter's ensuing discussions, I thus hope to raise enough difficulties for Paul's Inferential Theory to re-open the door for an Anscombean approach motivated by our need for an explanation of agent's knowledge.

I shall first discuss the structure of the Inferential Theory and its explanatory aims (§2). I turn then to Setiya's objection against this view and I argue that it begs the question given Paul's explanatory aims (§3). However, I think there are other worries one might raise against the Inferential Theory. In particular, that it fails to provide agents with a reliable non-observational evidential base for one's knowledge in action (§4) and that agent's knowledge as inferred knowledge fails to square well with the phenomenology (§5). Finally, I shall take the opportunity to remind ourselves of the reasons why Anscombe was concerned that a familiar picture in epistemology – which gives primacy to knowledge of what goes in our minds over what we do – could be carried over into the metaphysics of action (§6).

§2 The INFERENTIAL THEORY

The basic idea of Paul's Inferential Theory is to explain agent's knowledge as something an agent must have – if not as a necessary consequence of acting intentionally so at least in core cases of intentional agency – and that this knowledge is non-observational even though it is evidence-based and mediated like other kinds of empirical knowledge.[4] Paul therefore needs to argue that agents in the normal case will know the nature of their own intentional activities in virtue of inferring this knowledge from other things they know. Paul's solution is to say that we know what our actions are by virtue of inferring this knowledge from knowing our intentions and some suitable background conditions. In this way, the knowledge we have in action will be 'evidentially based on our knowledge of what we intend to be doing' (2009a, p. 4). Knowledge of my ability to Φ may, for instance, be 'supplemented by knowledge of some other proposition [such as the proposition that one has formed an intention to clench one's fist] from which I then infer that I must be clenching my fist', as Setiya says (2008b, p. 393). However, Paul nevertheless insists that this inferred and mediated knowledge is non-observational and, she argues, accounting for this will be enough to single out agent's knowledge as epistemically distinctive. In what follows, I will assume that knowledge is *immediate* (or direct) if it is non-evidence based and mediate otherwise.[5] I will also adopt Paul's terminology at this stage and treat inferential knowledge as beliefs or belief-like states formed by virtue of being 'derived from prior states that (ideally) evidentially support those beliefs' in epistemically favourable circumstances (2009a, p. 10). Being inferentially based is thus a way for knowledge to be mediated. What Paul tries to show,

[4] Paul's theory is a further development of a similar-looking theory developed by Grice (1971, pp. 278–9). See Paul (2009a, p. 11). There are serious problems with Grice's theory, a discussion of which, however, I must defer for dialectical reasons to an appendix below.
[5] See Cynthia Macdonald (1998, p. 131) and Pryor (2005, p. 183) for related notions of immediacy.

therefore, is that agent's knowledge can be mediated by being inferential and evidence based while nevertheless being non-observational. I shall spend the following section expounding details of this account and briefly note some of its key advantages.

According to Setiya, there is 'a profound conflict in action theory between those who follow Anscombe in seeing a close connection between intentional action and knowledge of what one is doing and those who do not' (2009a, p. 131). In particular, there are some who take such knowledge to pertain to the nature of these actions –thus requiring what I have defined as an Anscombean Theory – and those who resist this idea. Paul belongs to the latter. But although Paul takes a non-Anscombean stand in this 'profound conflict', she clearly wants to acknowledge the phenomenon of agent's knowledge and stand apart from the sceptics who deny agent's knowledge. Indeed, she is willing to grant much truth to the observations we made in chapter 2 about 'the pervasive presence of non-observational belief [or knowledge] in intentional action' as well as acknowledge that this is something a 'theory of action should account for' (2009a, p. 6). Where Paul draws the line is at the Anscombean suggestion that explaining agent's knowledge requires an understanding of intentional agency as being – or, at least, as constitutively involving – a distinctively epistemic relation.

In order to be able to draw the line where she does, Paul will have to disagree with Anscombeans over two crucial issues. First of all, she must deny that acting intentionally necessarily involves having knowledge or beliefs about what one is doing and why since general methodological principles would then pull us in the direction of an Anscombean Theory of action.[6] Secondly, she must deny agent's knowledge as the immediate and spontaneous result of acting intentionally since that would be enough for Anscombeanism to get some bite. In order to do that, Paul invokes an Inferential Theory that is meant to cover the 'kernel of truth' behind Anscombeanism and so explain 'the insight that we usually do know at least something about what we are doing intentionally, and know it without having to observe ourselves' without thereby implying the necessity of spontaneous knowledge in action (Paul 2009a, p. 3). Paul's immediate aim is therefore to model an agent's knowledge in action as knowledge inferred from knowledge of one's intention plus knowledge of suitable background conditions. Her hope is that it may nevertheless be possible to maintain the sense in which this knowledge 'doesn't', as Velleman says, 'require observation of the action in question' (2007, p. 20).[7] The project is therefore to come up with an account of knowledge in

[6] In chapter 2, §5 I argued on general methodological grounds that a necessary relation between acting intentionally and acting knowingly will be support for Anscombean Theories, or the idea that so acting just is a way of knowing.

[7] Of course, this is not to deny that your 'knowledge of what you're doing requires a background of observational knowledge', as Velleman points out (2007, p. 20). The point is that when you choose one of many possible actions and the fact that against the background of your

action as mediated and inferred knowledge without the agent 'becoming a spectator of one's own actions', as O'Shaughnessy puts it (1963, p. 391). If she succeeds, the sufficiency of the Inferential Theory will allow Paul to avoid the conclusion that we need an Anscombean Theory of action since her account will then effectively neutralize the primary source of motivation for an Anscombean approach. In turn, the sufficiency of the Inferential Theory may allow Paul to think of our intentions in action as 'distinctively practical, conative commitments that do not constitutively involve the belief [or knowledge] that one will do what one intends' (2009a, p. 11). Me and my partner may, for instance, make up our minds about going to the opera next week by committing ourselves to this course of action and treating it as reason-giving, so that our commitments and plans in turn will organize and 'guide our later conduct', as Bratman says (1999b, p. 2). Paul thereby defends what she calls a *Distinctive Practical Attitude* (DPA) theory of intention,[8] or the kind of theory that stands in stark contrast to Cognitivism about intention and views according to which intentions in action just are – or, at least, constitutively involve – belief-like attitudes.

Paul does not want to commit to a specific DPA account at this stage. However, she asks us to think of these DPA-intentions 'in terms of the functional role they play as settled objectives that shape further practical thinking, initiate and sustain action, and at least usually engage certain norms of practical rationality' (2009a, p. 12). The important thing is that DPA-intention, so conceived, will not essentially involve beliefs or belief-like attitudes; nor must an agent have beliefs about the content of what they DPA-intend to do as a matter of necessity. However, Paul thinks that the *DPA*-intention will nevertheless be 'intimately tied up with our beliefs about ourselves and about the world' since its occurrence in our cognitive economy will help us 'to achieve our goals and to aid in intra- and inter-personal coordination at a time and over time' and – which is the crucial part – because we are 'tending to believe, in the absence of evidence to the contrary, that we are doing what we intend to be doing' (Paul 2009a, p. 12). According to Paul, this tendency to infer what we are doing from what we intend to do also depends on knowledge of a battery of background experience and assumptions, such as, knowing that it is in one's ability to perform the intended action, that one's circumstances are conducive for the action, that one will follow through on one's intention by knowing one's history as an agent, and so on (Paul 2009a, p. 15). The idea is to

standing observational knowledge you know that you are Φ-ing rather than Ψ-ing 'doesn't seem to require observation of the action in question' or of other things from which one may conclude that this-and-that action is executed (Velleman 2007, p. 20).

[8] Different *DPA* views have been suggested and defended by different philosophers: Davidson, via his emphasis on an all-out judgement in action (1978); Frankfurt and his emphasis on second-order volitions (1971); Bratman, via his notion of intentions as elements in plans of action that aid intra and inter-personal coordination over time while being subject to norms that structure practical reasoning (1999b). In the same breath, one can mention Alfred Mele and his idea that intentions are executive attitudes towards plans (1992 chs. 8–11; 2003, ch. 1, §5). See Paul (2009a, p. 11n45 and n46; 2009b, p. 547n3).

say that when knowledge of background conditions is in place and in the absence of evidence to the contrary, an agent's knowledge of her own actions tends to materialize and thus be grounded in knowledge of her own intentions. Paul's first move is therefore to claim that the fact that we know our intentions serves as an evidence-base for inferential knowledge about our actions. In this respect, knowledge (or beliefs) about what we are doing is not guaranteed by the fact that we act intentionally; rather, such knowledge will only be likely to materialize inferentially from knowledge of what our DPA-intentions are.

An integral part of Paul's approach is thus to disagree with Anscombeans about the extent to which one should take the phenomenology of agent's knowledge at face value. As you may recall from our discussion in chapter 2, agent's knowledge is like perceptual knowledge and knowledge of our own minds in appearing to be *cognitively spontaneous*.[9] This knowledge of what I am doing and why – which I have as a result of acting intentionally – thus feels as if it is direct, non-inferential and unmediated, at least 'from the inside' or from the perspective of the agent who is acting intentionally. If I am heading for the pub as opposed to the church next door to the pub, it seems as if I know this immediately, directly and spontaneously. In other words, the following principle seems to be true:

> *Cognitively Spontaneous*: Expression of intentions/self-ascriptions of actions/intentions are psychologically non-inferred in the sense that they are not formed by any conscious process of inference from other beliefs.[10]

According to this principle, agent's knowledge is *psychologically* non-inferred, which means that the knowledge one has in action does not come about as the result of an inference you are consciously performing or have access to. 'You don't find out what you're doing,' as Velleman says; rather, we 'feel as if we're inventing what we do' (2007, p. 4, 47). Anscombeans want to take this phenomenology at face value as support for the claim that agent's knowledge really is immediate and 'not the conclusion of an inference,' as Stuart Hampshire puts it (1960, p. 70).[11] Some followers of Anscombe even think this falls out of the way she clarifies the notion of non-observational knowledge when she says that an agent knowing her action is in no need of 'separately describable sensations, having which is in

[9] Åsa M. Wikforss (2004, p. 272) makes a similar claim with respect to observational judgements that seem to give the perceiver a direct impression of what the room is like, what colour the ball is, and so on. Other philosophers have maintained that the same phenomenology appears to hold for knowing the meaning – and/or significance – of linguistic communication. As Elisabeth Fricker says, the interpreter 'hears the utterance not merely as sound, but as the speech act that it is' and this representation of meaning is 'not phenomenologically distinct from her hearing of the sounds' (2003, p. 325).

[10] This is a rendering of Elisabeth Fricker's *Minimal Phenomenon* as it applies to knowledge of one's own mind (1998, p. 168). For the connection between Fricker's application of these principles and agent's knowledge, see my discussion in chapter 2.

[11] Anscombe agrees when she says that 'in so far as one is observing, inferring etc. that Z is actually taking place, one's knowledge is not the knowledge that a man has of his intentional actions' (1957, p. 50).

some sense our criterion' for the relation to obtain (1957, p. 13; 1981, pp. 71–4). Setiya for one treats these clarifications as support for the idea that 'the crucial claim about the epistemology of knowledge in intentional action' is that when I Φ intentionally 'I know that I am doing it, and my belief [by virtue of which I know] is not inferred from sufficient prior evidence – that is, from evidence that precedes its formation' (Setiya 2008b, p. 393).[12] According to Paul, however, the impression that our knowledge in action is cognitively spontaneous may give us the wrong picture of what is really going on. Mere phenomenology is inadequate at this stage, she argues, since 'evidence-based information processing can take place rapidly and automatically at a non-conscious level' so that the knowledge may appear as unmediated, non-inferential and spontaneous while nevertheless being non-consciously inferred from something else one knows (Paul 2009a, p. 6). Paul's point is that an inferential model is not ruled out even if one accepts the phenomenology and the claim that knowledge in action is psychologically non-inferred since we perhaps explain these data on a model where non-observational knowledge of what we do intentionally is evidentially based and non-consciously inferred from knowledge of what we intend to be doing (Paul 2009a, p. 10).[13]

Paul's claim is that knowledge of our DPA-intentions tend to produce rapid, non-conscious inferences about the nature of what we are doing and that in favourable epistemic circumstances – that is, given the truth of what we are believing, the absence of evidence to the contrary, knowledge of suitable background conditions, and so on – these inferences will produce knowledge of what we are doing and that it will appear to be direct, immediate and spontaneous due to the non-conscious nature of these inferences. She recognizes that there are two immediate challenges for this view. The first is to make sure that, as Paul says, 'in core cases of intentional action, the evidential basis for non-observational knowledge will be in place, in the form of knowledge of what we intend to be doing' (2009a, p. 12). The other is to meet the objection that this is yet another 'two-factor thesis', as Falvey defines it (2000, p. 21), or that one has factorized knowledge of what we are doing as two components: one non-observational component (viz. knowledge of our intentions or what we try to do) and one that is observationally based (viz. knowledge of our actions). To combat these objections Paul argues, on the one hand, that knowing our intentions is something we can know without basing our judgement on evidence and observation; while, on the other hand, she goes on to argue that knowledge of suitable background conditions, though ultimately based on evidence

[12] Velleman entertains the same reading of the notion non-observational knowledge with respect to what we are trying to do (2007, pp. 21–3).
[13] A similar move can be made with respect to the idea that perceptual knowledge, knowledge of our minds, and linguistic interpretation are direct and unmediated: '[w]e are quite comfortable with this idea in the contexts of information-processing subsystems like visual perception or language processing' (Paul 2009a, p. 10). I shall return to this issue below.

and observation, is nevertheless non-observational when we use it to justify beliefs about what we are doing. Let me briefly deal with these aspects of her view.

According to Paul, our knowledge of what we DPA-intend to do is not based on evidence; rather, we may know the contents of our DPA-intentions non-evidentially. Given that our evidence-based knowledge of what we are doing relies on this non-evidentially based knowledge of our DPA-intentions, our knowledge in action will count as non-observational; or so Paul argues. A question of crucial importance is whether such knowledge of our intentions can be found to be a reliable evidential basis and something that 'will dependably be available in the relevant circumstances' and so be 'dependable and accurate enough to serve as the basis for non-observational knowledge of action' (Paul 2009a, p. 12). Paul has after all conceded to Anscombeans that action theory should account for the pervasiveness of such knowledge in core cases of intentional agency. While accounting for the kernel of truth behind Anscombeanism may allow of instances of benighted agency, these cases must be exceptional. The evidential base for the inference must thus be available to a certain degree so that it matches the range of cases for which we do have non-observational knowledge of our own actions. What Paul needs is therefore a story that shows us how we reliably and routinely have such an evidential base at our disposal in virtue of being agents.[14]

As I have argued earlier, we do enjoy a level of epistemic authority when we self-ascribe intentions or make up our minds with a view to action.[15] These speech acts are language-game authoritative and basic so it makes little sense to ask the agent for her epistemic source when she is making these self-referring ascriptions. Paul's gloss on these data is to emphasize the mental act of *deciding* what to do. According to Paul, the act of 'self-ascribing an intention on the basis of making a decision about what to do' is 'a fundamental way of knowing what we intend' because such decisions are 'conscious, relatively discrete mental acts' in contrast with the 'complex, diachronic, heavily dispositional' nature of intentions (2009a, p. 13). On her view, 'we undertake to shape our thought in action in just those ways characteristic of being in the state of intending to Φ' as soon as we have decided to Φ (Paul 2009a, p. 13). Making such a decision is therefore the chief way to make what we have settled on doing – viz. the content of our intentions – consciously available to us. As Paul says, we can know 'at any point what we intend' by 'making a decision about what to

[14] That agents have insight into their own agency has of course been questioned. A ground for scepticism may perhaps be that we 'routinely misidentify the motives behind our decisions, agonize over trying to figure out what it is we want, spend years not knowing that we're really in love, and so on' (Paul 2009a, p. 12).As Paul mentions, research by Richard Nisbett and Timothy Wilson (1977) may suggest that we are quite benighted in our agency. Also see Nisbett and Ross (1980). I have argued before (chapter 2, §3) that these data are far from conclusive.
[15] See chapter 2, §4. Also see Fricker (1998, p. 157).

do' (2009a, p. 13). We therefore will know our intention because this mental act of deciding both settles the matter and is known to us either because it is conscious at the moment, or we may recall previous decisions that have left some mark in our minds and memory. Since we may also assume that we will 'intend the actions we have decided on, and we are entitled to do so', being conscious of these decisions will give us knowledge of our intentions (Paul 2009a, p. 13). From this, knowledge of what one is doing can be arrived at inferentially, or so Paul argues, since we are also entitled to think that we will do what we intend.

The epistemic role of decisions is therefore to ground the 'intermediating between intentional action and knowledge of one's intentions' (Paul 2009a, p. 20). Keeping track of one's decisions may not always be required, of course. We might, for instance, get by with mental shortcuts via one's reasons for acting, memory and so on. However, decisions are there in case we do need to find out what we are up to and why. As Paul explains:

> *If the agent is acting intentionally, he will usually have decided at some point to act, and he can know on the basis of that decision what he intends to be doing. It is no that we must always refer back to the original decision in order to know of our intentions; we employ a variety of shortcuts in keeping track of our goals. Most importantly, we often simply keep track of our reasons for action, or of habitual patterns we have, without referring to anything mental. My suggestion is not that we must always keep track of our decisions in order to know what we intend, but rather that the option is there when other shortcuts fail. And if one does not know what one intends, one may always decide anew.*[16]

On the assumption that these conscious and relatively discrete mental acts are involved in core cases of intentional agency, Paul may therefore explain knowledge of our intentions in action as being non-evidentially grounded in our conscious awareness of our decisions. Invoking the mental act of decisions is thus meant to cover the first challenge to this view: namely, to provide a non-evidentially based evidence-base for knowledge of our own actions. That evidence-base is knowledge of our intentions, grounded as it is in conscious awareness of the decisions-act that provides our fundamental way of knowing our intentions.

For now let us assume that such mental acts of deciding can provide Paul with a reliable non-observationally based evidential base able to serve as a starting-point for the non-conscious, rapid inference and turn our attention to the other challenge to Paul's view.[17] According to Falvey, a 'two-factor thesis' is a view that factorizes knowledge of our own

[16] Paul (2009a, p. 14).
[17] It will come up again later when I turn to my criticism of her view. See §4.

actions into a non-observational and an observational component (2000, p. 21). Donnellan's view is a case in mind when he argues that 'our knowledge of our own intentional actions is complex, that it is divided up, so to speak, into elements of "direct awareness" (...) and other elements to which observation is relevant' (1963, p. 407). Perhaps the Inferential Theory can show how decisions provide non-observational knowledge of our intentions; however, if all we know non-observationally is what we intend to be doing and knowledge of one's actions requires observation we will have what Paul calls 'a disappointing deflation of Anscombe's thesis' (Paul 2009a, p. 16). As you may recall from our previous discussions, Anscombe takes our non-observational knowledge in action to be knowledge of what happens in order to allow that 'when the description of what happens is the very thing which I should say I was doing, then there is no distinction between my doing and the thing's happening' (Anscombe 1957, pp. 52–3). To make justice to that claim one's non-observational knowledge in action must go beyond knowing merely our mental states and DPA-intentions. A two-factor thesis will struggle to do that. The worry for Paul is that her model may be accused of factorizing knowledge in action in this way since on her view the justification of the inferential knowledge of one's action depends not only on knowledge of one's intentions but on observationally based knowledge of background conditions. After all, knowledge of what we intend is not enough to justify knowledge of what we will be doing or are doing. For that we also need to assume that the agent also knows that it is in her ability to perform the intended action, that her circumstances are conducive for the action, that she will follow through on her intention (which requires knowledge of one's history as an agent), and so on (Paul 2009a, p. 15). Since Paul accepts as a legitimate constraint on action theory that it explains our non-observational knowledge of our actions – and not merely knowledge of our intentions, tryings and so on – the question for Paul is whether knowledge of what one is doing can count as non-observational when it appears to rely on all this observationally-based knowledge.[18]

The first half of Paul's reply is to insist that background knowledge need not be observational when it is applied to specific actions even though their evidential source is observational. The reason is, she says, that 'the experience and observation on which the agent's belief is partly based is not experience or observation of the particular action in question' (Paul 2009a, p. 16). Like memory, background knowledge can be called upon to do service as part of one's justification for some *p* without thereby counting as an observation or as lending observational justification to the proposition in question. There is little reason to

[18] There are good reasons for avoiding the 'two-factor thesis' independently of having committed to the kernel of truth behind Anscombeanism. I discussed some of these in chapter 2, §3. See O'Brien's discussion of 'the Dual Component Model' (2007, pp. 171–6).

think that items of general or background knowledge – including events in one's own history that are related via memory – are inferred (consciously or unconsciously) from anything else even if they are ultimately grounded in something observational.[19] In a word, Paul suggests that some element, *e*, can be part of a subject's non-observational evidential base even if *e* in turn is epistemically based on some observational evidence, *e**.[20]

The other half of Paul's answer draws on features of the progressive. Both telic ('I am/was writing a thesis') and atelic progressives ('I am/was walking') are open in the sense that they may count as true without reaching an end-point. In atelic progressives this goes by default since one counts as having walked if 'I am/was walking' is true. In telic progressives, however, it may be true that 'I am/was writing a thesis' even though the history of the world never will see this process brought to an end so that I count as having written a thesis.[21] Paul thinks that these features of the progressive supports the non-observational status of knowing one's actions since 'the central condition' for the truth of the claim that an agent is/was writing a thesis is that she is/was writing with the intention of completing her thesis. This is in part due to the nature of the performance-verbs in these progressive statements, which allow the agent's intention to be 'a powerful determining factor in whether his behaviour indeed comes under the intended description' (Paul 2009a, p. 18). The agent might need, in addition, to have taken some minimal steps in the direction of completing the action, believe that it is not impossible for her to do, must be capable of doing it, and so forth. The crucial bit, however, seems to be the intention with which she is doing it. Given that the intention is the central condition that determines the nature of her action as a Φ-ing, knowledge of what one intends will justify the belief that one is Φ-ing since one will also have 'a default epistemic warrant to presuppose that our intentions will move us to act' (2009a, p. 18n69). You know that you are painting the wall yellow, writing the word 'action', pumping water, writing a thesis, and so on because you know the central condition for determining the nature of these activities as a Φ-ing and the truth of these progressive descriptions. In this way, then, agents can have non-observational knowledge of what happens; or so Paul argues (2009a, pp. 16–8).[22] It thus appears as if Paul is able to meet both the above challenges to her view.

There are three immediate advantages that should be noted right away about Paul's Inferential Theory. The first is that unlike cognitivist accounts of intention that define

[19] For a related point, see Teichman (2008, p. 17n6).
[20] On Paul's account non-observational and non-inferential knowledge are therefore separate notions since it will be possible to infer and come to have inferential knowledge of some *p* from *e* without *e* thereby being observational evidence for the truth of *p* (2009a, p. 16n61).
[21] See Paul (2009a, pp. 16–7). Also see Falvey (2000, p. 22). For the telic-atelic distinction, see Comrie (1976). I also discussed the progressive aspect of agent's knowledge in chapter 2, §2.
[22] The first and third of the above examples (also cited by Paul) are Anscombe's (1957); the second is Davidson's (1978).

intentions as involving, at least in part, a belief or belief-like cognitive attitude, the problem of wishful thinking or doxastic venturesomeness goes away. As we saw in the previous chapter, if an agent intends to Φ but lacks sufficient prior evidence or justification for intention-believing that she will Φ, cognitivist accounts imply that agents form beliefs for which they lack evidence and reasons. Since we frequently form intentions in such cases Cognitivism implies that we often form intention-beliefs without sufficient epistemic warrant and I argued that this is a persistent problem for cognitivists.[23] Paul's account, on the other hand, avoids this problem since there is no belief involved in her notion of an intention in action given that she takes them to consist exclusively of distinctively practical attitudes (DPAs).

The second point is that Paul's model makes the epistemology of agent's knowledge more ordinary by providing it with an evidential base whose reality is independent of the reality of the knowing. On this account, knowledge of what one is doing is brought about by a reliable belief-producing mechanism, a mechanism that relates the agent to states and events that exist independently of their being known that way. According to Paul, it is practical thinking and DPAs – and so not beliefs or belief-like cognitive states – that bring about the object of knowledge. For Paul, the epistemology and ontology of acting intentionally comes apart given that agent's knowledge is, *pace* Anscombe, not 'the cause of what it understands' (1957, p. 87).[24] According to Paul, the epistemic ground for knowing one's action is that one knows what one's intentions are; whereas its ontological ground is a practical commitment on the part of the agent to bring the action about. The result is that the object of knowledge is independent of the reality of the knowing. Agent's knowledge is therefore a matter of tracking truth even though those truths are the effects of one's own mind upon the world (viz. they materialize as a result of our decision). In this respect, the Inferential Theory is 'a demystification of the epistemic structure of non-observational knowledge' since it 'maintains that this knowledge is based on sufficient prior evidence' and so pertains to facts that are independent of the knowing (Paul 2009a, p. 22). This will make knowledge of one's actions less weird as it fits better with how we ordinarily conceive theoretical rationality.[25]

A final advantage is dialectical: with the Inferential Theory on the table we make room for a DPA account that may neutralize the motivation for Anscombean Theories of action and, in particular, for the cognitivist versions of the latter. The cognitivist's ability to 'explain non-observational knowledge has been touted as an advantage to this kind of view, while the *DPA*

[23] See chapter 3 and 4. Also, see Paul (2009a, §2; 2009b, p. 546).
[24] Anscombe attributes the claim to Aquinas (*ST* 1a.2ae.3.5.1).
[25] Paul emphasizes that this is a major difference between her view and Velleman's and Setiya's cognitivist accounts. The point is general: it is the difference between Anscombean Theories of action and non-Anscombean theories, as I defined them in chapter 2, §5.

camp has said very little about it', as Paul says (2009a, p. 22). The sufficiency of Paul's Inferential Theory is thus well-suited to provide DPA views with a story to tell about non-observational knowledge of action so as to avoid the Anscombean conclusion that the relation an agent bears to her actions insofar as they are intentional must be distinctively epistemic.

§3 SETIYA's OBJECTION

In the following section I shall begin my critical scrutiny of Paul's Inferential Theory with an objection raised by Setiya. Although, as I shall argue, his objection to Paul's account has something important to tell us, there are several reasons why it ultimately fails. It is a right step, but in the wrong direction, as it were. Setiya's criticism begins with acknowledging that we need a more 'persuasive way to make the point' (2008b, p. 394) against inferential models than the offhand rejection made by Anscombe and Hampshire on the basis of phenomenological considerations.[26] As we saw, Paul made a reply to this objection saying that appearance of immediacy and spontaneousness need not correspond with the way agent's knowledge really works. Setiya, therefore, looks for a more convincing consideration and he goes on to argue that the mistake with the Inferential Theory is its inability to account for agent's knowledge – or, as he prefers to put it, an agent's non-observational and spontaneous belief about what one is doing – as a necessary feature of acting intentionally:

> *The problem with such an account is that it cannot explain why it should be a necessary truth that doing something intentionally is doing it knowingly*. (...) *On the inferential model, the connection between intentional action and belief could only be a contingent fact.*[27]

Setiya's objection is thus rooted in the conviction that non-observational knowledge (or belief) is a necessary feature of acting intentionally. A contingent, although highly reliable, inferential connection will not suffice as an account of this necessity. After all, for all that the Inferential Theory is saying it is metaphysically possible for an agent to Φ intentionally, know that she so intends while nevertheless failing to 'put two and two together' and infer that she is in fact Φ-ing (Setiya 2008b, pp. 394–5). Inferential models can

[26] Anscombe famously dismissed an account of this kind as 'mad' since it suggests starting from the non-observational knowledge of what one will do in order to go from there to explain knowledge of what one does. Her reason for dismissing such accounts is that the only sense we could 'give to 'willing' is that in which I might stare at something and will it to move' (1957, pp. 50 and 52). Her objection seems to be that it is hard to understand willing something independently of what it is to act intentionally and that an account of agency and agent's knowledge must therefore start from knowledge of the action itself. That may or may not convince people, but it would be, as Setiya says, 'nice to have a more persuasive way to make the point' (Setiya 2008b, p. 394). Also, Paul points out that 'a natural way to make sense of the notion of willing is to think of it as the agent's intention in action' and that one can rely on later work in action theory on the notion of intention to further substantiate such a notion of 'willing' (Paul 2009a, p. 10n41).

[27] Setiya (2008b, p. 394).

at best ensure that we usually have insight into the nature of what we are doing intentionally *qua* being the agent responsible for the action. The alleged fault of the Inferential Theory is therefore its failure to rule out a kind of benighted intentional agency where an agent is Φ-ing intentionally without having a corresponding non-observational knowledge (or belief) about so Φ-ing. When we ask this agent what she is doing or why she is doing it she may be at a loss to know. In Setiya's view, this is exactly 'the possibility that had to be ruled out', and he thinks this is reason enough to dismiss this account (Setiya 2008b, p. 395).

Setiya launches this complaint specifically against an account which we find Grice defending and which is a predecessor of Paul's Inferential Theory.[28] On Grice's model an intention is defined as having a relevant belief about what one will do relative to and because of the fact that the agent is willing$_T$ to Φ. A willing$_T$ is a technical notion defined as an agent's conclusion that Φ-ing is the best thing to do when (1) the can Φ in those circumstances; and (2) the agent will be motivated to Φ rather than something incompatible with Φ-ing if she concludes in favour of Φ-ing. As an objection against this view, however, Setiya is off the mark. After all, Grice's model defines intending to Φ in terms of having a relevant belief relative to the agent willing$_T$ to Φ. In this respect, the Gricean account is what Paul calls a *Weak Cognitivist* account in that it takes a belief to be a constitutive element of intending; as Paul says, 'he [i.e. Grice] thought that believing one will Φ is part of being in the state of intending to Φ' (Paul 2009a, p. 11). Contrary to Setiya's assertion, the connection between intentionally Φ-ing and having a relevant belief is therefore not contingent on Grice's model. Even though Grice's model faces other problems, it does not fall prey to Setiya's objection.

Setiya's objection may seem better suited to Paul's Inferential Theory since that account explicitly asserts that 'the agent will *tend* to believe, partly on the basis of his intention to Φ, that he is in fact Φ-ing' (2009a, p. 14 [original emphasis]). A tendency may, as a matter of course, be disrupted in certain circumstances and Paul has no apparatus for denying that these actions would still count as intentional. If that is the possibility Setiya thinks we have a general reason to rule out, the Inferential Theory could therefore be dismissed.[29] The problem with Setiya's objection, however, is that whether or not agent's knowledge (or belief) is a matter of necessity is what, by and large, one disagrees over along this fault-line. Both sides agree that non-observational knowledge (or belief) is a pervasive phenomenon in intentional agency and that it cries out for explanation by action theory.

[28] See Grice (1971, pp. 278–9) and Paul (2009a, p. 11). I discuss Grice's position in an appendix to this chapter.
[29] Setiya's dialectics thus resembles the *Bottom Up* strategy against 'normative interpretation of the recognition and adoption of reasons involved in acting on them', which begins with general constraints on a theory of action in order to show that these constraints rule out the relevant interpretation (2007b, p. 22). I discussed this strategy above, chapter 1, §8.

However, Paul knows that as soon as one accepts the connection as a matter of necessity the journey will begin towards a theory where an agent's knowledge in action pertains to the nature of intentional agency.[30] Paul must therefore positively deny the existence of such a necessity. Consequently, she is quite open to the idea that one may act intentionally and yet thoughtlessly and benightedly, as in the case in which the agent is 'zoning out' by

> *... driving home on autopilot.* (...) [*T*]*he driver takes a certain route because it is the best way home, uses his turn signal in the appropriate places for reasons of safety, raises the garage door in order to park his car, and so forth. To express the fact that his turning and signalling are purposive, have a means-end structure, and are performed in response to reasons, it is natural to say that he turns and signals intentionally. If his driving is habitual enough, however, he may have* no belief *that he is doing these things, and might only be able to discover that he is doing them by observation.*[31]

According to Paul, such a case indicates that the actions we group as intentional and explain by reference to an agent's reasons need not converge with the actions we must have non-observational beliefs about. Paul refers to research on the application of the adverb 'intentionally' which apparently shows that we use it for a variety of purposes.[32] Our usage of this adverb is, among other things, meant to 'capture the fact that an episode of behavior had a goal-directed, means-end structure and exhibited responsiveness to reasons' even though the related non-observational belief is absent, she says (2009a, p. 5). Setiya's general constraint is therefore not something whose force Paul must accept; in fact, cases like the zoned-out driver may count against the universal validity of Setiya's general constraint.

There is reason to be cautious about data from such usage of the adverb 'intentional', though. All it might show is that there is a complex relationship between what it is to act for reasons and the behaviour that subjects judge to be intentional under certain circumstances. These judgements are presumably sensitive to pragmatic concerns – such as issues of praise and blame, responsibility, and so on – so it might be questionable whether one really is tapping into the relevant sensitivity of the concept of an intentional action rather than other

[30] I argued (chapter 2, §5) that this follows from general methodological principles which require that we explain necessary connections between *F*s and *G*s (for any *F*s and *G*s) by appeal to their nature. Also see Setiya (2007b, p. 27n10; 2008b, p. 395).

[31] Paul (2009a, p. 5). Paul also mentions cases of mindless expert-coping (zoning in). A case in mind is the chess Grandmaster discussed by Hubert L. Dreyfus, who may play lightning chess, thereby displaying a way of coping that depends 'entirely on perception and not at all on analysis and comparison of alternatives' (2005, p. 53). In particular, Dreyfus argues, so coping is a non-conceptual activity. If so, a Grandmaster copes with his situation in a way characteristic of expertise by zoning in and doing without the governance of a belief-system. Also see Dreyfus (2007a; 2007b). For defence of the idea that expert-coping is nevertheless conceptual, see McDowell (2007a; 2007b).

[32] Paul refers to Joshua Knobe (2003). Knobe (2003; 2007) and Knobe and Sean Kelly (2009) argue that since ascriptions of intentional actions are sensitive to concerns about moral or ethical responsibility this will have a bearing on the concept of intentional actions.

possible concerns.[33] In general, one should also be very cautious of inferring conclusions about how they think about intentional actions – or the meaning of this concept – from the way subjects talk about and label activities on questionnaires. People often talk as if books and libraries contained 'knowledge', that the plane on autopilot 'knows' where it's going, or that the malfunctioning laptop is in 'a cranky mood' this morning. Whether they take this mind-talk at face value is another question. Finally, we should note that even if we did take these results at face value they could count against the belief-condition on acting intentionally without thereby ruling out other plausible epistemic conditions on intentional action.

In my view, this all goes to show that in order to settle these matters the merits of the Inferential Theory must be tested on what Paul takes on as an explanatory commitment: namely, to account for the non-observational knowledge an agent has of her actions as a result of acting intentionally in core cases of intentional agency. A general constraint on action theory cannot be used to rule against some theory in favour of another if the constraint itself is on dispute between the two theories. To get further progress from here, one could perhaps 'address the deeper sources,' as Setiya suggests in his reply to Paul (Setiya 2009a, p. 131), for supporting the truth of the necessity-claim. That would necessitate a more systematic discussion and Setiya seems to think it would count to his advantage given his systematic arguments elsewhere in favour of the claim that intentions are beliefs: viz. that this hypothesis gives the best account of instrumental reason, that it explains the impossibility of believing at will and can account for knowledge of what we intend.[34] Of course, if those arguments work, we will have reason to think that non-observational beliefs really are necessary elements of acting intentionally. But, as we have already seen, his position is deeply problematic.[35] The primary battleground must therefore be the kernel of truth behind Anscombeanism – viz. that agents acting intentionally paradigmatically have non-observational knowledge of what they are doing and that their knowledge appears to be cognitively spontaneous and immediate.

In the next three sections I shall raise problems for the Inferential Theory even on this count. First, I shall question Paul's evidential base for the inference (§4). After that, I will question the inference, as Paul conceives it (§5). Finally, I will take the opportunity to reflect about these worries and how they may point towards an Anscombean approach (§6).

[33] See Eric Wiland (2007). Wiland suggests that subjects seem sensitive to pragmatic considerations of blameworthiness. It is therefore difficult to judge whether Knobe is tapping on to the category of intentional actions, or actions for which agents are blameworthy – which may or may not overlap – when he finds cases judged as intentional actions but where the belief-condition fails to hold.
[34] See respectively Setiya (2007a; 2008a; and 2011).
[35] See chapters 3 and 4.

§4 PROBLEMS with the EVIDENTIAL BASE

In the following section, I want to raise certain worries pertaining to Paul's evidential base for the inference that is meant to non-observationally ground our knowledge of our own actions. I shall focus on two related worries: (1) that even though the conscious act of deciding what to do plays a role in our life as agents, it cannot ground non-observational knowledge of our intentions the way Paul thinks; and (2) that even if we were to grant that Paul can account for knowledge of our intentions, knowledge of our actions will not follow.

According to Paul, the non-observational basis for our knowledge in action is that we can know what we intend to do without evidence or observation because our decisions are 'conscious, relatively discrete mental acts', the occurrence of which entitles us to believe of ourselves that we intend to do what we have decided to do (Paul 2009a, p. 13). Admittedly, it is never entirely clear exactly what a decision is on Paul's account, she describes them as mental acts of a kind which, insofar as they occur, allow us to 'settle the question for ourselves of what to do' (2009a, p. 13). It is clear that she thinks of them as different in nature from both intentions and practical judgements. On the one hand, she explicitly contrasts decisions with the 'complex, diachronic, heavily dispositional' nature of intentions (2009a, p. 13), while, on the other hand, rejecting views that identify decisions as practical normative judgements. In her view, it is not 'a general constraint that we must decide what to do in light of what we take ourselves to have sufficient practical reason to do' (2009a, p. 13). Unfortunately, Paul does not spell out much of the details of this story, nor does she provide any arguments for making the controversial choice that this notion is non-evaluative and independent of practical judgements. As we have seen already, many philosophers disagree with the idea that the actions chosen by virtue of such non-evaluative decisions can be intentional actions at all.[36] According to Paul, her non-evaluative notion of a decision is, for instance, in disagreement with Shah's claim (2008) that an 'explicit deliberative question whether to Φ must be settled by answering the normative question of whether one ought to Φ' (Paul 2009a, p. 13n51). Paul's position will also go against all the theorists who endorse principles such as the *Guise of the Good*, or against those who invoke some constitutive normative aim or principle that agents must aim for insofar as they act intentionally and for reasons. Paul might, however, have in mind a notion such as Bratman's decision to treat something as one's reason, which he defines as treating something as 'end-setting' and 'potentially justifying' (1996, p. 198). As Bratman argues, his notion allows room for cases of

[36] See chapter 1, §7.

practical irrationality where the agent is at odds with her best reasons or deeper values and principles. A decision to treat some *R* as reason-giving and then act on it will be enough for the agent to act for this reason; however, at the time of action one's decision need not cohere with one's general policies, values and principles such as they occur in the agent's own practical judgements. A drug addict may, for instance, cave in, as it were, and resentfully treat her desire for a drug as her reason for taking it *this time*. In that case, 'one's will would be divided,' as Bratman says (1996, p. 201), and the addict would act for a reason she fails to approve of and identify with. When the addict caves in and treats her desire as a reason she nevertheless treats it as end-setting and potentially justifying this time. This notion, far from being obvious, is at least a possible candidate notion of a decision.

Suppose that some such notion can be had: what I take to be a serious worry about Paul's framework is its idea that such decisions can play the central epistemic role in our agency that Paul ascribes to them. In particular, I doubt whether the states an agent has – by virtue of which she decides to do something intentionally, treat some fact as reason-giving, form intentions and so on – will need to involve such conscious, discrete mental acts. Rather, I think there is a range of cases where no such decision need occur, while the agent is nevertheless able to perform intentionally and have non-observational knowledge of what she is doing. We have silently formed intentions, as Hampshire says:

> *When asked what I am going to do on a certain occasion in the future, I may find myself giving a definite answer without the least hesitation and without ever having reviewed the question before; this, I now recognize, had all along been my intention, just as I had all along believed without question that a certain statement is true. Any human mind is the locus of unquestioned and silently formed intentions and of unquestioned and silently formed beliefs.*[37]

Settling on a course of action and forming an intention – either by doing something directly with an intention in action, or by forming an intention for the future – appears to be a process that is more like the characterization Paul gives of the nature of intentions: namely, as involving states that are like our intentions in being complex, diachronic and heavily dispositional. Settling for actions may be something that dawns on you and comes about gradually or over long stretches of time. It often involves complex cooperation between your will, judgements, beliefs, desires, and, at least sometimes, conscious decisions. Discrete, mental acts such as Paul's decisions seem to play a less central role in such processes.

[37] Hampshire (1960, p. 101).

An expression of the idea that the mental act of decision-making plays a more minor role in our active life is Frankfurt's unrelenting reminder that there are limits to what we can decide at will. As Frankfurt points out, an agent can call into the 'vasty deeps' of herself for her will to be as she has decided to just to find that she is unable to make her will follow suit 'when the chips are down' (1992, p. 10). 'We do not control, by our voluntary command, the spirits within our own vasty deeps,' Frankfurt goes on to say, and so '[w]e cannot have, simply for the asking, whatever will we want' (1992, p. 10).[38] Paul claims to acknowledge Frankfurt's insight; however, I do not think she appreciates its philosophical consequences. What Frankfurt's observation reveals, is that our intentions in action – and, in consequence, our intentional actions – may often be rather independent of our conscious decision-making. Intentions really are complex and they may often be 'like beliefs', as Hampshire points out, in the sense that they are 'not always and necessarily the outcome of a process of thought or of a datable act of decision' (1960, p. 101). Rather, intentions will often 'effortlessly form themselves in my mind without conscious and controlled deliberation,' as Hampshire goes on to say (1960, p. 101). In a way, Paul has put the cart before the horse. Intentions are meant to play a pivotal role in actions that are voluntary and intentional. It is therefore intentions that allow an agent a degree of control. Decisions are mental acts over which we have control in this way. If anything, it is these mental acts that seem to depend on our ability to act as we intend rather than the other way around. A decision is a conscious, voluntary mental act; forming an intention, on the other hand, is more like the process of forming a belief.

To appreciate this distinction – and its philosophical importance – it is worthwhile comparing the relationship between the mental act of deciding and intention-forming with how judgements stand to belief-formation. Judgements are by many philosophers seen as the fundamental way to form a belief and this is presumably similar to how Paul takes decisions to be the fundamental way to form intentions. However, as some philosophers have pointed out, judging that *P* is true does not always pave the ground for forming the relevant belief. Just consider the following case from Christopher Peacocke:

> *Someone may judge that undergraduate degrees from countries other than their own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that*

[38] As Plato points out in the *Republic*, the relationship between spirit (will) and reason calling it to heel is like a dog with respect to the shepherd (*R* 440d). As any shepherd knows, the dog's training determines whether it is obedient or not when its master calls. The same holds for the will: our control here is never total, although proper training will presumably increase it.

> *effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all.*[39]

To appreciate what is going on here we must recognize that judgements and beliefs are different ontological categories. Making a judgement is probably best rendered as a mental action – or a *cognitive mental act* – which is the act of putting a proposition forward in one's mind as true. That operation is an act or a process rather than a state because it is an event and something we do for a reason. Moreover, it is cognitive because this event or action is to present something to one's mind as being true. A belief, on the other hand, is a mental state the occurrence of which represents something as true. We may follow Shah and Velleman in characterizing this as a *cognitive attitude* (2005, p. 503). Such belief-states will probably normally occur as a result of judging; however, in a given case judging may fail to lead to the formation of the relevant belief, as we saw in Peacocke's case.[40] Moreover, beliefs may, as Hampshire pointed out, form themselves effortlessly and without judgements.

The same reasoning applies to decisions and intentions. Their difference in character – in particular the contrast between the 'conscious, relatively discrete mental acts' of decisions and the 'complex, diachronic, heavily dispositional' nature of intentions (Paul 2009a, p. 13) – is an expression of the fact that they too belong to different ontological categories. Decisions are what we may call *practical* mental acts – or alternatively *conative* mental acts – whereby an agent puts forward some content or proposition as a pattern to follow. As opposed to the cognitive way of representing some proposition – viz. where one treats the content as a fact and *as true* – the practical mental act is the act of representing something as *to be made true*.[41] Intentions are not mental acts like that; rather they seem more similar to states and attitudes in being something whose occurrence represents something as to be done. The latter may often come about as a result of the former; but, just as beliefs stand to judgements, intentions may not obey one's decisions nor do they need them in order to obtain. 'Any human mind,' as Hampshire says, 'is the locus of unquestioned and silently formed intentions and of unquestioned and silently formed beliefs' (1960, p. 101).

[39] Peacocke (1998, p. 90).
[40] As Quassim Cassam points out in his discussion of this case, belief-formation may be influenced by other forces, say non-rational forces such as 'self-deception, prejudice and phobias' (2010, p. 82). An unshakeable prejudice may simply hinder me from taking my own reasonable judgements on this matter to heart, as it were. As Cassam says, I may 'mentally affirm that undergraduate degrees from countries other than my own are of an equal standard to my own and yet my attitude towards this proposition is not the attitude of belief, as evidenced by my hiring decisions and letters of recommendation' (2010, p. 82). But even if one's beliefs fail to track one's judgements in these cases, there is no need to think that one thereby fails to know what one reasonably judges to be true. The deviance between judgements and beliefs thus opens a way for knowing without believing. Also see Shan and Velleman (2005, p. 508).
[41] For a similar distinction between cognitive attitudes and conative attitudes, see Velleman (1992a, p. 105).

The possible recalcitrance of the will with respect to the agent's call into her own 'vasty deeps' is therefore an accurate image of how intentions may stand with respect to the mental act of deciding something. It might be, as Aristotle says, that we engage in active deliberation when 'the event is obscure, and with things in which it is indeterminate' (*NE* 1112b9-11). It even seems plausible that decisions, choices and chosen actions need to have 'involved reason and thought', as Aristotle goes on to say (*NE* 1112a15), and that intentional agency therefore displays 'rationality in operation' (McDowell 2008a, p. 211). However, it is clearly not the case that 'what has been decided' – viz. what settles the course of action – must have been 'decided on by previous deliberation' involving the conscious act of deciding what is to be done (Aristotle *NE* 1112a15). The intentions will not always need such discrete and datable mental acts in order to be formed and give rise to intentional actions; nor will our will always obey these decisions. Such decisions are therefore, on the one hand, not sufficient ground for knowledge of what our intentions are, while, on the other, we seem perfectly able to non-observationally know our intentions without them. In neither case, will the mental act of deciding what to do – viz. as a conscious, discrete mental act – play the epistemic role cut out for it by Paul's Inferential Theory by 'intermediating between intentional action and knowledge of one's intentions' (2009a, p. 20).[42] Paul has therefore not provided a reliable evidential basis for non-observational knowledge of our intentions in our decisions.

Note that by raising this worry, I do not disagree with Paul that there is something about the way we make up our minds and settle for some action when the action is intentional that makes it available to us. I argued in chapter 1 that certain unwitting decisions do not give rise to intentional actions even though they guide our behaviour. Consider the following case:

> The UNWITTING DECISION: *Suppose that I have a long-anticipated meeting with an old friend for the purpose of resolving some minor difference; but that as we talk, his offhand comments provoke me to raise my voice in progressively sharper replies, until we part in anger. Later reflection leads me to realize that accumulated grievances had crystallized in my mind, during the weeks before our meeting, into a resolution to sever our friendship over the matters at hand, and that this resolution is what gave the hurtful edge to my remarks. In short, I may conclude that desires of mine caused a decision, which in turn caused the corresponding*

[42] Of course, if an agent regulates well and puts herself in order – in a way reminiscent of how Plato defines the just person (*R* 444d) – a conscious decision will settle the matter and make sure that an intention to Φ is formed, while this in turn will give rise to an intentional performance. However, an agent's will may not always be obedient and her conscious decisions may be powerless and futile. The critical point against Paul is therefore that the latter kind of agency – indeed, the more common one – fails to locate the mental act of decision in a place in one's cognitive economy where it can play the epistemic role it is meant to do on the Inferential Theory.

> *behavior; and I may acknowledge that these mental states were thereby exerting their normal motivational force, unabetted by any strange perturbation or compulsion.*[43]

In this case, an agent settles on Φ-ing and carries out this decision. Yet too much is going on behind the agent's back for the action – viz. to sever the friendship – to be a case of intentional agency.[44] The agent is surely doing something intentionally in this case; however, the intentional action is not to sever the friendship since that decision is only something the agent realizes later after having reflected on his own, surprising behaviour. The basis for this interpretation of the case is epistemic: its agent has to find out what he is doing and why by inferring from considerations about his behaviour. The case thus illustrates two things: first of all, it displays the widespread assumption that an intentional agent should not need to make such discoveries about himself. Secondly, it shows that resolving to do something is not always transparent to its agent. According to Bratman, '[c]oncerning any mental act or occurrence, even a decision, one can raise the question of whether or not the agent identifies with it' (1996, 193). In light of Velleman's case, one can ask of any mental act or occurrence, even a decision, as to whether the agent knows of it in such a way as would reveal intentional agency. Settling what to do is not always self-reporting. Yet there is an assumption that settling what to do in core cases of intentional agency is a process that makes the content of the decisions and intentions available to its agent. In a word, we assume that an agent must have taken due account of her action insofar as it is intentional.[45]

Paul is therefore perfectly in her right to assume that the states that play the relevant functional and monitoring role in mediating intentional agency are, at least in most cases, epistemically available and known by their agent. An agent relates to her intention in action by being, as Hornsby says, 'especially well placed to become aware of it' (2005, p. 121). Where I disagree with Paul is in thinking that decisions – as conscious mental acts – will be in the right place to account for these epistemological features of our intentional agency. Rather, I go with Peter Strawson in thinking that 'a man may very well know that he is going to do (...) and do it, without having raised the question to himself at all, without indeed having thought about it' (Strawson 1954, p. 179).[46] The mental act of deciding something is not always around when we are acting intentionally; yet we may still have non-observational knowledge of what we are thereby doing or what our intentions in action are. On the other

[43] Velleman (1992b, pp. 126–7).
[44] We found similar cases such as the story of *Freud's Inkstand* and the quitting *Law Student*. See chapter 1, §6.
[45] Unwitting decisions may have serious consequences even though they fail to produce intentional actions. An example is Isabel Archer, the heroine of Henry James' *The Portrait of a Lady*, who bitterly characterized her husband at a crucial point in their increasingly claustrophobic marriage by saying: '[H]e was of course unwitting; he himself had not taken account of his intention.' (James 1907-9, p. 436)
[46] Strawson is reaching this conclusion in part by interpreting Wittgenstein's discussion of action (*PI* §611–60).

hand, their occurrence will not always be sufficient for justifying beliefs about our intentions given the possible recalcitrance of our will. For all I have argued, Bratman might still be right in thinking that we need to decide to treat something as our reason in order to act intentionally (1996, p. 198); but if that is true, these reason-giving decisions cannot be equivalent to the conscious mental acts that Paul talks about. We therefore still lack an explanation telling us why such intentions do report themselves to their agent insofar as one is acting intentionally. With those words let us move on to my second worry, namely, that knowledge of our intentions and our previous experiences will not justify beliefs about our actions.

According to O'Brien's discussion of inferentialist models of non-observational knowledge in action, their chief problem is in justifying beliefs about our actions (O'Brien 2007, pp. 176–7). Recall that the knowledge an agent has of her actions is on Paul's Inferential Theory supposed to be mediated and indirect by being inferentially justified by knowledge of our own intentions plus knowledge of suitable background conditions. We must therefore assume that the agent has knowledge of her ability to perform the intended action – which, among other things, includes knowing that 'my motor system is in working order' (O'Brien 2007, p. 177) – that one's circumstances are conducive for the action, that one will follow through on one's intention by knowing one's history as an agent, and so on (Paul 2009a, p. 15). In order for this view to, on the one hand, not collapse into what Falvey defines as a 'two-factor thesis' (2000, p. 21), while avoiding invoking an Anscombean kind of spontaneous or immediate knowledge, it is important that one's knowledge of what one is doing gets 'no further epistemic support (...) by my doing anything' (O'Brien 2007, p. 177). It is thus important for the Inferential Theory that warrant for the beliefs we have about what we are doing intentionally depends on nothing but the inferential relationship between that belief and the knowledge we have of our intentions and suitable background conditions.

A question O'Brien raises is whether that justifying relationship holds in general. Indeed, as she avers, 'all too often, even if I could do what I intend to do, I do not' (2007, p. 177). Intending to Φ plus knowledge of these background conditions may in a range of cases at most ground a justified hypothesis that one will Φ or is Φ-ing. To reach the level of epistemic warrant so that we also know what we are doing would seem to require further epistemic support. Of course, from a naïve perspective epistemic support appears to come from the action itself. As O'Brien says, it 'seems that I know that as a result of my raising my arm' rather than being inferred from 'my knowing that I intend to raise my arm' (2007, p. 177). But for that to happen the additional support must come by observation or

spontaneously from the fact that one is doing it; and the Inferential Theory does not leave room for any of these options. After all, if it needed to invoke observation, it would collapse into a version of the two-factor thesis we briefly discussed above. Moreover, were it to require spontaneous knowledge it would succumb to an Anscombean approach. Since the Inferential Theory cannot have either of that, it therefore seems to have a problem in justifying the beliefs we have about our own actions so as to make them knowledgeable.

One thing is certain: we do not appear to have merely justified hypotheses about what we are doing insofar as we are acting intentionally. In the core cases we know what we are doing (and why) and we are not merely having fairly good justification that falls short of being knowledge. There would then be no reason why such justified hypotheses could not be met by challenges of the kind 'How do you know?' and I have already argued that it is a very odd challenge to level against someone who declares what she is doing intentionally or what she will be doing. In fact, I argued that if such challenges to an indicative speech act were well placed, they would undermine a reading of this speech act as expressing one's intention.[47] This consequence of the Inferential Theory seems to make such challenges legitimate. It will therefore not be able to explain why they intuitively cancel a reading of the speech act as an expression of intention. The epistemic worry raised by O'Brien is therefore support for 'the naive thought', as O'Brien says, that 'we know what we are doing *because* we are doing it' and not because we know something else that indicates the presence of an action of ours (2007, p. 182). O'Brien's epistemic worry will thus put more pressure on the sufficiency of the evidential basis that is offered by the Inferential Theory.

I therefore conclude that we have reason to be worried about the evidential base that is meant to ground knowledge of our own actions. On the one hand, I have argued that Paul's account fails to explain why we have non-observational knowledge of our intentions insofar as they guide, accompany, or give rise to our intentional actions. On the other, I have argued that even if we can overcome this difficulty, knowledge of our intentions plus knowledge of certain background conditions is not enough to support the beliefs we have about our own actions so that they amount to knowledge. Let me now turn to the inference itself.

§5 A PROBLEM with the INFERENCE

As I said above, it is important for Paul that she can explain the kernel of truth behind Anscombeanism. A significant part of that task is to account for the fact that my knowledge of

[47] See chapter 2, §4.

my own actions appears to be non-inferential, direct and immediate. 'It does not *seem* to me that I know I am raising my arm as a result of inferring that I am from my knowing that I intend to raise my arm', as O'Brien says; 'it rather seems that I know that as a result of my raising my arm' (2007, p. 177). The inferentialist clearly cannot take these phenomenological data at face value; however, they may perhaps explain why our knowledge occurs directly and immediately despite coming about as the result of inferences. Paul tries to accommodate these data by arguing that our knowledge appears immediate and direct because the 'evidence-based information processing' whereby we know our actions 'take place rapidly and automatically at a non-conscious level' (Paul 2009a, p. 6). In what follows, I shall question the adequacy of this explanation. In particular, I will focus on the differences there are between the knowledge we have of our actions and the cases where it is theoretically fruitful to assume that, despite appearances to the contrary, one's knowledge is inferentially based.[48]

In support of the idea that knowledge of our actions is inferential but nevertheless appears immediate because the inferences are rapid, automatic and non-conscious, Paul appeals to the explanatory success of similar models in other areas of knowledge, such as visual perception and language processing. Paul claims that '[w]e are quite comfortable with this idea in the contexts of information-processing subsystems' for these domains of knowledge (2009a, p. 10). The inferences that ground beliefs about our actions may similarly be rapidly and non-consciously 'derived from prior states that (ideally) evidentially support those beliefs' (2009a, p. 10). If this move works, knowledge of our actions may appear immediate and spontaneous – just like visual knowledge, or the knowledge we have of the significance and meaning of someone's speech acts – while nevertheless being 'evidentially based on our knowledge of what we intend to be doing' (Paul 2009a, p. 4). I take it that Paul's suggestion is not based exclusively on an analogy between these different ways of knowing; nevertheless, I think there are some important differences here that should make us suspicious of the whole idea that knowledge of our own actions is inferentially based in this way.

I shall focus on the case of linguistic communication and thus leave visual perception aside. The phenomenology in linguistic communication appears to be just as direct and immediate as it is in visual perception and knowledge of our own actions. In the case of interpretation, as Fricker says, a competent speaker 'hears the utterance not merely as sound, but as the speech act that it is' and the 'representation of meaning is not phenomenologically

[48] The points in this section owe their existence to the resistance that Trine Antonsen gave me when I first considered conceding to Paul the idea that these inferences could be rapid, automatic and non-conscious.

distinct from her hearing of the sounds' (2003, p. 325).[49] An interpreter 'hears its meaning in the utterance itself' and 'experiences it as a semantically laden event' (Fricker 2003, p. 325). Despite this 'fact of phenomenology', however, our 'effortless enjoyment of [such] an understanding-experience' may require that a 'very complex processing must take place subpersonally to enable this' (Fricker 2003, p. 325). In particular, it may turn out that the interpreter has and unconsciously invokes inferentially 'metalinguistic beliefs about word meanings' and how they combine to generate sentence meanings (Fricker 2003, p. 365).[50] Experiencing immediately and directly that some speaker's utterance, *U*, expresses her deep concern for Kaddafi's regime in Libya is therefore not decisive evidence against a model of linguistic interpretation wherein this meaning and significance is non-consciously inferred from knowing the word meaning of words in *U*, rules for generating sentence meaning by combination, taking account of context sensitivity and so on.

Hornsby makes similar observations in the case of speech production since 'the phenomenology of hearing has a counterpart for speaking' (2005, p. 112).[51] Our knowledge of what we are saying appears just as immediate and non-inferential as hearing what someone else is saying. As Hornsby says, 'it seems (as I shall put it) that you directly *voice your thoughts*' (2005, p. 112). However, it does not follow that we can simply rule out the possibility that an agent engaged in speech production 'may invoke knowledge unconsciously', as Hornsby points out (2005, p. 120). 'What cannot be taken for granted', Hornsby goes on to say, is 'that our experience of speaking informs us of what knowledge we exercise as speakers' (2005, p. 120). Being able to knowingly voice thought *P* in language, *L*, by virtue of uttering *U* in some context *C* may well be an ability that appears to give immediate and non-inferred knowledge of the fact that one is voicing *P* rather than, say, *Q* or *R*; however, it may nevertheless be undergirded by rapid, non-conscious inferential processes.

What will decide these issues, however, is whether an inferential base is theoretically required in order to explain the linguistic and semantic competence one enjoys as a competent speaker of some language, *L*. I have no intention of deciding this issue for the case of linguistic communication. What I am pointing out here is only that there might be theoretical gains in ascribing such rapid, non-conscious inferences in order to account for someone's ability to understand and produce utterances in *L* when they are fully competent speakers of *L*. A reason why some such thought may be intuitive with respect to language is that linguistic

[49] These points apply to one's *home* language, which 'servers you as a mother tongue ordinarily serves a person' (Hornsby 2005, p. 111).
[50] Fricker opposes the latter model in favour of the idea that in the basic case our understanding of what others are saying is grounded in the enjoyment of 'correct understanding-experiences of the meaning of heard utterances' (2003, p. 328).
[51] Hornsby explicitly mentions Fricker's observations with respect to interpretation and hearing (Hornsby 2005, p. 112).

competence may seem to require knowledge of several complex components, rules of grammar and combination, and so on.[52] There might therefore, presumably, be theoretical needs for making the hypothesis that undergirding one's abilities as a competent speaker are the rapid, non-conscious inferences one makes from a body of largely tacit and non-conscious linguistic and semantic knowledge.[53] The question is whether similar considerations apply to the knowledge we have of our own actions. I think not. Let me briefly say why.

The difference between what one is supposed to infer knowledge of one's actions from and the inferential base serving linguistic production and interpretation is striking, however. In action the agent knows what her intentions are. There is similarly a fair amount of transparency with respect to the background conditions an agent must know. In the case of linguistic communication, on the other hand, the required knowledge or inferential base is largely hidden from a speaker's view. It is tacit, implicit, or non-conscious knowledge or something a speaker manifests in her linguistic behaviour rather than having conscious access to it. In fact, she might not even know or recognize an adequate description of this alleged knowledge – say of some syntactic or semantic rule – when it is presented to her even though her behaviour clearly reveals some reliable sensitivity to it. Just to take an example: a speaker will always treat an ambiguous string, such as 'I saw her duck and swallow', as combining 'duck' and 'swallow' either as a noun-phrase conjunction or a verb-phrase conjunction:

(1) I saw her $[_{NP}[_{N}$duck] and $[_{N}$swallow]]
(2) I saw her $[_{VP}[_{V}$duck] and $[_{V}$swallow]]

What one will never do is treat or hear (1) as four-way ambiguous, or as follows:

(3) ?I saw her $[[_{N}$duck] and $[_{V}$swallow]]
(4) ?I saw her $[[_{V}$duck] and $[_{N}$swallow]]

The explanation for that is, as Barry Smith points out, 'that the internalized grammar that shapes and conditions our conscious experiences of speech respects a co-ordination constraint' and this constraint restricts us to conjoining expressions of only the same syntactic category (Smith 2006, p. 959).[54] Speakers do not know this rule; however, they abide by it with striking reliance thereby revealing their competence with respect to this rule. Similar considerations apply to many of the other things we know when we know a language, in

[52] As Dummett argues, knowledge of a language is not a '*merely* practical' capacity (1978; 1991, pp. 93–7; 1993, pp. x–xi).
[53] Both Hornsby (2005) and Fricker (2003) are opting for non-inferential models of respectively speech production and interpretation.
[54] For both the examples and discussion I rely on Smith (2006).

particular the knowledge on which we may be said to draw when we engage as speakers and interpreters. We rely on the fact that some word 'cat' means or encodes the concept *CAT*, for instance; however, we may not notice that it was this word that we used to encode this concept in order to say what we wanted to. Similarly, one may often find oneself remembering all the things someone said without remembering the words used.

That is a striking difference, however, to the way we know the alleged inferential base that Paul takes to ground knowledge of our own actions. Neither our intentions nor the background conditions mentioned by Paul are things we have little or no access to. At least, this holds with remarkable effect of the knowledge we have of our intentions in action. We know what they are insofar as they guide our intentional behaviour and agents are able to recognize – and in most cases even able to provide – the relevant description of their intention in doing something. Our knowledge here does not seem anything like the knowledge we have of word-meanings and combination rules when we produce or interpret utterances within our own language. No one would suggest that this is something we know tacitly, non-consciously or implicitly; rather, knowledge of our intentions counts as one of the paradigmatic things we are, to use Hornsby's words, 'especially well placed to become aware of' (2005, p. 121).[55] Its agent need not necessarily be consciously aware of her intentions; but if she were to attend to, or to reflect upon, what her intention is this is something she could find out about herself with little or no effort. When asked I may, as Hampshire points out, 'find myself giving a definite answer without the least hesitation and without ever having reviewed the question before' (1960, p. 101). This feature is in stark contrast with how the alleged inferential base that grounds our linguistic performances where our knowledge may be hard to come by, surprising and something it is hard to recognize the truth of. In contrast, the only thing that appears to be in the dark when it comes to the knowledge we have of our actions is the alleged fact that we reach this knowledge via an inference from knowing our intentions. On Paul's model both the decision and the intention appear to be things to which an agent has fairly good access in virtue of being the agent who acts. We thus seem to know in a relatively conscious way both what our premises are and what our conclusion is; the only thing missing is the fact that we infer what we are doing from what our intention is plus background conditions.

Another related difference is that the inference that is allegedly involved in knowing our actions seems to carry much less an informational load in its inferential process compared with the information processed by the information-processing subsystems that ground our

[55] Hornsby speaks of the knowledge we have of our own actions; however, the same point seems to apply equally well with respect to knowledge of our intentions, whether they be future-directed or intentions in action.

competence in speech. As cognitive scientists can tell us, the inferences and information-processing we may access require more processing and cognitive resources than if similar processes occur at a rapid, automatic and non-conscious level. Indeed, our nature and cognitive systems may be imposing certain constraints on us – in particular given how fast or slow our reasoning is in the given area – furnishing evidence for the idea that some reasoning process is non-conscious. Linguistic processing is a case in mind. It appears to be complex and involve much fast and frugal processing, so there might be a theoretical motivation for modelling it as a non-conscious inferential system. There clearly are many such processes involved as we move about and perform complex intentional activities. Riding my bike, for instance, will involve a plethora of processes which I am in no need of consciously attending. The non-conscious nature of these processes may allow this conscious 'I' to rather concentrate on the final pages of his thesis, enjoy the landscape, or think about what it would be good to have for dinner tomorrow. However, there are no analogous reasons why the information-processing that takes me from knowledge of my intentions to knowledge of my actions should be non-conscious and automatic. After all, the inferential step seems fairly straightforward given how Paul has characterized the premises. If her story adequately presents how we come to know our actions, it would require only a minimal amount of self-awareness and inferential skills to become consciously aware of the inferential steps that are supposed to be going on at an automatic and non-conscious level. This inferential process will not stretch our limits as conscious information-processing systems. There thus appear to be no information-processing constraints arising from the nature of our cognitive systems to explain and motivate why this particular inference should be so intractably hidden from view.

What these two points indicate is that there is no theoretical motivation similar to that of our linguistic performances that supports invoking non-conscious and automatic inferences to ground knowledge of our own actions. The inference is not complicated and would not bear too high a toll on our cognitive system if we assumed it to be conscious; nor does it contain information that we will fail to have conscious access to if we reflect on our situation. It thus appears as if the only motivation for suggesting that this inference is non-conscious is that it would enable the Inferential Theory to adequately respect the phenomenology. Paul is right in thinking that we simply cannot assume that the phenomenology we have in acting intentionally reveals its true nature, but there should be solid theoretical reason for invoking non-conscious inferential processes. In the case of intentional agency, I cannot see these reasons. On the contrary, this alleged inferential process would appear to be a paradigmatic

example of what we could carry out at a conscious level. The fact that agents systematically fail to gain conscious access to such an inferential process thus speaks in favour of there being nothing there to access. In other words, we should rather consider taking the phenomenology at face value: it tells us that there is no inference by which we come to know our actions.

§6 An ANSCOMBEAN WORRY

What one may find attractive with Anscombe's project in *Intention* is, among other things, her effort to surpass a Cartesian way of thinking about our knowledge of ourselves and the world. By the time Anscombe was writing *Intention*, she was resisting a temptation to defend what she formerly encapsulated in the slogan 'I do what happens' (1957, p. 52). Nevertheless, as McDowell points out, she did not give up 'the line of thought that the slogan was designed to capture' (2010, p. 423). That line of thought is inspired by Wittgenstein's critical attitude towards Cartesian epistemology (Teichman 2008, p. 17). For Anscombe's part, this critical attitude turns up as an effort to resist interiorizing what is known in intentional action 'so that it does not include what is actually happening' (McDowell 2010, p. 430). In this section, I suggest that the above problems with the Inferential Theory could function as a reminder for why non-observational knowledge of our own actions should be grounded in what we do – where what we do includes what happens – rather than being grounded in something that is merely in our own minds or safely within the borders of what Quine used to refer to as 'our sensory surfaces' (1960, p. 1). Moreover, I will suggest that the disadvantages of the Inferential Theory together with the unattractiveness of the two-factor thesis are indicators that Quine and other empiricists were quite wrong in thinking that 'we know external things only mediately through our senses' (Quine 1960, p. 1). On the contrary, we should consider seriously the idea that I take Anscombe to be pursuing: namely, that we may know facts about the empirical world solely in virtue of participating in our actions.

In *Intention*, Anscombe seems to treat our knowledge in action as epistemically on a par with the knowledge we have of our own minds. I can know that I am writing a thesis in philosophy with just as much confidence as I know that I'm thinking about my grandmother. You can muster all the evidence you want for thinking that I'm really doing something else – such as pointing out that much of my time is spent pacing through the hallways, reading Saul Bellow rather than philosophy papers, updating Facebook, and so on. If my writing of a thesis is an intentional action of mine, however, I should be no less confident that this is what I'm really doing than that my grandmother really is on my mind. This is not the same as saying

that as agents and minds we have infallible knowledge of ourselves, or that we cannot reliably find out what someone else is doing or thinking from the third-person perspective. The point is rather that there is a conceptual and necessary connection between 'a subject acting [intentionally] and her knowing what she is doing', as O'Brien says, in the sense that so acting rules out a certain form of epistemic disassociation and self-blindness (2007, p. 169).

By treating knowledge of one's actions as epistemically on a par with knowledge of our own minds, Anscombe moves upstream against the Cartesian tradition by rejecting the idea that the things that are in our minds are cognitively and epistemically closer to us than every item of knowledge that falls outside our minds. For Cartesians, the mind – or at least relevant parts of it – is a cognitive home for us since it provides a realm where, as Timothy Williamson says, everything that is in it 'lies open to our view' (2000, p. 93). If the Cartesians are right about this, the states of these relevant parts of our minds are luminous conditions, which, according to Williamson, means that the subject is in position to know that they obtain in any case where they do obtain in one's mind (2000, p. 95). One may, as Williamson does, doubt whether there are such thing as cognitive homes since thought – even introspection – must always engage with 'conditions whose *esse* is distinct from their *percipi* as soon as it engages with any conditions at all' (2000, p. 113). A true Cartesian will then try to show that at least certain parts of our minds prove special in this respect. If we take intentional actions to include what happens and are publicly available events, Cartesians would deny that the conditions determining the nature of our actions are luminous. Its location outside our minds will make sure that we are not always in a position to know what the nature of our actions is. Our intentional agency will not be a realm that could provide us with such a cognitive home.

The Inferential Theory does not commit to the idea that our minds are determined by luminous states; however, it fits nicely in with this familiar way of thinking since it takes knowledge of our actions to be derivative and inferred from knowing our intentions. That the world is arranged the way I intend it to be – viz. that I really am doing what I intend to do – is, on the Inferential Theory, something I know as a secondary reflection from knowing mental items, such as my decisions, intentions, plus beliefs about relevant background conditions. Anscombe, on the other hand, rejects this familiar image when she suggests that the 'only description that I clearly know of what I am doing may be of something that is at distance from me' (1957, p. 53). Her radical idea is to treat this distanced object of knowledge – viz. the knowledge of what one is doing as it includes knowledge of what happens – to be non-inferentially available without being known 'mediately through our senses' (Quine 1960, p.

1). In this way Anscombe seems to approach a position from which it could be just as natural to take the realm of intentional agency as providing a sort of cognitive home for us. Her move could thus be taken to approximate how Cartesians think we must be in position to know aspects of our own minds.[56] For Anscombe, intentional agency is a realm whose *esse* is not distinct from their *percipi*, as it were. It does not mean that we know all features of the things we do. What it means is that insofar as some aspect belongs under its intentional description the agent should know that it pertains to her action. Against that background, however, it makes just as much sense to start directly to talk about the knowledge we have of our own actions rather than derivatively from the knowledge we have of our own minds.[57]

At several junctures of the *Intention* it becomes clear that Anscombe perceives a serious threat in the familiar image. Just consider what she says in the following passage:

> *A man can form an intention which he then does nothing to carry out, either because he is prevented or because he changes his mind; but the intention itself can be complete, although it remains a purely interior thing. All this conspires to make us think that if we want to know a man's intentions it is into the content of his mind, and only into these, that we must enquire, and hence, that if we wish to understand what intention is, we must be investigating something whose existence is purely in the sphere of the mind; and that although intention issues in actions, and the way this happens also presents interesting questions, still what physically takes place, i.e. what a man actually does, is the very last thing we need to consider in our enquiry. Whereas I wish to say it is the first.*[58]

What Anscombe is urging us to resist here is a familiar way of philosophizing that starts from what one takes to be cognitively more at home to us and then proceeds to extend our philosophical understanding to realms that are perceived to be more obscure. Giving epistemic primacy to knowledge of our own minds as opposed to knowledge of our actions thus seems to carry over into the way we think about these things philosophically. A possible result of the Cartesian epistemology is meta-philosophical: it might provide a direction to our philosophical inquiries.[59] In particular, it may ask us to consider intentions and what goes on in our minds as being philosophically primary with respect to what goes on in the external

[56] Wittgenstein goes to the other extreme and suggests that it is not only our minds that have luminous conditions; rather, he seems to be entertaining the idea that, at least for the realms that are of interest for philosophers, 'everything lies open to view' (*PI* §126). For a similar interpretation, see Williamson (2000, p. 93n1).
[57] To some philosophers this aspect of Anscombe's view makes it tempting to include actions and the realm of intentional agency as a part for our minds. O'Brien, for instance, has argued that talk about actions belong to our basic psychological repertoire (2007, p. 130). The fact that actions extend beyond the limits of our body may thus be taken as a reason for thinking that our minds extend beyond those limits as well; at least, one might argue that the conditions of our intentional agency that make up our cognitive home should be recognized as parts of our mind due to their alleged luminosity. The limits of the mind is not, however, my topic so I shall leave it at this.
[58] Anscombe (1957, p. 9).
[59] See Lex Newman (2010).

world, which is where our actions belong. Anscombe's worry is thus that the familiar epistemology will incline towards an account of actions in terms of what is considered philosophically more primary: viz. the mental states and acts that take place in our own minds. Thereby it will lend support to an analysis or explanation of intentional actions as a metaphysical complex or hybrid constituted by mental states, -acts and -events in combination with events and circumstances that are external to our minds. This epistemology thus fits with what O'Brien describes 'a view of actions themselves as complex combinations of bodily movements with other more basic psychological phenomena' (2007, p. 129). In this respect, what McDowell has labelled a latent interiorization in epistemology signals not only an epistemic 'withdrawal from the external world' (1995a, pp. 395–7; 2010, p. 430); it may move us towards a philosophical understanding of the nature of intentional actions as a metaphysical composite or hybrid constituted by internal and external elements.[60]

There are two things one might be worried about concerning this familiar epistemological image, at least from the Anscombean perspective. On the one hand, one might fear that such an interiorization of knowledge of our own actions – viz. by taking our knowledge to pertain at the first stage to our intentions, decisions and so on, and only secondarily (viz. by inference or observation) to our actions themselves – may have consequences for how we go about in our inquiries into the nature of action. In particular, it may, as I just mentioned, lend weight to the idea that actions are not basic elements in our philosophical inquiries but metaphysical composites that cry out for explanation and analysis in terms of its more basic building blocks.[61] On the other hand, we may fear that the familiar image risks distorting the kind of standing in the space of reasons that knowledge of our own actions – as it is enjoyed by agents who act intentionally – arguably is (McDowell 1995a, pp. 395–7). In particular, it will suggest that we take knowledge of our own actions to be a kind of knowledge we have at one remove from the knowledge that comes first to us. An attractive feature of Anscombe's project – one that she herself occasionally struggles to be reconciled with[62] – is that she pushes us beyond this to allow, as O'Brien says, that 'our most basic awareness of ourselves is as performers of actions, mental and physical' (2007, p. 3).

[60] The notion of a 'metaphysical hybrid' is used by Williamson to characterize a view about the nature of knowledge that is entertained by similar reductionist programmes in the metaphysics of knowledge where knowledge is accounted for as 'a mixture of mental states together with mind-independent conditions on the external world' and where the reductionist programme tries to break 'the mixture down into its elements' (Williamson 1995, p. 533). Also see, Williamson (2000, chs. 1, 2 and 3) and O'Brien (2007, ch. 8).

[61] One may, of course, have independent reasons for thinking that the correct metaphysics of action is to treat actions as metaphysical complex constructs constituted out of more basic building blocks. In that case, this worry will not announce itself. For them, however, the challenge remains to account for the knowledge we have of our own actions in the wake of the problems I have raised for the Inferential Theory and the cognitivist approaches, and the problems that pertain to the different two-factor theses.

[62] Anscombe may be read as taking someone writing "I am a fool" on the blackboard to have practical knowledge even if she fails to be writing it. To the extent that she does, however, her notion of practical knowledge is vulnerable to the accusation of being a 'funny sort of

The problems pertaining to the Inferential Theory could serve as a reminder of the inadequacy of this familiar epistemological image and its latent interiorization. As we have seen, Paul's model struggles to give an account of the knowledge we have of our intentions in action insofar as we are acting intentionally. Moreover, an inference from knowledge of our intentions plus suitable background conditions will in many cases not be enough to ground the knowledge we have of our own actions. Both these points may be taken to suggest that there is something crucial that is contributed by the action itself when we know our actions. Perhaps what we need, then, in order to understand the epistemology and knowledge we have of our own actions is to recognize what it would mean to say 'we know what we are doing *because* we are doing it' (O'Brien 2007, p. 182). In other words, our discussion may be taken to suggest that we consider whether knowledge of our own actions is primary in the sense that acting provides an epistemic contribution of its own. Given that the epistemic contribution of one's doing something cannot be conveyed via observation or as mediated by our senses – after all, that will only throw us back to problems pertaining to the two-factor thesis – we should seriously consider the idea that agents may know facts about the empirical world solely in virtue of participating in their own actions. At least, I will venture to explore this image further in the ensuing chapter. Perhaps there really is, as O'Shaughnessy puts it, an 'awareness of creating' and that the will really 'provides us with an answer as to how we effect what we do effect' (1963, p. 387). That is, I think it is time we consider whether the act of willing and participation, all by itself, can ground knowledge of our own actions.

§7 CONCLUSION

I therefore conclude that we have reason to question the adequacy of the Inferential Theory as an account of agent's knowledge. On the one hand, it fails to provide an account of the evidential basis for such knowledge given that it fails to explain why we know our intentions while at the same time its alleged inferential basis is too thin to justify beliefs about our actions. On the other hand, we found that the non-conscious inference by virtue of which an agent allegedly comes to know what she is doing cannot respect the phenomenology. The Inferential Theory thus fails as an account of the kernel of truth behind Anscombeanism. I therefore think it is time to reconsider whether the action itself can provide a non-observational epistemic contribution. It is also to this idea we shall turn next.

knowledge' (1957, p. 82). I agree with McDowell and think that if the agent fails to be writing 'I am a fool' on the blackboard, she also fails to express *knowledge* by saying 'I am writing "I am a fool" on the blackboard'. The point is that practical knowledge must be *factive*. Whether one has it depends on whether one *is* actually doing what one takes oneself to be doing. See McDowell (2010, pp. 429–30). McDowell also mentions that Eylem Özaltum suggests a reading of Anscombe as giving voice to an opposing interlocutor in these passages.

§8 APPENDIX: GRICE'S INFERENTIALIST ACCOUNT

Paul's Inferential Theory is a development of an account once suggested by Grice. Let me briefly say what this account is and mention why I think it offers little help to the problems we are facing.

According to Grice, knowledge of what one intends to do – and, by extension, knowledge of what one is doing – is inferred from an agent's prior knowledge of what Grice somewhat technically defined as an agent's willing$_T$ do something. A willing$_T$ is an agent's conclusion that Φ-ing is the best, most favourable or valuable thing to do when the following conditions hold: (1) the agent can Φ in those circumstances; and (2) the agent will be motivated to Φ rather than something incompatible with it if she concludes in favour of Φ-ing. Grice then defined an intention in action as an agent willing$_T$ to Φ at the same time as she believes that she will Φ – where the latter 'will' is not the technical notion but the progressive content of one's belief about what one is about to do – when this belief is a result of so willing$_T$. Non-evidential knowledge of what one will$_T$ do was thus thought by Grice to give rise to the evidence-based belief that one will Φ against a background of experience. The latter would then be a piece of inferentially-based knowledge when the circumstances of reaching this belief are epistemically favourable.[63] By extension Grice's model could thus recommend itself as a way to explain how agents in favourable cases know what they are doing. An attractive feature of Grice's model, in this respect, is namely that it harbours the promise of providing an evidence-based notion of knowledge in action. After all, one's knowledge of what one is doing is derived from one's knowledge of what one will$_T$. Moreover, his model seems to be able to retain a notion of decision that is distinctively practical given that an agent settles what she will do in virtue of willing$_T$ to do something. The required belief comes along inferentially as the result of having decided what one will$_T$ do. Grice may therefore retain a notion of willing$_T$ that is distinctively practical which allows him an object of knowledge whose existence and nature is independent of the act of believing.

The key difference between Grice's inferential account and Paul's Inferential Theory, however, is that on Grice's notion an intention in action will constitutively involve a belief (viz. the agent's belief that she will be Φ-ing). In this respect, he is not entirely in line with what Paul calls our 'naive way of thinking' on which we think that 'intending is not a theoretical [or doxastic] matter at all but a [strictly] practical one' (2009b, p. 546). His model is therefore best categorized as a hybrid view that combines an inferential model of knowledge in action with the distinctively Anscombean idea that epistemic access – which Grice takes to be best characterized as a sort of belief rather than knowledge – pertains as a matter of necessity to intentional agency.[64] Paul suggests that we label this view as *Weak Cognitivism*:[65] it is cognitivist because it takes an intention to be, at least in part, constituted by a belief; however, it is only weakly cognitivist since it treats the formation of the belief as secondary to performing an act of willing$_T$. Despite diverging sharply from how Grice defined intentions, then, Paul's *DPA*-intention is meant to cover the same epistemic role as Grice's notion of a willing$_T$: namely to provide a non-observational evidential base for indirect and inferred knowledge about what one is doing and why. Paul's proposal thus differs from Grice's model in two major ways: first, by offering a different set of

[63] See Grice (1971, pp. 278–9). Also, see Paul's characterization of Grice's position (2009a, p. 11).
[64] In chapter 2, I argued that Anscombeans, broadly conceived, require an agent to have epistemic access to the nature of her own action insofar as she acts intentionally. If that epistemic access is best rendered as a sort of belief – as, for instance, Setiya thinks – Grice's model will count as an Anscombean Theory of action since it would be a necessary feature of having an intention, as it is defined by Grice, that one has a corresponding belief about what one is doing or will be doing. Anscombe, of course, thought that spontaneous knowledge is required for acting intentionally and I will argue in chapter 6 that we have reason to go back to Anscombe's original view.
[65] Views that take an intention to just be a sort of belief are labelled *Strong Cognitivism*. See chapter 3, §2.

propositions that an agent must know in addition to knowing her abilities; second, by giving up on the cognitivist idea that intention constitutively involves belief-like attitudes.

Grice's reason for believing in a tight connection between intending and believing – and thus, by extension, a connection with knowledge – is the observation that it is odd to say that 'I intend to Φ, but I do not believe that I will Φ' and he thought it could be explained if we can somehow see that intending implies believing. In other words, Grice thought that linguistic data akin to Moore's Paradox supports the truth of what is known as the *Strong Belief Hypothesis* (or SBT), and that it was up to a theory of intention to explain its truth:[66]

> *SBT*: *A* intends to Φ only if *A* believes that she Φ-es (or will Φ).[67]

As Davidson points out, appeal to such linguistic data will not give you that intending implies believing since these data can just as well come from the fact that asserting implies believing (1978, p. 91).[68] However, one may still think that *SBT* is true and thus that it stands in need of explanation by action theory and that Grice's model is a way to carry out this project.[69] Grice's theory, if it works, may for instance be used to explain and derive the truth of *Belief*, which Setiya suggested as the way to render Anscombe's basic insight (see chapter 3, §2 and §4).

One problem with Grice's model is, according to Paul, that it invokes the technical and artificial notion 'willing$_T$'. On Grice's view, willing$_T$ to Φ constitutes an intention to Φ when the willing$_T$ is suitably connected with a belief. If we could avoid such a technical notion, we would be better off; or so Paul argues (2009a, p. 11). Worse, however, is it that Grice never gives us any reason for thinking that agents must have knowledge about what one will$_T$: he simply seems to assume that this is so. Given the artificiality of this notion, however, it is hard to see that he could come up with an explanation for this connection. The joint strength of these problems, then, lends weight to the impression that Grice's notion is little more than stipulation. Another worry is that as a cognitivist view it implies that intentions and intentional actions must involve beliefs. However, there is reason to think that intentions may execute and guide our behaviour in the appropriate way even in the absence of beliefs. As Paul points out, we attribute intentions and give reason-giving explanations because we want to capture 'the fact that an episode of behavior had a goal-directed, means-end structure and exhibited responsiveness to reasons' (2009a, p. 5). It is not clear that this requires beliefs. When I, for instance, go to work in the morning I do all sorts of things habitually and zoned-out, as it were, giving little thought to what I'm doing and what my intentions are. Moreover, small children, sophisticated animals, and autistic subjects may all provide instances of agency whose subject lacks the ability to form the relevant beliefs. We should, as Susan Hurley points out (2001), make sure that we are not overintellectualizing our account of intentional agency so

[66] Moore's Paradox is the name for a peculiarity that pertains to assertions of the form '*P* and I do not believe *P*'. As G.E. Moore noted, this phenomenon is strange given that such a sentence may perfectly well be true; however, it is nevertheless always inappropriate to assert such a thing (1966, p. 78). Moore also recognized that one will get a similar result by replacing 'believe' with 'know' (1962, p. 277). Some philosophers have taken this phenomenon in support of the idea that the speech act of assertion implies a norm of belief or knowledge, since violation of such norms would explain why these sentences could never be asserted without being somewhat to blame. After all, one would then convey that one takes oneself to believe/know what one denies believing/knowing. See Williamson (2000, ch. 11).

What Grice is capitalizing on is that 'I intend to Φ, but I do not believe I will Φ' gives rise to a similar oddity. The difference between Grice's view and knowledge/belief accounts of assertion, however, is that Grice has defined an intention as constitutively involving a belief. The sentence 'I intend to Φ, but I do not believe I will Φ' will thus be contradictory given how Grice defines intentions.

[67] See John Brunero (2009, p. 312). The *Strong Belief Thesis* should not be confused with the related 'Weak Belief Thesis' that Wallace defends (2001; 2006), according to which intending to Φ involves believing that Φ-ing is *possible*.

[68] This is also a point that Paul makes (2009a, p. 11). The irony of this is that it is Grice who taught us that the connection between the meaning of an expression – and thus its semantic relations – and its linguistic use is indirect (John MacFarlane 2007, p. 204).

[69] For defenders of *SBT*, see Hampshire and Hart (1958), Grice (1971), Velleman (1985; 2007) and Harman (1976; 1986). Setiya signs up to something similar as the most plausible way to formulate the truth behind Anscombe's observations in *Intention* (2007a, pp. 663–64; 2007b, pp. 26, 41 and 48–56; 2008b, p. 390). For counterarguments, see Davidson (1978), Bratman (1999b) and, at least on some interpretations, Anscombe (1957). In chapter 3 I discussed how some of these counterarguments could be dealt with by qualifying the principles.

that we are not ruling out a range of intelligent responses and behaviours that should be classified as being performed for reasons. Vindicating *SBT*/*Belief* might not be such a good thing, after all.

I also think a third worry could be added given that Grice's notion of a 'willing$_T$' is an evaluative notion. To will$_T$ is to reach a conclusion about what one thinks would be the best thing to do, all things considered.[70] The willing$_T$ is therefore an all-things-considered judgement about what it is good to do. A problem for this model is that there are agents who are suffering from weakness of will or akrasia who will Φ intentionally by forming and executing an intention in action that goes against their all-things-considered judgements. In that case, one might say that the agent wills$_T$ some ψ-ing but ends up Φ-ing instead as a result of willing (in the non-evaluative sense) to Φ. Again, we should remember Frankfurt's observation that an agent may call into the 'vasty deeps' of herself for her will to be as she has decided to just to find that she is unable to make her will follow suit 'when the chips are down' (1992, p. 10). As Davidson has convincingly argued, one can explain such an akratic's behaviour as intentional only by invoking a non-evaluative notion of an intention that is independent of the evaluative all-things considered judgement. For his own part, Davidson invoked the notion of an all-out judgement, whose function it is to decide finally and absolutely that some ψ-ing or another is the best action. Davidson then argued that an agent's intention in action is a *sui generis* pro-attitude that may or may not obey one's evaluations. What goes on in akratic agents is that the all-out judgement and the all-things-considered judgement disagree, but that the all-out judgement wins out in the end (Davidson 1970a).

Whether or not we agree with Davidson's account of weakness of will, it should be clear that Grice's model falls prey to the phenomenon that Davidson deals with. For Grice the belief that is a constitutive element of an intention is inferred from the willing$_T$ and he portrays the latter as an evaluative all-things-considered judgement where one's motivation is somehow baked into this evaluative notion. But in cases of akrasia the connection between what one does intentionally and what one will$_T$ do is disrupted. What one does in such cases is not 'compatible' with what one judges as the best thing to do. As a result, no intention, at least as it is defined by Grice, can come about in such cases and consequently these cases cannot be instances of intentional agency which, of course, goes against the widespread assumption that akratic actions are intentional. If we bypass this problem, however, Grice would also face problems in explaining their knowledge in action. Akratic agents, although irrational, are not so thoroughly benighted that they also fail to know what they are doing. I may for instance fail to exercise and instead eat cake even though I judge exercising to be the better action. It does not follow from this that I will have to eat cake without knowing what I'm doing. A further problem for Grice is therefore that knowledge of what I'm doing as an akratic agent cannot be derived or inferred from knowledge of what I will$_T$ do. After all, what I will$_T$ is to exercise whereas I ended up intentionally eating cake instead. Grice's model can therefore not explain the knowledge an agent has of her own action insofar as it is akratic. Followers of Grice would either have to deny that weakness of will is a case of intentional agency, or thoroughly revise the model; however, this is nothing less than the familiar choice between Scylla and Charybdis. I thus conclude that there is little reason to expect Grice's theory to help solve the problems we are struggling with.

[70] As Paul says, Grice's notion of a 'willing is something like a conclusion about what would be the best thing to do' (2009a, p. 11).

6

The Metaphysics of Practical Knowledge

> *Can it be that there is something that modern philosophy has blankly misunderstood: namely what ancient and medieval philosophers meant by practical knowledge. Certainly in modern philosophy we have an incorrigibly contemplative conception of knowledge.* G.E.M. Anscombe[1]

> Visitor: *Well, divide all cases of knowledge in this way, calling the one sort practical knowledge, the other purely theoretical.*
> Young Socrates: *I grant you these as two classes of that single thing, knowledge, taken as a whole.*
>
> Plato, *Statesman*, 258e

§1 INTRODUCTION

What we have learned so far is that central to an understanding of intentional action is that its agent takes or treats something as her reason for acting (chapter 1). We have also learned that in order to understand intentional agency we need to explain the necessity or pervasiveness of an agent's non-observational knowledge in action (chapter 2). As I have also argued, the necessity of an agent's non-observational knowledge in action would be a reason to think that taking as one's reason is – or, at least, constitutively involves – a distinctively epistemic relation. In the last few chapters we have seen two main efforts at explaining the pervasiveness of knowledge which, in their own way, qualified these claims. Setiya gave up the necessity-claim with respect to knowledge, but nevertheless went on to defend a similar claim for the necessity of having a belief or belief-like cognitive attitude about what one is doing, and then used that claim as his springboard towards a cognitivist account of taking as one's reason that identified this attitude as a belief-like cognitive state (chapter 3). Paul, on the other hand, gave up any such necessity-claims in order to retain the idea that taking as one's reason is a practical matter – and so not involving an epistemic relation – thus

[1] Anscombe (1957, p. 57).

effectively signing up for a *distinctively practical attitude* (DPA) theory of intentions that conceives this process as not essentially involving belief-like cognitive attitudes (chapter 5).[2] She then sought to account for non-observational knowledge in core cases of agency as being inferred from knowledge of what one is *DPA*-intending plus suitable background conditions.

Setiya's account failed, as we saw, because it located belief-like cognitive attitudes about what one is doing as essential elements of the relation taking as one's reason. If that is an essential part of an intention in action, agents would be epistemically at fault for deciding to do things for which they lack epistemic justification or reasons for believing that they will do or are doing. Since agents are not blameworthy in this sense, Setiya's failure suggests that we should reject the claim that taking as one's reason is a belief-like attitude.[3] Paul, on the other hand, appears to have a better notion of what it is to treat something as one's reason; however, she failed to account for an agent's knowledge in action because the inferential model both failed to provide knowledge of one's actions in the relevant range of cases and its reliance on non-conscious inferences failed to find suitable motivation. What Paul's failure suggests is therefore that contrary to her assumption agent's knowledge really is spontaneous.

In what follows, I will suggest that we can, on the one hand, agree with Paul in conceiving the nature of intentional agency and the relation taking as one's reason as a distinctively practical relation, while nevertheless accepting that knowledge in action is a spontaneous, necessary feature of acting intentionally that pertains to its nature. In order to do that, I will defend the idea that taking as one's reason is an epistemic, knowledge-conducive relation even though we conceive of it as a distinctively practical attitude that does not essentially involve belief-like cognitive attitudes about what one is doing and why. This means I will dispute a widespread assumption: namely, that knowledge can be had only by virtue of the subject having *cognitive attitudes* or executing *cognitive mental acts*, or the kinds of mental states or acts the occurrence of which represents something *as being true*.[4] On this

[2] A DPA theory of intentions defines intentions in terms of 'the functional role they play as settled objectives that shape further practical thinking, initiate and sustain action, and at least usually engage certain norms of practical rationality', as Paul says (2009a, p. 12). Different *DPA* views have been suggested: Davidson via his emphasis on an all-out judgement in action (1978); and Bratman via his notion of intentions as elements in plans of action that aid intra- and inter-personal coordination over time while being subject to norms that structure practical reasoning (1999b). In the same breath Paul mentions Alfred Mele (1992). See Paul (2009a, p. 11n45 and n46; 2009b, p. 547n3).

[3] When an agent takes something as her reason she decides what she should treat as a reason-giving consideration on this occasion. She may then decide to treat some of her beliefs, desires, wants, or values as reasons for acting. One may argue that cognitive attitudes are necessary conditions for this process. Setiya's account involves more than this claim, however: he suggests that cognitive attitudes belong as essential elements in the nature of the very attitude by which one takes something as one's reason. What I have argued (chs. 3 and 4) is that intentional agency has no room for belief-like cognitive attitudes in that relation. That would be entirely compatible with the truth of the claim that one can only take some *R* as one's reason if one has a belief or relevant cognitive towards *R*. Not that I think that claim *is* true.

[4] In what follows, I will follow Velleman (1992a, p. 105) and Shah and Velleman (2005, p. 503) in taking cognitive attitudes to be mental states whose content is 'grasped as patterned after the world' (Velleman 1992a, p. 105). A belief is, of course, a paradigmatic example, but so are also conjectures, assumptions, and impressions and so on. Similarly, cognitive mental acts are acts whereby its agent is putting forward a proposition or content in one's mind as being true. It is an act or a process rather than a state because it is an event and something we do for a reason. It is cognitive because this event or action is to present something as being true. Judgements are paradigm instances of that and they often lead to the formation of corresponding cognitive attitudes. In contrast to this distinctively cognitive frame of mind are what I shall refer

orthodox view a subject must take the world to be a certain way and it is their so taking it that will raise it to the level of knowledge if it results from an appropriately exercise of an ability to tell what the world is like when 'what one takes to be so is so' (McDowell 1995a, p. 400). What this view rules out is that knowledge can be had solely in virtue of a distinctively practical frame of mind whereby one wills some action, commits to plans, actions and reasons, forms and executes one's intentions, and so on. Below I will reject this assumption by employing a framework for thinking about the metaphysics of knowledge that is offered by Keith Hossack (2007). What I will argue is that his framework is compatible with the idea that we can have knowledge solely by virtue of a distinctively practical frame of mind. In turn, I will argue that this allows us to conceive taking as one's reason as both a distinctively practical attitude and an epistemic relation with the result that the cognitivist and the inferentialist approaches will become obsolete. According to Frank Ramsey's heuristic maxim, it is often the case that 'the truth lies not in one of the two disputed views but in some third possibility which has not yet been thought of, which we can only discover by rejecting something assumed as obvious by both disputants' (1925, pp. 11–2). Until now we have implicitly assumed that knowledge and epistemic relations must involve the subject taking something as being a certain way. That assumption has led our contestants to choose whether taking as one's reason is to take oneself to be acting in a certain way, or to decide to be so acting. If we reject this 'ancient juxtaposition', as Danto puts it (1973, p. 1), a third option becomes available: that taking as one's reason is *both* epistemic and distinctively practical.

I shall first (§2) expound Hossack's framework before I argue (§3) that it allows us an elegant model for accepting practical knowledge as knowing in virtue of distinctively practical attitudes or mental acts. Then (§4) I will argue for the necessity of practical knowledge, so conceived, in intentional agency by defining the will – or the faculty the exercise of which allows an agent to act intentionally – as an epistemic faculty, a power of the mind which all by itself may cause its agent to know. I shall then (§5) show how this framework can accommodate some of the distinctive features associated with agents' knowledge and practical knowledge, such as Anscombe's claims that an agent's knowledge in

---

to as *practical* (or *conative*) attitudes, whose contents are 'grasped as a pattern for the world to follow' (Velleman 1992a, p. 105). In short, someone having a cognitive attitude towards *P* regards *P as true* whereas having a practical/conative attitude to *P* is to regard it as to *be made true* (Velleman 1992a, p. 105). Paralleling cognitive mental acts we may conceive *practical* mental acts – or, alternatively, *conative* mental acts – as acts whereby an agent puts forward some content or proposition as a pattern to follow. As opposed to the cognitive way of representing some proposition – viz. where one treats the content as a fact and *as true* – the practical mental act is the act of representing something as *to be made true*. Making a decision about what to do is to perform such a mental act.

Note that the latter need not be equated with value judgements. Dennis Stampe's way of characterizing this distinction – where belief and similar states are conceived as representing something '*as obtaining'* in contrast to desires and conative attitudes that represent something as '*the obtaining of which would be good*' (Stampe 1987, p. 355) – therefore risks being inadequate unless, that is, representing something as being good is treated in a technical, deflationary sense as equivalent to representing something as to be made true.

action is: (1) spontaneous; (2) that it is 'the cause of what it understands'; and (3) that insofar as something goes wrong in one's exercise of the capacities that yields practical knowledge 'the mistake is in the performance, not in the judgement' (1957, p. 82, 87). Finally, I shall see how this account leaves action theory and the issue of Ethical Rationalism (§6).[5]

§2 HOSSACK'S FRAMEWORK

Our main question in what follows is whether we may know some facts – in particular, facts about what we are doing and our reason for doing it – in virtue of a distinctively non-cognitive and practical mode of thinking. That is, we shall ask whether we may have what I shall follow Anscombe in referring to as *practical knowledge*. Anscombe credited this phrase to Aquinas, for whom it described God's knowledge of His creation. As Velleman says, 'God knows what the world is like, but not by dint of having found out; He knows what the world is like because it is just as He meant it to be' (2007, p. xiv). Anscombe's nod to medieval theology thus suggests that she thought of intentional action as something that 'realizes the agent's knowledge of it, just as the creation realizes God's omniscience' (Velleman 2007, p. xv). Both Setiya and Velleman are eager, as it were, to 'explain' Anscombe's claims in order to tame this notion and even 'set aside the practical dimension of the knowledge that Anscombe describes' by virtue of emphasizing the idea that the knowledgeable intention-belief 'causes its own content to be true' by 'causing my action' (Setiya 2007b, pp. 25, 43).[6] Given my rejection of Cognitivism, the question I am raising now is not whether we have non-observational knowledge of our own actions because an intention-belief has verified itself by causing its content to become true. Rather, I am asking whether we can know some fact in virtue of having a distinctively practical attitude or mental act, which are acts and attitudes whose content does not represent how things *are* arranged or *as being* in such-and-such a way, but rather as how they are *to be* arranged and are to *be made*.[7] What I will argue is that Hossack's framework for thinking about the metaphysics of knowledge allows room for an affirmative answer to that question and thus paves the ground for a distinctively practical way of knowing. I will spend this section expounding this view and comparing it with alternatives.

In what follows, I shall take knowledge as primarily a way to characterize the mental condition of an individual mind when she stands in a suitable epistemic relation to something

[5] As I defined it in chapter 1 §3, Ethical Rationalism is a strategy in moral philosophy that seeks to vindicate the normative or rational authority of morality from action theory by virtue of first deriving independent standards of practical reason from the nature of what it is to act intentionally and then argue that the moral life recommends itself on those standards of practical reason.
[6] Also see Velleman (2007, pp. xviii–xix, ch. 1).
[7] For this contrast between cognitive and practical cognition, see Velleman (2000b, pp. 24–6).

that is the case. What we do when we ascribe knowledge to some subject, *S*, is thus primarily to ascribe *S* with a knowledge-relation to some fact, *f*.[8] Given that *S* is a mind – or thinking thing – it is natural to think that there are some mental events or acts of *S* in virtue of which *S* partakes in the knowledge-relation and thus in virtue of which she has knowledge of *f*. I may, for instance, have perceptual knowledge of the fact that my chair is red. It would be natural then to think that I have this knowledge in virtue of processes or events going on in my mind that in turn relate to processes in my sensual apparatus as it responds to external stimuli, or 'through the effects,' as Quine says, that objects 'help to induce at our sensory surfaces' (1960, p. 1). In this respect, it is natural to think that a knowledge-report, such as '*S* knows *P*', attributes mental states to *S* just as '*S* believes *P*' or '*S* desires *P*' would do (Williamson 2000, p. 6). According to Hossack, however, there is a fundamental disagreement in the metaphysics of knowledge between the vast majority of philosophers who take knowledge to be *composed* or *constituted* by the mental states or acts in virtue of which knowledge obtains and those very few who, like Hossack himself, take this relationship between knowledge and it provenance to be *causal*. As I will argue in chapter 7, we have independent reasons in the metaphysics of knowledge for going with Hossack and the minority on this issue; for now, however, I shall rest content with showing that Hossack's framework allows for a distinctively practical way of knowing since it will be possible to have knowledge, as Hossack conceives it, in virtue of having distinctively practical attitudes and undergoing practical mental acts.

There are three main models in the metaphysics of knowledge depending on what decision one makes on the following two issues: (1) whether one conceives the relationship between the state of knowledge and its provenance of knowledge – viz. the mental states or acts in virtue of which one knows something – as a *constitutive* or a *causal* relation; and (2) whether one takes knowledge to be a metaphysical composite (or hybrid) relation, or rather treats it as a basic or simple relation.[9] If you follow the vast majority of philosophers, you will vouch for a constitutive relation in (2), thereby endorsing the following:

[8] This way of talking about knowledge is opposed to the kind of communal relationship one may have in mind when one speaks of all the knowledge that is stored in the British Library, or of what some age or culture knows, or the state of knowledge within a particular field of scientific inquiry. When we talk about such knowledge-relations the focus is not on what an individual mind knows; rather, it is on the kind of impersonal knowing-relation denoted by phrases such as 'It is known that *p*'. Although there is a relationship between personal and impersonal knowledge, the latter may not even require an individual who knows that *p*. As Hyman says, '[t]he invention of writing means that each of us can cease to remember without all of us, collectively, ceasing to know' (1999, p. 434).

[9] I follow Williamson in treating the notion of a 'metaphysical hybrid' as characterizing views about the nature of knowledge that are entertained by reductionist programmes in the metaphysics of knowledge. On these views knowledge is accounted for as 'a mixture of mental states together with mind-independent conditions on the external world', where the aim of the reductionist programme is to break 'the mixture down into its elements' (Williamson 1995, p. 533). McDowell tries to resists a similar view when he opposes what he calls 'the hybrid view', or the view that knowledge is decomposable into some condition for which we are completely responsible, and some surrounding conditions for which we are not (1995a, p. 405). Also see Williamson (2000, chs. 1, 2 and 3) and O'Brien (2007, ch. 8).

> *Constitutive Thesis*: Knowledge is constituted by being in the right psychological or mental state in a favourable context.

Those very few who have sympathies for the causal view, however, will rather endorse this:

> *Causal Thesis*: Knowledge is caused by being in the right psychological or mental state(s) in a favourable context.[10]

As Hossack points out, the basic disagreement between these views is a disagreement about how to read the 'in-virtue-of' relation that holds between the state of knowledge and its provenance of knowledge, viz. the combination of mental states or acts and conditions of one's external circumstances in virtue of which one knows something. The *Constitutive Thesis* reads this 'in-virtue-of' relation as pointing out constitutive elements or components of the knowledge-relation. The *Causal Thesis*, on the other hand, reads the 'in-virtue-of' relation as signifying a causal relationship. Obviously, it follows from some claim such as 'knowing is constituted by psychological or mental state, *M*, plus …' that knowledge is a kind of *M*. Just as obvious is it that this will not follow from – although it would perhaps be compatible with – the claim that 'knowing is caused by psychological or mental state, *M*, plus …'.

Turning to our second choice in (2) it becomes clear that the *Constitutive Thesis* may come in two different versions. Some philosophers will here take it that knowledge is a metaphysical complex or hybrid constituted by more basic building blocks or element. One may for instance think that knowledge is the conjunction of one or more qualified mental states (such as a justified belief, reliably formed belief, and so on) and external factors (such as the truth of what one believes). Proponents of the *Standard Tripartite Analysis*, for instance, take knowledge to be constituted by justified true beliefs, whereas reliabilists conceive knowledge as reliable true beliefs.[11] I shall refer to such views collectively as *Non-Trivial* versions of the *Constitutive Thesis*. Non-trivialists take the mental states, acts, or processes that constitute an agent knowing something as not including the state of knowing. If I may use Dancy's notion, it is not 'a bare truth' that we know something (2000, p. 162); rather, we know something in virtue of the fact that other conditions hold whose correct

[10] Hossack (2007, p. 11).

[11] For a version of the *Standard* analysis, see Roderick Chisholm (1957, p. 16). Chisholm employs the term 'accepts that' rather than 'believes that'. He is careful to add, though, that he nevertheless opposes the predominate doctrine that knowledge is a species of belief (1957, pp. 17–8). Also see Danto (1968; 1973, p. 4). A more recent defence of the idea that knowledge can be so analyzed and is a species of belief, can be found in Haddock (2010, ch. 9). Reliabilists disagree with the requirement of justification, putting emphasis rather on the truth-conduciveness of the belief-forming process. They thus substitute justification with the criterion that the belief must be reliably true – viz. reliably formed, reliably tracking the truth of the belief, and so on. See Unger (1968), David Armstrong (1973), Goldman (1975), Dretske (1971; 1981) and Nozick (1981). For a more recent defence of Reliabilism, see Sherrilyn Roush (2005). On *Pragmatist* or *Instrumentalist* notions of knowledge the factivity-condition – i.e. the truth of the belief – is replaced by some other success-condition.

description does not involve mentioning knowledge or its cognate terms. The fact that we believe some *P*, *P* is true, and so on will then be the conditions that constitute one knowing something and they are thus treated as its realizers which come before knowledge in the order of philosophical explanation. If one takes knowledge to come about in virtue of such a suitably qualified belief – viz. in the constitutive sense of 'in virtue of' – one will be vouching in favour of what Hossack calls a 'doxastic priority' theory in the metaphysics of knowledge as opposed to an 'epistemic priority' theory where knowledge comes before belief and one may therefore explain belief as a knowledge-causing state (2007, p. 11).[12]

Being a non-trivialist is not the only way to hold the *Constitutive Thesis*, however. Other philosophers, such as Williamson, have refused to accept that knowledge comes second in the order of explanation. According to them, the answer on (2) is that knowledge is a metaphysically non-composite and simple relation that is basic and fundamental in the order of philosophical inquiry. However, Williamson still signs up to what I shall refer to as a *Trivial* version of the *Constitutive Thesis* by treating knowledge as a relation constituted by a factive mental state, *M*, where *M* is the mental state of knowing being in which is both necessary and sufficient for knowing something (2000, pp. 21, 37):

> *Knowledge as Mental*: There is a state, *M*, such that *M* is a mental state and being in *M* is both necessary and sufficient for knowing some particular fact or proposition and the state of knowledge is constituted by being in *M*.

Proponents of the *Constitutive Thesis* thus split into two opposing camps depending on how they stand on the issue raised by (2); however, they agree on the issue raised by (1) in taking the relationship between the state of knowing and its provenance to be constitutive.

There is no reason why proponents of the *Causal Thesis*, on the other hand, should divide into two camps over the issue raised by (2). The reason is that opting for the *Causal Thesis* makes available an important theoretical distinction between the state of knowing and its provenance. The provenance of some instance of knowledge is the specific *way of knowing* or mode of knowledge-acquisition that is exercised on an instance, thereby allowing the subject to know. In the paradigmatic case this involves an exercise of an epistemic faculty whose successful outcome will be that its subject knows something.[13] Knowledge is thus a sort of crowning achievement of a successful exercise of a way of knowing. Perceptual

[12] Also see Williamson (2000, *Introduction*).
[13] I will follow Hossack in treating a faculty as a power of the mind and an epistemic faculty as a subset of that which, among other things, has the power to cause its agent to know something (Hossack 2007, p. 13). For a suitable notion of knowledge as a state established by an exercise of a way of telling (or an ability to tell), see Millar (2007, p. 187)

knowledge will, for instance, materialize as the result of an appropriate exercise of one's sensual apparatus under suitable conditions. You may see that the ball is red all over and by so seeing judge that the ball is red or form the relevant belief, thereby acquiring knowledge about the ball's surface and the fact that it is red. Alternatively, I may come to know some fact, *f*, by virtue of inferring and coming to believe the corresponding proposition, *P*, from something else I know (e.g. from knowing *P*&*Q*). The *Causal Thesis* is a metaphysical claim about the relationship between such ways of knowing – viz. the sensual impressions, inferences, judgements, beliefs and other relevant mental states or acts – and one's state of knowing something. What the thesis claims is that the state of knowledge is caused by its way of knowing rather than constituted by it. On the *Causal Thesis* the belief, the sensual impressions, the inferences and so on are all located at the level of the provenance of knowledge whose success may be crowned by causing its subject to know some *f*. There will therefore be no theoretical advantage to suggesting a candidate composite, $C_1\&C_2\&,...,\&C_n$, as accounting for the nature of knowledge if one holds the *Causal Thesis*. After all, such a theory would only succumb to a particularly odd version of the *Constitutive Thesis* where the provenance of knowledge would not constitute knowledge; however, $C_1\&C_2\&,...,\&C_n$ would.

We should note that the *Causal Thesis* is orthogonal to the traditional sorts of causal views that have been offered in epistemology. A causal theory of perception (or perceptual knowledge), for instance, is offered as an analysis of perception in causal terms. Such a theory will presumably appeal to the causal factors and relationships holding between the perceived object and our sensual apparatus in order to account for the nature of perception and our perceptual knowledge. These causal accounts are therefore offered as a sort of non-trivial *Constitutive Thesis* and analysis either for some species of knowledge, such as perception or inference, or as a general theory of the state of knowledge itself.[14] Such accounts are therefore causal theories only in the sense that they take the non-trivial constitutive elements of knowledge to be analysable in causal terms. The *Causal Thesis*, however, is a view about the metaphysics of knowledge and how this state relates to its provenance. Proponents of the *Causal Thesis* may very well take these causal views – at least, when they aspire to say something about a particular species of knowledge rather than knowledge in general – as saying something right and important about the level of provenance or our ways of knowing. They will, however, insist that the elements mentioned in such a story only give us an idea of the nature of this way of knowing rather than the nature of knowledge itself. What this brings

[14] For an instance of the former, see Grice (1961). Goldman (1967) offers what he calls a 'causal theory of knowing' in the latter sense.

out is that the *Causal Thesis* makes an important distinction between the state of knowledge and the elements that go into the different ways of knowing. In turn, this allows us to reconsider whether it would theoretically better to locate some features that have been traditionally associated with the state of knowing as pertaining to its way of knowing rather than to the knowledge-relation itself. Since that will have a crucial bearing on our issue I shall briefly illustrate it with respect to content and direction of fit.

Hossack has argued that there is no need in the metaphysics of knowledge to think that the knowledge-relation itself involves representational content. Content and representation are things for which one has theoretical need only when accounting for the ways of knowing. The mental states and acts whereby we know something will only bring about knowledge when our epistemic faculties operate properly under suitable circumstances. We need a level of representation to account for the nature of these processes because they can go wrong by introducing some kind of mismatch between our representations and the world which precludes the individual from knowing what these states purport to represent. On the *Constitutive Thesis* knowledge would inherit this theoretical need for representations from its ways of knowing since on that thesis knowledge is constituted by whatever must go into the nature of these modes of knowledge-acquisition. The *Causal Thesis*, on the other hand, can relegate representations and content to these ways of knowing while pointing to the factivity of knowledge – viz. the fact that knowledge cannot be wrong in this sense since it only obtains if it relates the subject to a fact – as a reason for thinking that representations are not theoretically called for in the metaphysics of knowledge. The job of representing is, in a way, sufficiently carried out by the mental states and acts that constitute our ways of knowing. The knowledge-relation itself, on the other hand, can be thought of as connecting the individual subject directly to the facts themselves without representational intermediaries. There might therefore be no obvious theoretical need in the metaphysics of knowledge for treating it as a propositional attitude or as a state that connects minds with propositional contents.[15]

Of course, our knowledge-attributions seem to be sensitive to the contents whereby an agent knows some fact. To take an example: I may report that my colleague knows that the tallest philosopher at the institute is rather dull, thereby conveying that he knows this fact, *f* – namely that some property, *F*, applies to some person, *P*. However, my report will also convey that my colleague knows this fact under a description. For him it is the 'the tallest philosopher' that picks out *P* and he knows that whoever is singled out by this description

[15] This move will presumably require us to invoke a *sui generis* ontological category of facts that may stand as *relata* of the knowledge-relation. Hossack argues that metaphysics provides good reasons for invoking facts independently of this. I return to that in chapter 7.

instantiates the property of idiocy. Given the report's sensitivity to a description it does not necessarily convey that he knows that he himself, given that he is the tallest philosopher, is dull. To do that one would need to add that he also knows that he is the tallest philosopher. There will thus be need for content to make room for these distinctions in knowledge-reports. We should not infer from this, however, that the knowledge-relation conveyed by the report is a relation between minds and contents. The need for content is due to the function of the reports – viz. its need to convey the fact known and the provenance of knowledge – and does not necessarily pertain to the knowledge-relation itself. It follows, that the *that*-clause used to report someone's knowledge by saying '*S* knows that ...' does not report the object of knowledge in the way that a that-clause reports the object of propositional attitudes when we say things like '*S* believes that ...'. Rather, the clause in the knowledge-report gives two things: (1) the content of the mental act or state that caused the knowledge (viz. the content of the mental act in virtue of which *A* knows *f*); and (2) the fact that one thereby knows. Hossack suggests that we analyze the logical form of knowledge-reports as follows (2007, pp. 6–7):

> '*S* knows that *A*' $=_{df}$ $(\exists x)(\exists p)$ (*x* is a mental act & *f* is a fact & content(*x*) = *that-A* & *that-A* is a mode of presentation of *f* & *S* knows of *f* in virtue of *x*)

A similar consideration applies to what is usually referred to as a state's *direction of fit*. As it is discussed in the literature, direction of fit is a metaphorical way of describing conditions of adequacy – or, as it is often put, the locus of responsibility – pertaining to psychological states, and it specifies how the subject stands with respect to the content of the state. If the state is a belief whose content *that-A* is an indicative mode of representation for the fact, *f*, the subject having the belief regards *that-A* as an adequate representation mirroring how the world is. In this respect, a belief displays a cognitive or from world-to-mind direction of fit where the mind treats its contents as descriptions of how the world is arranged. The opposing conative or practical from mind-to-world direction of fit is illustrated by desires, conative attitudes and distinctively practical attitudes since its subject will regard their contents more as recipes, orders, or directions for how the world is to be arranged.

The direction-of-fit metaphor is invoked to make a systematic distinction between how a subject will respond in different circumstances in a complex interaction between the contents of her attitudes, the attitudes' direction of fit and regulative principles that pertain to the attitudes in question. In particular, the direction of fit of some mental state or event, *M*, may enable us to see how a subject will respond should the content fail to match the world. If

the attitude displays a from world-to-mind direction of fit, the subject will typically adjust her attitude towards this content so that her world-to-mind attitudes correspond better with how the world is in fact arranged. On the other hand, if the attitude is mind-to-world the world's failure to correspond with the content of the attitude may motivate the subject to put an extra effort into arranging the world in accordance with the attitude. One can then narrow down a description of the specific state in question by depicting how the subject is disposed to regulate her relation and behaviour to its content. Usually, such a story gets told in terms of complex dispositional properties of the subject. In the case of belief, such a story often tries to cash out a certain minimal sensitivity to truth that lays bare the subject's disposition to regulate her acceptance of the relevant proposition. Velleman, for instance, has emphasized that the acceptance of some proposition when it is the content of one's belief 'is regulated in ways designed to promote the acceptance of the truth' and thus on 'its being regulated in a way designed to make it track the truth' (1992a, p. 113).[16] In short, a belief must be regulated in tune with sound epistemic principles. If the opposite or practical direction of fit pertains to some state, *M*, a complex of disposition properties will similarly account for the nature of the state and how it influences the subject's regulation of its content.

Given the distinction between knowledge and its provenance, however, a direction of fit is theoretically obsolete with respect to the state of knowledge itself. A direction-of-fit attribution to the state of knowledge will not contribute anything to our understanding of this state. After all, knowledge is not a relation with conditions of adequacy that a knower must satisfy in order to know properly. If the subject fails to regulate her actions properly, there will be no knowledge at all. There is reason to regulate one's behaviour with respect to *M* in order to know or retain one's knowledge, but it makes no point to ascertain whether the state *qua* knowledge fits the world or whether the world fits it. Attributions of such directions are thus only theoretically useful for the states and acts in virtue of which one knows. A consequence of the *Causal Thesis* and the metaphysical distinction between knowledge and its provenance is therefore that there is no theoretical need in the metaphysics of knowledge to tie the state of knowing in with a particular kind of content or a particular way of representing something. If knowledge was constituted by the mental states, acts, or attitudes, M_1, M_2, ..., in virtue of which some subject knows something, knowledge would inherit features from its constituent components. Philosophers who take knowledge to be constituted by a sort of

[16] If beliefs can be defined, at least in part, in terms of their specific function in our cognitive economy, a state's failure to satisfy the complex dispositional story counts as evidence in favour of dismissing the state as a belief. If someone is in utter disregard for the facts one could choose to label the subject's state as a state of irrational obstinacy rather than a state of believing.

belief will treat knowledge as a state with a from world-to-mind direction of fit that pertains to propositions or some truth-evaluable string of conceptual content. But, as far as the *Causal Thesis* is concerned, knowledge is a relation whose causal antecedents involve these features; having said that, there is no theoretical compulsion to impose these features on to the knowledge-relation itself. Indeed, as I will argue later, there are reasons in the metaphysics of knowledge for resisting such moves since by not saddling the state of knowledge with a particular direction of fit, a certain kind of content, and so on one can retain a sense of unity to the knowledge-relation in the face of a vast plurality of ways of knowing. This framework can thus handle that one may know facts in the world by means of very different ways of representing those facts. An animal may, for instance, know features of its environment by non-conceptual representation whereas we may come to know those features by conceptually representing them. Moreover, the *Causal Thesis* allows us to think that knowledge may come about in virtue of a vast variety of mental acts and states. A virtue of the *Causal Thesis* is therefore that the unity of knowledge is compatible with 'the great variety of mental acts and mental states in virtue of which we have knowledge', as Hossack says (2007, p. 12).

Hossack takes these considerations to count in favour of the idea that rather than treating knowledge as a mental state – or as a composite of mental states and/or external factors – we may conceive knowledge as a 'direct, unmediated mental *relation* to the world' that connects minds directly with the facts one knows without requiring intermediaries (Hossack 2003, p. 193 [my emphasis]). According to Hossack, we should thus think of knowledge as a metaphysically basic and non-composite mental relation that holds between minds and facts. On his framework, knowledge is therefore mental in the sense that what it connects are *minds* with facts; but it is metaphysically speaking a different category from the mental states and mental acts that compose the mental life and activities of an agent, as it were, given that the state of knowledge does not involve contents or a direction of fit but directly connects minds with the world. In a word, its *relata* are minds and facts unlike the relata of ordinary mental states (or attitudes) and mental acts that involve minds and some representational content. Hossack is at liberty to do that given that he accepts the *Causal Thesis* with its claim that all it takes for the 'in-virtue-of' relation to obtain between some mental or psychological state, *M*, and the fact that *S* knows some fact, *f*, is that *M* is among the causal antecedents of the knowledge-relation. On this story *S* knows *f* when the fact that *S* has *M* under suitable circumstances cause *S* to know *f*. Given that there is only a causal relationship between such *M*-states and *S* knowing something rather than a constitutive

relationship, one is free to entertain the idea that the knowledge-relation is something metaphysically other than a mental state or a composite involving such states.

Hossack thereby underlines the independence of knowledge with respect to its provenance. In fact, having knowledge will be conceptually independent of having a psychology that consists of the kinds of mental states that we have and which relate us to the world via representational contents and ways of relating to these contents. Knowledge may thus obtain by myriad of ways of being a mind. We may imagine, Hossack says, an omniscient knower who has no use for beliefs, judgements, or perceptions since it already knows everything (2007, p. 267). One may even question whether such a being – though it is a mind by virtue of knowing things – will have a need for a psychology at all. So on the *Causal Thesis* knowledge need not itself be a mental state or composed of such states; rather, knowledge can be treated as the mental relation that connects minds with facts.[17]

I think there is something deeply right about Hossack's view and I shall discuss it in more detail below.[18] For our current purpose, however, I want to concentrate on a simpler claim: namely, that the metaphysical distance between the state of knowledge and ways of knowing makes conceptual room for an important plurality in ways of knowing. In particular, I think this framework invites the idea that one may know in virtue of mental acts and mental states (or attitudes) that are distinctively practical. In other words, it allows us to dispute the widespread assumption that knowledge can be had only by virtue of a subject having cognitive attitudes or executing cognitive mental acts, viz. the kinds of mental states or acts the occurrence of which represents something as being true.

§3 KNOWLEDGE through AGENT PARTICIPATION

On an instance of knowledge-acquisition it may be that a subject, *S*, knows some fact, *f*, in virtue of – and so, at least in part, due to – cognitive states or a corresponding exercise of *S*'s cognitive abilities. The truth of the *Causal Thesis*, however, allows us to accept such stories without forcing us to model our understanding of knowledge on these cognitive or receptive ways of knowing. In a word, the *Causal Thesis* leaves room for something rather

[17] In turn, one may go on to argue that knowledge is the very ground for mindedness and having a full-blown mental life. Hossack compares the relationship between knowledge and mindedness with how the metaphysically fundamental relation of betweenness (viz. the relation of being between) might be said to be 'the very essence of matter' (2007, p. xi). In its role as the essence of matter, betweenness is not a relation that consists of the bits of matter in virtue of which the relation obtains at an instance of something, *b*, being between *a* and *c*. Betweenness can thus be instantiated by a myriad of ways of being matter. Similarly, knowledge is instantiated by many ways of being a mind. If knowledge is fundamental in this sense, one can argue that it is the ground for mindedness and having a mental life in the first place.

[18] The key advantage of the *Causal Thesis* is that it allows us to retain the unity of the knowledge-relation despite a plurality of ways of knowing. In particular, the unity of the knowledge-relation does not require a mental state or mental act, *M* – nor, for that matter, a specific kind of content – that is uniform across instances of knowing. Knowledge as mental relation is a way to explain that unity. I argue in favour of this view in chapter 7; but, as we shall soon see (§4), this feature of Hossack's view allows us to characterize intentional actions.

important in the metaphysics of knowledge: that we may conceive the nature of knowledge as something different from the nature of a cognitive mental state or cognitive mental act (or a composite of such states and other factors). In this section, I argue that this feature of the *Causal Thesis* allows us to form an elegant model of the spontaneous knowledge an agent has of her actions in virtue of acting intentionally. I shall refer to it as *practical knowledge* and I argue that an agent may know what she is doing and why through *agent participation*.

Up to this chapter we have reviewed efforts to explain the fact that agents know their own actions in virtue of acting intentionally which have either invoked suitable mental states and attitudes as constitutive components of the complex attitude an agent forms whereby she takes some *R* as her reason for acting or forms an intention; or, else, they have sought to vindicate our knowledge in action as inferentially based on knowing crucial mental elements (e.g. the decision or intention) of the action. What I think has partly guided the proponents of these views is the widespread assumption that knowledge is – either wholly or in part – constituted by mental states or mental acts that are cognitive in the sense that they have a from world-to-mind direction of fit so that their content purports to represent the world as it *is arranged*.[19] For Setiya and Velleman this assumption resulted in a conception of an agent's choice and intention in action as a belief-like cognitive attitude which is cognitive and belief-like in that it aims to represent how the world is arranged although what Velleman calls its 'direction of guidance' (2000b, p. 25) – viz. whether the attitude cause or is caused by what it represents – is practical in the sense that this attitude causes its own content to be true. In this respect, choice would be, as Velleman says, 'a case of practical cognition' (2000b, p. 25).[20] Similarly, Paul tried to account for agent's knowledge as beliefs about what one is doing that were supposed to be inferentially based on knowing one's intentions. What none of them considered seriously was whether an agent may spontaneously and non-observationally know what her action is and why she is doing it in virtue of having *distinctively practical* attitudes or by performing *practical* mental acts. That is, they do not consider the option of knowing in virtue of attitudes and acts where not only the direction of guidance is practical but whose

[19] Following Velleman (1992a, p. 105) and Shah and Velleman (2005, p. 503) I take an attitude, mental state, or mental act to be cognitive insofar as its content represents how the world *is* arranged. This is in contrast to practical (or conative) attitudes, mental states, or mental acts which represent how the world is *to be* arranged. Also see above §1.

[20] This suggestion marks an important break with tradition since philosophers have usually portrayed knowledge exclusively as the outcome or product of processes that take place in what we may categorize as our *receptive faculties*. Roughly speaking, a faculty is a power of the mind and the power of these receptive faculties is that they are in one way or other sensitive and able to register the occurrence of facts. Usually, it is assumed that such monitoring concerns a range of *f*s whose reality is independent of the reality of our receptive processes and that it is in virtue of among other things *f*'s (causal) influence on these processes that we come to know *f*. This is also why we categorize them as receptive: they receive what they know. Consequently, the object of knowledge must be given to the subject by virtue of the subject being affected by *f*, such as via testimony, by one's sensual apparatus, or as an inference from something previously known, and so on. The possibility that Velleman and Setiya entertain is that knowledge may come about by virtue of the fact that *f* is caused by the attitude whereby one knows *f*. As will become clear, I agree with Velleman and Setiya about this; however, unlike Velleman and Setiya I think the attitude whereby one knows in this way is a distinctively practical attitude or mental act.

direction of fit is practical in purporting to represent how the world is *to be* arranged. That is the possibility I recognize as genuinely *practical knowledge*.

I think this option is often overlooked because of a tacit assumption in the metaphysics of knowledge that favours some version or another of the *Constitutive Thesis*. What all constitution views have in common is that they understand knowledge as fundamentally constituted by – either wholly or in part – the mental or psychological state, mental act, or what have you in virtue of which the agent knows something. If proponents of this thesis were to recognize two different ways of knowing, K_1 and K_2, they would have to consider whether K_1 and K_2 are, to use W. E. Johnson's vocabulary (1921), determinants to the same determinable knowledge (or *K*). A range of determinants comprises determinants of the same determinable not necessarily because of some common property shared by the determinants, but rather because of some principled difference-making relation holding among the determinants. At least, that is how Johnson characterises it:

> *In fact, the several colours are put into the same group and given the same name colour, not on the ground of any partial agreement, but on the ground of the special kind of difference which distinguishes one colour from another; whereas no such difference exists between a colour and a shape.*[21]

A proponent of the *Constitutive Thesis* would have to say something similar about K_1 and K_2 insofar as one takes them to be, as Plato put it, 'two classes of that single thing, knowledge, taken as a whole' (*St* 258e). Knowledge acquired through a perceptual channel – such as seeing or hearing – is, for instance, structurally different from knowledge acquired through non-perceptual channels – such as knowledge of one's own mind, proprioception, or other kinds of knowledge of luminal facts. These will presumably be regarded as two different species and/or determinants of knowledge – as for instance sensible and non-sensible knowledge (Haddock 2010, ch. 9) – since one may account for them as different routes towards acquiring the same knowledgeable state. What that knowledgeable state is depends on one's version of the *Constitutive Thesis*. If you vouch for Williamson's model, you would have to connect perception and our non-perceptual means of knowing with the production of the mental state of knowing. If, on the other hand, you signed up for some non-trivial analysis of the state of knowledge your aim would be to connect perception and our non-perceptual channels with the complex state one takes knowledge to be constituted by.

[21] Johnson (1921, p. 176).

I agree with Plato and take it that an account of the agent's knowledge in action should strive to preserve knowledge in action as a determinable of 'that single thing, knowledge'. What we should notice is that it is impossible for a proponent of the *Constitutive Thesis* to respect this unitary virtue while acknowledging genuine practical knowledge, or knowing in virtue of distinctively practical attitudes or mental acts. After all, for a proponent of the *Constitutive Thesis* this 'in-virtue-of' relation will amount to the following claim:

> *Constitutive Model*: We know what we are doing and why we are doing it in virtue of practical attitudes or mental acts because in a suitable context these attitudes/acts constitute knowledge.

For proponents of the *Constitutive Thesis* to recognize genuinely practical knowledge in this sense – that is, to make room for knowledge in virtue of distinctively practical attitudes or mental acts – would mean that they would have to find a way to align this specific constitution-relation with the other constitution relations that pertain to more ordinary ways of knowing. However, since such a genuinely practical way of knowing will involve mental states that are so radically different from the states and processes involved in other ways of knowing – such as perception, memory, proprioception, conscious self-awareness, and so on – it seems impossible to incorporate this alleged knowledge within the patterns of differentiations that determine the other ways of knowing as species of knowledge. The reason most proponents of a *Constitutive Thesis* do not seriously consider genuinely practical knowledge is therefore that knowing in virtue of distinctively practical attitudes and mental acts would appear on their framework to require a *sui generis* knowledge-relation: viz. the kind of knowledge that is constituted by practical mental states or acts.

This problem is obvious if one takes knowledge to be constituted by a sort of belief or belief-like cognitive attitude. After all, you would then need to locate a belief as a constitutive element of the practical activity in order for the latter to constitute knowledge all by itself. In a word, these non-trivial versions of the *Constitutive Thesis* will drive people to conceive the relevant practical attitudes or mental acts as, at least in part, being constituted by beliefs or belief-like cognitive attitudes. In this respect, it would take us towards a cognitivist approach where the practical attitudes or acts in question would at most be practical in the sense that what Velleman calls their 'direction of guidance' would go from mind to world (2000b, p. 25). In other words, these models can at most yield a version of the above *Constitutive Model*

where Velleman's notion of 'practical cognition' constitutes knowledge; and we have already seen that this is no adequate model of agent's knowledge.[22]

More surprising, perhaps, is the fact that this persists as a problem on Williamson's trivial version of the *Constitutive Thesis*. Williamson does not take knowledge to be a sort of belief or belief-like cognitive attitude. Rather, he treats knowledge as a metaphysical simple or primitive by accounting for it as being constituted by a non-composite factive mental state or attitude of knowing, being in which is both necessary and sufficient for having knowledge (Williamson 2000, 1.4). It is important to notice that even though this framework takes knowledge to be a non-composite simple it nevertheless offers what Williamson calls 'a modest positive account' of this state (2000, p. 33). The crucial claim of that account is that knowledge is defined as the most general *factive stative attitude* (or FSA), which means that knowledge is 'that which one has to a proposition if one has any factive stative attitude to it all' (2000, p. 34). Like seeing and remembering, knowing is *factive* in being defined as a propositional attitude one necessarily has to all and only truths. It is moreover a propositional attitude that is *stative* in that it constitutes a state of the subject rather than a process such as forgetting. Williamson's suggestion is thus to say that for any Ψ that is a factive stative attitude it follows from '*S* Ψ-es that *P*' that '*S* knows that *P*'.

For this framework to render the truth of the *Constitutive Model* the relevant practical attitudes or mental acts must involve some relevant FSA towards the proposition that one is so acting. The problem is that there are no obvious states that are involved in – or constitutive elements of – acting intentionally that appear as promising FSAs. What the agent has agent-knowledge of is 'what happens'. Thus, such an FSA would need to have the fact – or, alternatively, the true proposition – that one is so acting as its object. There clearly appears to be room for FSAs in explanations of intentional actions. As Gibbons argues, convincingly in my view, knowledge of means plays a causal role – at least as causal background condition – for explaining someone's actions (2001). It may even be, as Hornsby argues, that knowing one's reasons plays an indispensable role in intentional agency (2008); or more radically, as Hyman suggests, that unless you know *R* it cannot be your reason for Φ-ing (1999, §4). However, none of the knowledgeable states that play indispensable explanatory roles in intentional agency seem to be distinctively practical states or attitudes that have one's doing something as their object. In fact, it appears that the gap between the practical mental states that are candidate constitutive elements of acting intentionally – such as wanting something,

[22] See chapters 3 and 4.

valuing something, and, if the following are states, intending to Φ, having decided to Φ, and so on – leaves an indispensable gap between the occurrence of these states and the agent actually doing what the attitude represents as to be done. After all, this is why most philosophers try to devise accounts of what it is to act intentionally that combine a story of such practical states and acts with a certain way in which they cause or bring about movements of the body, behaviour and so on. The perceived gap between the occurrence of such states and the fact that one is acting intentionally is thus evidence that there are no relevant FSAs constitutive of acting intentionally. An agent surely has knowledge of what she is doing insofar as she is doing it intentionally; however, that knowledge cannot be constituted by an FSA that both has the fact that one is so Φ-ing as its object and is an essential practical element of acting intentionally. FSAs are no aid to the *Constitutive Model*.[23]

A genuinely practical way of knowing – knowledge coming about in virtue of practical attitudes, states, or acts – therefore seems to require its own determinable within the framework of a *Constitutive Thesis*. A problem with that is that it will threaten its claim to be recognizable as knowledge. Just as arguing in favour of the existence of a colour, *X*, that fails to fall under the determinable colour would likely lead to the conclusion that *X* is no colour at all but belongs to some other determinable, the idea that practical knowledge requires its own determinable will probably provoke the reaction that it fails to be knowledge.[24]

A key advantage of the *Causal Thesis*, then, is that it allows us to recognize genuinely practical knowledge while avoiding the pitfall of invoking a new determinable. The framework of the *Causal Thesis* opens up an important conceptual distance between the nature of the knowledge-relation and its provenance or ways of knowing. That distance not only removes the theoretical compulsion to tie the nature of knowledge in with the nature of

[23] It should be noted that there is no suggestion in Williamson's work that he accepts anything like the *Constitutive Model*, or that he thinks his framework can adequately account for something like knowledge by willing. What I am entertaining here is only the idea of taking Williamson's framework as a starting point for accounting for the *Constitutive Model*.

A possible variant is perhaps available to someone endorsing what Hyman calls 'the minority view', or the idea that we may understand knowledge in a sort of an anti-intellectualist way as 'a species of ability' rather than 'a species of belief' (1999, p. 435). According to Hyman, this view can 'trace its pedigree' to Plato (*Th* 196c–199c) and Aristotle (*OS* 417a21), but also to Wittgenstein's remarks that '[t]he grammar of the word "knows" is evidently closely related to that of "can", "is able to"' (*PI* §150); or, for that matter to, Ryle's claim that '"know" is of the same family as skill words' (1949, p. 129). Hyman picks up this view and suggests that we may define knowledge as 'the ability to do things, or refrain from doing things, or believe, or want, or doubt things, for reasons that are facts' (1999, p. 441). The knowledge one has in virtue of willing something (or agent-participation) can perhaps be accounted for in terms of the ability one has to do things for reasons that are facts, an ability resulting from the agent-participating activities one undergoes in acting intentionally. One could then argue that insofar as one is Φ-ing intentionally this fact – viz. that one is so Φ-ing – becomes available as a reason for doing/refraining from doing other things, believing, wanting, or doubting things, and so on. Since this ability constitutes knowledge we may have knowledge by willing without needing to invoke problematic mental states. The problem for this view, however, is, as Williamson points out, that it cannot refer to a single failure to know to explain the correlation between a range of things I cannot do, refrain from doing and so on for a reason that is a fact (2000, p. 64n1). If I fail to know some *R*, it is true that I cannot Φ_1 for the reason that *R*, Φ_2 for the reason that *R*, or refrain from Φ_1-ing for the reason that *R*, refrain from Φ_2-ing for the reason that *R*, and so on. These incapacities will have little in common except for the fact that I fail to know *R*. However, if these incapacities define what it is to know *R* – viz. since this long list of incapacities makes up the portfolio of the ability that defines this knowledge – we cannot explain the correlation between them as my failure to know *R*. I therefore think we can safely put aside the idea that knowledge as ability can provide a version of the *Constitutive Model*.

[24] A case in mind is the standard interpretation of Ryle as claiming that knowledge-how is a *sui generis* type of knowledge (1949, ch. 2).

distinctively cognitive attitudes and processes so that we can, as we saw in the previous section, avoid connecting the nature of knowledge with a specific direction of fit, a specific representational content, mental state or act, and so on. It also allows room for a distinctively practical way of knowing. On this framework the relationship between the knowledge-relation and a way of knowing is causal rather than constitutive. We may therefore respect the distinctively practical dimension of an agent's knowledge in action by locating this practical dimension in the causal antecedents of knowledge. We can then say that some *S* has practical knowledge insofar as its mode of knowledge-acquisition – viz. its way of knowing – consists of mental attitudes, states, or acts that are distinctively practical. Given the *Causal Thesis*, those features of the causal antecedents of the knowledge-relation will not bear on how we understand the nature of this relation itself. On this framework we are thus in position to recognize the practical dimension of the knowledge one has in action without needing to invoke a new *sui generis* determinable. It thus looks as if the *Causal Thesis* is a suitable framework for a genuine notion of practical knowledge since it can allow practical mental states and acts to go into that 'great variety of mental acts and mental states in virtue of which we have knowledge' without threatening the unity of knowledge or requiring a new determinable (Hossack 2007, p. 12). When an agent has practical knowledge of her own actions we can therefore say that she has ordinary knowledge of the relevant fact, *f*, but that it is practical in that it obtains in virtue of (viz. by being caused by) a practical way of knowing.

It is this possibility within the framework of the *Causal Thesis* that we can capitalize on in accounting for an agent's knowledge in action. As you may recall from chapter 2, we followed Anscombeans in characterizing an agent's knowledge in action as follows:

> *Agents' Knowledge* (AK)*:* If *A* is Φ-ing intentionally (or for reason *R*), then *A must* either be Φ-ing while having non-observational, non-inferential and spontaneous knowledge that she is so Φ-ing, or else she is so Φ-ing by virtue of doing other things – which is not merely trying to Φ – for which this knowledge-condition holds.[25]

Through our discussions since chapter 2 we have seen efforts to qualify *AK* by weakening the intimate connection between acting intentionally and acting knowingly, giving up the idea of spontaneous knowledge, or by letting go of the claim to know. Let me now suggest that we

[25] Agents' knowledge is non-observational in the sense that it does not need observational clues or evidence in order to obtain. It is, moreover, direct and non-inferential because knowledge of what you are doing and why is not derived or inferred from other things you know about the world or yourself. Finally, it is spontaneous because it obtains in virtue of the fact that one is an agent (in this respect it seems true, as Anscombe reports Aquinas as saying, that agents' knowledge is 'the cause of what it understands' (1957, p. 87)). I discussed the principle of *AK* as the most plausible rendering of *Anscombe's Dictum*. See chapter 2, §3.

move back to the original principle since I think the framework of the *Causal Thesis* provides a suitable framework for explaining the truth of *AK*.

According to O'Brien, 'actions are things we know, not by observing them, or by reflecting about them, or accepting some representation of them'; on the contrary, she says, we know our actions 'by actively engaging in them' (2007, p. 183). What O'Brien is driving at here is the claim that *qua* being intentional agents we can know by virtue of willing our actions or participating in the events that are our actions.[26] Suppose we characterize this knowledge as *knowing through willing* or *through (agent) participation*. Suppose further that we agree with Paul and other distinctively practical attitude theorists in thinking that an exercise of our will – viz. an exercise of the faculty that enables an agent to act intentionally or for reasons – is a distinctively practical matter and that the attitude or mental act whereby an agent takes something as reason is a practical attitude or act. What I suggest is that armed with the *Causal Thesis* we may account for our knowing through participation as practical knowledge. This would be knowledge of what we are doing and why, which is practical since its causal antecedents are the practical elements that constitute our active participation in the events that are our actions. I shall call this the *Causal Model*:

> *Causal Model*: We know what we are doing and why we are doing it in virtue of acting, willing, or participating in our action because in a suitable context this distinctively practical activity of the agent causes her to know these things.

The *Causal Model* is an instance of the *Causal Thesis*. It is cashing out that on this framework saying that one knows in virtue of participating in one's action means that one's participation is the causal antecedent for the fact that the agent knows her action. Given that participation is distinctively practical, knowledge through participation will be a practical way of knowing:

> *The Causal Model for Practical Knowledge* (CMPK): An agent having practical knowledge of her own intentional action – viz. that she is Φ-ing intentionally or for some reason *R* – is caused by the agent's distinctively practical participation in the action (viz. when the action obtains by her exercising of the will or by the agent willing the action).

[26] On O'Brien's framework this possibility turns on whether 'the action can function unmediated as the reason for its own ascription', something she goes on to argue that it does. On her view, '[a]ctions carried out in a certain way are conscious and able to stand without further representation as the reasons for their own ascription', a consciousness she takes to come about because the agent has what she calls a 'sense of control' in agency (2007, pp. 183–4). A problem with O'Brien's view, however, is that she accounts for this control in action as something we get by 'acting on the basis of an evaluation of the possibilities open to us' and 'when we act on the basis of our evaluation of possible actions' (2007, p. 184). If 'sense of control' is restricted in this way, however, O'Brien's model will not cover the knowledge an akratic agent has of her own actions when she acts intentionally and knowingly but fails to act on the basis of her evaluation. A point in favour of the view I am defending here is that it covers akratic instances of intentional agency. For some defence of the claim that akratic actions are intentional and voluntary, see Amelie Rorty (1980; 1981) and Davidson (1970a; 1982a).

What I suggest is therefore that this is the kind of knowledge that an agent has of her own actions insofar as she is performing them intentionally. Knowledge, as it is portrayed by the *Causal Model* and *CMPK*, seems to provide the right structure for agent's knowledge given that it will have the good fortune of being non-observational, non-inferential and spontaneous. After all, if one's knowledge in action can come about all and only as the result of participating in one's actions – viz. by being caused by the distinctively practical relation one has to the events that are one's intentional actions – it would explain why an agent acting intentionally is in no need of making observations or inferences in order to know what she is doing and why. Moreover, if the knowledge portrayed by the *Causal Model* is what occurs whenever agents act intentionally, we would have an elegant explanation for the truth of *AK* that does not oblige us to impose alien cognitive attitudes or mental acts in the process of acting intentionally. If our knowledge in action comes about through such participation – that is, even if we conceive this participation as essentially involving all and only practical attitudes, states, or mental acts – there will be no reason in the epistemology of intentional agency for invoking attitudes, states, or acts with a cognitive direction of fit in an account of intentional agency. The advantage of this framework is thus that it can explain *AK* without dictating that intentional agency must contain alien cognitive elements, FSAs and so on.

If agent's knowledge is modelled on the *Causal Model*, we might therefore vindicate the truth of *AK* from within a distinctively practical conception of intentional agency and taking as one's reason. This would allow us to meet the challenge Setiya raised against such views when he questioned their ability to explain the necessity of an agent's knowledge in action (2007b, p. 41).[27] Moreover, by vindicating *AK* from within a distinctively practical framework, we can avoid the epistemic problems pertaining to the different cognitivist views since there are no corresponding epistemic requirements pertaining to the practical mental states, attitudes, or acts. Finally, the view would render Paul's inferential model obsolete. I take all this to favour the *Causal Model* as an account of the knowledge we have in intentional agency by virtue of being an intentional agent who exercises her will. This model will probably require more development and work; however, given its advantage I think it is time to put it forward as a promising working hypothesis. Let me now turn to see, first (§4), whether we have reason to believe that this way of knowing will necessarily occur insofar as

[27] Setiya argues against that taking as one's reason cannot be accounted for by saying that an agent takes some reason, *R*, as a good reason to Φ since that will not help explaining why action done for reasons must satisfy *Belief*. The general point is that he thinks practical commitments cannot explain why agents' knowledge pertains to intentional agency as a matter of necessity. Also see Paul (2009b, p. 547).

we are acting intentionally, and second, how this framework can account for certain features associated with Anscombe's introduction of the term practical knowledge (§5).

§4 The WILL as an EPISTEMIC FACULTY

Bertrand Russell once said that he was 'inclined to think that volition differs from desire logically, in a way strictly analogous to that in which perception differs from belief' (1918, p. 228). It is a shame that he thought it would take him 'too far from logic to discuss this view'; however, I think we should be in position now to speculate whether there is some truth to the idea that volition stands to desire as perception and knowledge stand to belief. In the last section, I concluded with a conditional claim: namely, that we could explain the truth of *AK* if practical knowledge, as it is portrayed by the *Causal Model* and *CMPK*, occurs whenever an agent acts intentionally. In this section, I shall defend the antecedent of this claim by virtue of suggesting a framework where volition or agent participation – viz. the performance by virtue of which an agent acts intentionally – is an epistemic relation by being the distinctive mental act of what I shall recognize as an epistemic faculty: namely, the will. In particular, I claim that the will – viz. the faculty an exercise of which enables agents to act intentionally – is an epistemic faculty in the sense that its function can be defined in terms of knowledge. That will not only allow us to demarcate a class of interesting particulars – viz. the range of individuals or agents who have a capacity for intentional agency – but it will also explain the necessity of having practical knowledge insofar as one is acting intentionally.

In the previous chapters we have dealt with the will as a faculty or power of the mind the exercise of which enables us to engage in practical thought and deliberation, in forming intentions and, if all turns out well, in acting intentionally.[28] Having a will thus enables us to make what Aristotle characterized as a choice (*prohairesis*), which is an act or process he thought was restricted to the rational animals given that its exercise requires 'reason and thought' (*NE* 1112a15-6; 1111b6-7). In accordance with this line of thinking, many philosophers talk about the will as something whose possession and exercise make the difference between intentional agency and other incarnations of agency.[29] According to Ginet, people are 'the paradigms of enwilled agents' and the specimen of agency we are after – viz.

[28] The will is a capacity for self-determination the exercise of which allows agents to take something as their reason – e.g. a desire or some fact, *f* – and thereby act for those reasons. Also see Wallace (1999, p. 59; 2001; 2004) and Sebastian Rödl (2010, p. 139).

[29] It is somewhat common – at least within the confines of the Aristotelian vineyard – to make a distinction between actions that are voluntary (*hekousios*) and actions that are chosen (*prohairesis*). As Aristotle defines the *hekousion* it is actions whose 'moving principle is in the agent himself, he being aware of the particular circumstances of the action' (*NE* 1111a23-4). That category is wider than the category of *prohaireton* or chosen actions. All chosen actions are *hekousios*, he says, however 'the latter extends more widely' to include both children and non-rational animals and actions on the spur of the moment (*NE* 1111b4-10). Anscombe operates with a similar distinction between intentional and merely voluntary actions; however, she seems to include actions done on the spur of the moment as intentional (1957, p. 89). For further discussion of Aristotle's distinctions, see John L. Ackrill (1978) and Sarah Broadie (1991, ch. 3).

intentional actions or actions done for reasons – 'applies only to beings who have wills' (1990, p. ix). On the action-theoretic side it is thus the will that grounds our primary object of inquiry.[30] The will is thus commonly defined by philosophers as the faculty or 'the ability to choose and perform intentional actions' (Raz 1999, p. 48).

In general, I treat a faculty as a power of the mind of any sort. Without implying compartmentalization in any interesting sense, I have assumed that the will is a faculty whose function it is to engage in the kind of thinking we recognize as thinking with a view to action and decision-making. Such thought-processes will involve distinctively practical attitudes (DPAs), practical mental acts, the action itself when all goes well and an intention in action. Sometimes such thinking enables an agent to go about Φ-ing directly. At other times she forms a future-directed intention that bides for its time to come. Now, I will follow Hossack also in assuming that among our faculties there is a sub-set that comprises *epistemic faculties*, by which I shall mean a 'power of the mind to cause itself to know' (Hossack 2007, p. 13). A faculty, *F*, of an individual mind or agent, *A*, is therefore epistemic insofar as *F* has the power to cause its bearer, *A*, to know some fact, *f*. The mental states, attitudes, or acts that are exercised by *F* constitute *F*'s distinctive way of knowing. These ways of knowing – such as perceiving, reasoning, or remembering – are different ways of apprehending some fact. An act of apprehension is thus the distinctive contribution of an epistemic faculty. The operations of an epistemic faculty will thus cause the agent to know something by virtue of enacting an apprehension-act towards some fact under circumstances that are favourable to the faculty in question. You will, for instance, need light for operations of the visual apparatus to be of any use. With the possible exception of knowledge of one's own mind, the possession of an epistemic faculty is thus a necessary means for some individual to partake in knowledge-relations.[31] What I suggest is that the will is one of those epistemic faculties and that we may regard its proper exercise of as causing its agent to know what she is doing and why.

It is important to realize that our ways of knowing and the corresponding acts of apprehension in our epistemic faculties play a vital role in accounting for the fact that knowledge is not a state one could be in by just any configuration of causal relations.

[30] As I mentioned in chapter 1, there is an echo of this emphasis in ethics and moral theory where it is common to connect enwilled agency as the basis for moral accountability or responsibility. Susan Wolf, for instance, says that 'only an agent who has a will—that is, who has desires, goals, or purposes and the ability to control her behavior in accordance with them—can be responsible for anything at all' (Wolf 1990, p. 7). In particular, she goes on to emphasize the role of the will in being, at least potentially, '*effective* in determining the existence of these events or properties' if the agent can be said to be responsible for them. In other words, both intentional agency and ethical responsibility seem to require a form of control that agents only have if they possess a 'potentially effective will' (Wolf 1990, p. 10).

[31] I leave knowledge of one's own mind as a possible exception because it seems right to conceive consciousness as knowledge we have without the operation of a corresponding faculty. In other words, there is '*no* faculty of consciousness' (Hossack 2003, p. 195). Knowledge of one's own mind may nevertheless depend on one having such epistemic faculties, if not only for the reason that an individual 'which never knows anything is not a mind' and thus has no mental states or events to know about (Hossack 2007, p. xi).

Knowing the name of your spouse's great grandfather is not achievable by merely getting a bump on the head, for instance. Such a relation will only materialize by virtue of the work of a relevant epistemic faculty. A mathematical truth gets known through testimony, or by the exercise of one's own mathematical intuition and capabilities of deduction. In either case, there will an act of apprehension on part of the relevant epistemic faculty and on Hossack's framework this mental act will cause its agent to know this mathematical fact conditional on the act of apprehension being properly executed under suitable circumstances.

One may define the nature of such faculties in terms of their distinctive ways of causing their bearer to know. Perception can thus be defined and distinguished from memory in being 'the mind's power to cause itself to know facts about its surroundings' via the mental acts of looking, sensing, and hearing and so on. Similarly, memory can be accounted for as one's power to yield knowledge of facts about the past via the act of remembering. In general, an epistemic faculty thus causes its bearer to know by the mental act that is specific to the faculty in question. The model of an epistemic faculty is thus as follows:

> *Epistemic Faculty*: Some agent, *A*, knows some fact, *f*, by means of faculty, *F*, when *F*'s distinctive mental operations, *F*-ing – forming some mental state, performing a mental act, and so on – operate under favourable circumstances thereby causing *A* to know *f*.

Similarly, we may characterize the will as an epistemic faculty as follows:

> *The Will as Epistemic*: Some agent, *A*, knows some fact, *f*, by means of the will when the will's distinctive operations – viz. forming practical mental states, performing practical mental acts, and so on – operate under favourable circumstances thereby causing *A* to know *f*.

To understand the will as an epistemic faculty means to define it, at least in part, in terms of its ability to cause knowledge in favourable circumstances. Now I think there are good reasons for seriously considering the will as an epistemic faculty. The fact that the products of the will – viz. decisions, prospective intentions, intentions in action, intentional actions – all seem to be things we know insofar as we are doing something intentionally can, I think, be adduced in favour of this hypothesis. The will is a faculty whose proper operation grants its agent with 'a sense of control', as O'Brien says (2007, p. 184). The reason is that the will makes sure that, as Velleman puts it, the agent gets 'appropriately connected to the facts' that one thereby knows (2007, p. xiv). After all, it is presumably right to think that the exercise of the will's distinctive operations – viz. its act of will or agent participation –

requires knowledge of how to do the relevant things. That is to say, insofar as you are Φ-ing intentionally you will, at least at the time of Φ-ing, know how to Φ.[32] That knowledge-how will – whether one conceives its nature as propositional knowledge, or as a non-propositional, complex dispositional capacity – grant its bearer a reliable and predictable capacity for, at least under suitable circumstances, executing the performance one decides to enact. You may know how to play the *Wanderer* piece on the piano. If so, you will reliably be able to produce this performance. However, only under suitable circumstances: it requires the presence of a piano, that your hands are not tied behind your back, and that you are conscious, and so on. But when the circumstances are suitable for an exercise of a relevant action the idea is that an agent is in a position to know what she is doing and that all it takes to get into that position, as it were, is for her to do the thing by virtue of exercising her will. The epistemically suitable circumstances of the will, as defined by the above principle, are thus intimately bound up with suitable circumstances for the enwilled exercise of different actions.

According to Hossack, we may use the notion of the epistemic faculty – defined as a capacity to bring knowledge about – to make a distinction between minded beings and creatures merely having a capacity for undergoing functionally similar operations that fall short of generating knowledge. One may, for instance, use an epistemic criterion – viz. one's power to cause knowledge – to define mental states such as beliefs by requiring that 'a system of functional states can be a system of beliefs only if at least some of them cause knowledge' (Hossack 2007, p. 12). A functional characterization of these states – viz. the way beliefs guide our actions, are updated in the light of experience, and so on – will thus not be enough to capture the nature of what it is to be or have a belief-state. States with such a functional role can be understood as belief-states, as Hossack says, 'only if some of them at least cause knowledge' (2007, p. 266). In a word, Hossack employs the possession of actual knowledge as a way of differentiating a certain form of mindedness from merely having what can be recognized as functionally on a par with having a mind. Consequently, he treats a mind as 'that which knows' something and adds, contrapositively, that 'that which never knows anything is not a mind' (2007, p. xi, 194).[33] For Hossack knowledge – or the basic relation between minds and facts – is thus constitutive of and the essence of mindedness.

[32] I guess that in most cases an agent also knows how to Φ prior to her Φ-ing. However, that is not necessary. Knowing how to Φ may come about as the result of one's decision or intention-formation. It is nevertheless true that Φ-ing intentionally requires knowing how to Φ.

[33] Hossack is not alone in emphasizing an intimate relationship between having a mind and having knowledge. Williamson for instances entertains the 'conjecture' that '[i]f a creature could not engage in such processes without some capacity of success, we may conjecture that nothing could have a mind without having a capacity for knowledge' (2000, p. 48). Since Williamson takes knowledge to be a mental state, however, he is precluded from defining a system as a system of mental states in terms of knowledge. In this respect, Hossack's framework is better suited than Williamson's to illuminate the notion of a mind. Since Hossack conceives knowledge not as a mental state but a basic

To take an example: if we assume that beliefs are mental states, some creature, *C*, may only have a belief-state, *Bp*, on the condition that *C* is a knower and thus that some of its psychological states are brought about by an appropriate exercise of an epistemic faculty. Suppose that some creature, *C**, fails to be a knower in this sense, but that she has a state, *B**, towards some *p* and *B*p* is functionally identical to how *Bp* functions in the cognitive economy of *C*. Unlike *C*, however, *C** is *not* a genuine knower of facts. Hossack will then deny that *B** – even though it enables its bearer to respond to environmental changes in a way that involves a minimal kind of psychology – can be the mental state of belief. A virtue of this view is that it makes it 'a substantive question of fact' whether some given individual thing or agent can be said to possess not only functional states and operations but a mind. It would, for instance, allow us to deny that the fact that a computer program has states that are functional duplicates of the states of a living, conscious mind is inductive evidence for the claim that the computer program instantiates an individual mind. We could deny this on the grounds that the operations of the computer program fail to cause the computer to know things.[34]

Regardless of whether or not this is a plausible account of belief-states – or other mental states – I do think the framework can be put to our advantage on the issue of intentional agency and knowledge in action. We should take care to notice that Hossack may only use knowledge to define mental states and acts – or mindedness - because he takes knowledge to be metaphysically distinct from such mental states and acts. In his view, knowledge is a 'direct, unmediated mental relation to the world' that relates the bearer of the faculty directly to facts without intermediaries such as representations, contents, propositions, directions of fit, other mental states or events, and so on (Hossack 2003, p. 193). Knowledge is not itself a mental state or composed by such states. It obtains in virtue of the fact that the mind is constituted in certain ways by its mental states, attitudes and mental acts that thereby suitably relate the individual to some fact, *f*, in favourable circumstances, $C_1 \vee C_2 \vee,...,\vee C_n$. In this respect, the existence of knowledge on an instance is, of course, due to operations of the mind; however, the truth of the *Causal Thesis* offers a reading of this 'due-to' and 'in-virtue-of' relation where the relationship between the knowledge-relation and operations of the mind is causal rather than constitutive. Given Hossack's view of what the nature of knowledge is, it follows that he can define some *F* as an epistemic faculty insofar as it has the power to bring

mental relation it allows him to define mental states in terms of knowledge, whereas on Williamson's framework it is impossible to further illuminate the mind via the state of knowledge since knowledge just is a mental state. See Hossack (2007, chs. 5 and 6).

[34] See Hossack's discussion (2007, ch. 8).

about this metaphysically simple mental relation that connects minds directly with the facts. Moreover, he can then define mindedness in terms of having such faculties.

If we turn to intentional agency and the claim that the will is an epistemic faculty, we may use knowledge in a similar role in order to understand the kind of mindedness that is distinctive of intentional agency. So far I have affirmed at several junctures that intentional actions can obtain when their agent is absent-minded or not consciously attentive to her own activities; however, the idea running through the entire thesis has nevertheless been to inquire about the nature of the sort of mindedness or awareness that persists as a constitutive feature of the actions that are intentional. I have referred to this mindedness as agent's knowledge and characterized it in terms of the principle *AK*. Agent's knowledge makes the difference between intentional activities and those that are merely motivated in a way that is strikingly similar to how Hossack takes knowledge to help make a distinction between what is a belief and what is merely a functional duplicate. Like Hossack, we can say that some agent, *A*, is an intentional agent or acting intentionally only on the condition that *A* is a creature with a will and thus has a capacity for having practical knowledge of one's action. Some agent, *A**, that is Φ-ing may be a functional duplicate of how *A* is Φ-ing with the only difference being that *A** lacks the ability to know in a practical way that she is so Φ-ing. *A** will then not be able to Φ intentionally. *A**'s performance may resemble *A*'s action, but is nevertheless of an entirely different nature despite being a functional duplicate of *A*'s activities. That distinction is due to a difference in the nature of the faculties possessed by *A* and *A**. In particular, that the action in *A*'s case but not in the case of *A** is brought about by an epistemic faculty – viz. the will.

The will is, as I have said earlier, 'the ability to choose and perform intentional actions' (Raz 1999, p. 48). Given the claim that the will is an epistemic faculty we can now say what possessing a will involves: namely, that it allows its bearer to exhibit a crucial sort of mindedness in action. The proposal is thus that the will – viz. the faculty whose exercise enables agents to act intentionally – is distinctive in being a faculty that has the power to cause its agent to have knowledge of the nature of what she is doing in virtue of so doing it. What I suggest is thus that we understand the will as a faculty that causes the agent to know what she is doing by virtue of exercising her distinctive mental act of *willing* something or choosing something (*prohairesis*) insofar as she executes this act of apprehension under favourable circumstances. Analogous to Hossack's accounting for a system of states as beliefs, we can therefore define a system of motivational acts and performances as intentional agency only insofar as the operations of the will whereby its agent acts cause the agent to

know the object of these acts of willing. Given that the objects of these willings are actions we may thus attribute some creature with a will only insofar as it knows its actions in virtue of so willing them. In a word, I suggest that it is an individual's capacity for bringing about knowledge of what she is doing that grants her the mindedness that marks intentional agency:

> *The Will*: Some agent, *A*, has a will insofar as she knows that she is Φ-ing for some reason, *R*, by virtue of willing the Φ-ing for *R* under favourable circumstances (viz. insofar as so willing causes knowledge of this fact).

The knowledge one gets by virtue of participation – viz. by willing, acting and so on – is practical knowledge. We can therefore say that some agent, *A*, is an intentional agent only insofar as *A* has this capacity for practical knowledge of her own actions, as it is portrayed by *The Will*. In agents like us, practical thinking may produce intentional actions if the behaviour is performed under certain conditions. We are able to take our circumstances into account and decide to treat or take something as our reason for acting and then execute the action under the guidance of our intention to so act. We can also, however, imagine agents being capable of similar operations that produce comparable movements and behaviour without ever being able to rise to the occasion, as it were, where they treat something as their reason for so Φ-ing. If an agent were only capable of the kind of behind-the-back agency that was revealed in the case of *Freud's Inkstand* – viz. where Freud carried out a purposive action but where both the nature of his act and his motivation were subconscious and hidden from view – we could recognize its operations as being functionally on a par with much of what is going on in our own agency. Nevertheless, there would be a crucial difference in the faculty that governs these operations and the faculty that we exercise when we act intentionally. That difference is the distinction between intentional agency and merely motivated behaviour or action. What I suggest is thus that we distinguish our agency from such rigidly Freudian agents by seeing our agency as the exercise of a faculty whose distinctive feature is that it can cause its agent to have practical knowledge of what she is doing and why: namely, the will.

By this principle we enable ourselves to define or demarcate a range of important particulars or individuals: namely, the individuals that are intentional agents. That, however, will only allow us to explain the necessity of having a will insofar as one is acting intentionally and thus that so acting must go together with possessing the epistemic capacity whose power it is to cause itself to have practical knowledge by virtue of willing or participating in one's Φ-ing. What I need to explain as being the truth of *AK* is that so acting

necessarily is an instance of such knowledge-production. Now it is a distinctive feature of the will that it grants its bearer with an ability to go beyond producing mere behaviour or merely purposive and merely motivated actions, and so rise to the level of intentional agency. What I suggest is that this feature of the will gets explained by its knowledge-producing ability.

Recall that in chapter 1 Velleman defined the problem of action as locating 'an agent at work amid the workings of the mind' in order to explain the mechanism by which 'an agent participates in [their action] in the sense that he does it' (1992b, pp. 128n5, 131). In relation to this he says something which in my view reveals the true difficulty and nature of this task:

> *For let there be mental states and events in abundance – motives, reasons, intentions, plans – and let them be connected, both to one another and to external behavior, by robust causal relations; still, the question will remain how the existence and relations of these items can amount to a person's causing something rather than merely to something's happening to him, albeit something mental.*[35]

The irony is that, after having said this, Velleman goes looking for just another mental state: namely, 'a mechanism modifying the motivational forces already at work' (2000b, p. 12). He therefore still thinks that the job can be done in action theory if we can only locate 'mental events and states that are functionally identical to the agent, in the sense that they play the causal role that ordinarily parlance attributes to him' (1992b, p. 137).

I think a different lesson could be drawn when we consider Hossack's framework and the way he suggests that we treat mental states and mindedness. In particular, I think we should take Velleman at his word when he rejects the utility of invoking mental states and events in abundance. Perhaps what we need is to invoke a more basic mental relation, something whose nature is not that of being a mental state, attitude, or mental act. In other words, perhaps knowledge will do. What I have in mind is that from the conception of the will as an epistemic faculty we can use its function as a knowledge-producer to define its successful operations. We can capitalize on this model of an epistemic faculty and say that an act of will or an instance of practical thought is successful insofar as it causes its agent to know. Now an instance of such practical thinking – viz. a string of thinking with a view to action – is, of course, successfully carried out if it produces an intentional action by virtue of treating some *R* as one's reason for so acting. We would then also say that when the agent carries out these operations in the right way her performance is intentional and done for a

[35] Velleman (1992b, p. 131).

reason. The result is that an agent acts intentionally if, and only if, she has practical knowledge of what she is doing and why by virtue of exercising her will. What we might be in position to specify when the will is defined as an epistemic faculty is therefore what makes it the case that some action is intentional. The crucial bit is whether the will operates and thereby causes its agent to know what she is doing and why. The claim that the will is an epistemic faculty is thus a way to explain instances of successful practical thought – viz. intentional actions – as those actions the performance of which causes its agent to know them.

Now, at this stage a caveat is due. I do not take myself to have ultimately established the framework I am favouring. Accomplishing that would require thorough comparisons with other candidate views about agent's knowledge and intentional agency. What I have suggested is rather an interesting possibility whose ultimate viability must and will be considered on another occasion. That said, the problems pertaining to both the cognitivist approach proposed by Setiya and the inferential model championed by Paul do speak in favour of the idea that agent's knowledge is practical knowledge in that it comes about in virtue of distinctively practical attitudes and mental acts. That possibility, in turn, speaks in favour of Hossack's *Causal Thesis* given the difficulties of accommodating practical knowledge under the framework of the *Constitutive Thesis*. Finally, I have suggested that Hossack's conception of the nature of knowledge – namely, as a mental relation holding between minds and facts without intermediaries – would allow us to define the nature of the will and the range of actions that are intentional. All this speaks in favour of the framework I am operating with. I shall return to argue more directly in chapter 7 in favour of Hossack's view with respect to the competitor models in the metaphysics of knowledge. For the rest of this chapter, however, I shall discuss how the framework can accommodate features of practical knowledge (§5) before discussing its wider consequences for action theory (§6).

§5 FEATURES of PRACTICAL KNOWLEDGE

In this section I will discuss the ability of the view to accommodate some of the characteristics of the knowledge we have in action that we have encountered on our way. In particular, I will focus on how it works as a model of spontaneous knowledge in allowing for immediate, non-observational knowledge of what happens. However, I will also say something about how this framework allows us to see why it is correct to say that a mistake in yielding such knowledge is a mistake 'in the performance, not in the judgement' and that it is

'the cause of what it understands' (Anscombe 1957, pp. 82, 87). I shall also explain how this view can accommodate Anscombe's claim that this knowledge is sensitive to descriptions.

Let me begin to address these issues by way of responding to what Thor Grunbaum (2009) takes to be a sort of 'paradox' for defenders of the notion of practical knowledge. According to Grunbaum, there is a tension between what he labels the factual (or external) and practical (or internal) aspects of practical knowledge (2009, §3). Anscombe's discussions in *Intention* – at least as I have interpreted them so far – singled out four important features of practical knowledge: (1) that as a kind of knowledge it must be factive and thus relate its agent to a fact; (2) that this knowledge pertains to what I do and so will have as its object something that is 'at a distance' (recall her slogan 'I do what happens'); (3) that it is something we acquire spontaneously and non-observationally; and (4) that having knowledge of what we are doing is to have knowledge both of the relevant action-type and the reason why that act is performed.[36] Grunbaum takes (1) and (2) to articulate the factual constraints of practical knowledge whereas (3) and (4) articulate its practical constraints. He then goes on to argue that these constraints together with a couple of 'commonsense assumptions' will allow us to derive inconsistent claims (Grunbaum 2009, §3). In particular, he takes the truth of (1) and (2) to exert pressure on (3) given the commonsense assumption that 'our only means of acquiring knowledge of events and facts in the world beyond our body' (2009, §3) – viz. of things that are at a distance from the agent – is either (a) by observation (or testimony); or (b) by inference. Knowledge of what we are doing when that includes distal events thus seem to require ways of knowing that involve (a) or (b) (or a combination of the two) with the result that (3) is rejected. On the other hand, constraint (4) may provide basis for an argument if we combine it with the intuitive claim that observation, testimony, and inference are the wrong kinds of knowledge. That is something we have been arguing in favour of for quite some time now so this argument will also count in favour of affirming (3). The constraints (1), (2) and (4), plus a couple of plausible assumptions, can therefore both be used to affirm and reject (3).

According to Grunbaum, 'most contemporary theories of practical knowledge' are distinguished by their 'ways of responding [at least implicitly] to this paradox of practical knowledge' (2009, §4). Proponents of the so-called 'two-factor thesis' will, for instance, respond by claiming that the so-called paradox trades on an ambiguity.[37] According to them,

[36] Anscombe seems to accept the factivity-condition in point (1) when she says that 'I have called such a statement [viz. the reply 'I am opening the window'] knowledge all along; and *precisely because in such a case what I say is true*' (1957, p. 51 [my emphasis]). Point (2) is emphasized at several instances throughout *Intention* and grounds the kind of thinking that is behind her slogan 'I do what happens'; see in particular (1957, pp. 52–4). Point (3) is also evident from her discussions; see (1957, pp. 50–4) and my discussion in chapter 2, §4. Point (4) encapsulates my discussion in chapter 2, §2; also see Anscombe (1957, pp. 57, 80; 1983, pp.179–80).

[37] See Falvey (2000, p. 21).

non-observationally knowing our actions could only mean that we have non-observational knowledge of what we intend to do or are trying to do; however, that is not inconsistent with assuming that the agent is in need of observation/testimony/inference to yield knowledge of the full action if the nature of the action depends on events beyond our bodies. I may thus non-observationally know that I am trying to spill my coffee by shaking my cup, but must call on observation/testimony/inference to yield knowledge of the fact that I spilled my coffee. Since I have already raised critical points against this view I will not bring it up for serious scrutiny again; however, it is useful to see it fit in with the dialectics Grunbaum puts up.[38]

Another way to go is to hold on to the one-factor, spontaneous knowledge by weakening the externality of the account. One could, for instance, give up (1) and take knowledge of what one is doing to have other practical constraints rather than factivity; however, that comes at the cost of requiring a new, *sui generis* and perhaps 'funny sort of knowledge', as Anscombe says (1957, p. 82).[39] A better way to go would be if one could respect the factivity condition in point (1) but weaken the account's threatening externality by virtue of qualifying (2). Anscombe seems to be driven in this direction when she speculates whether knowledge of what one is doing – such as when I am writing 'I am a fool' on the blackboard – is the sort of knowledge that is 'independent of what actually happens' (Anscombe 1957, p. 82). Thompson (2011) suggests that passages like these support a reading of Anscombe's account where the objects of spontaneous knowledge are exclusively given by contents that portray its agent's activities in the progressive and imperfective mode. What they know *qua* being an agent is what they are doing *in medias res*, as it were, and these contents portray what the agent is doing rather than what she has done or achieved. Our practical knowledge of what we are doing is therefore exclusively given by propositions whose truth will not be altered if the big H-bomb goes off and seriously interrupts all our activities. Its character as knowledge will therefore not be disrupted by the bomb.[40]

There is clearly some truth to this. As an example goes in Grunbaum (2009, §4.1): if I decide to make myself a pot of tea, the correct description of my activities – viz. filling the kettle with water, finding the teapot, taking down a cup, and so on – may still be that I am making tea even if I mistakenly pour coffee in the teapot rather than adding tea leaves in a teabag. It is not as if I fail to know what I am doing – viz. making tea – even if on this occasion no tea is actually made. In response to my mistake I may pour the coffee out and

[38] See chapter 2, §3.
[39] As the question goes in the *Republic*, 'for how could something that is not be known?' (Plato *R* 477a)
[40] Thompson thinks this is the way to read Anscombe as many of her examples describe what agent's are or was doing.

take greater care this time to make sure it is the tea leaves that get in the teapot rather than coffee. Or I may change my mind altogether and just settle for coffee. The right thing to say would nevertheless be: 'I was making myself a cup of tea when I suddenly realized I had made coffee instead.' My mistake is the mistake that Anscombe reports Theophrastus as saying 'is in the performance, not in the judgement' (1957, p. 82).[41] The problem with this account, however, is its tension with the idea that knowledge of what I am doing can be about something that is at a distance from me and pertain to what happens. In other words, it is in tension with the idea that 'it is the actual accomplishment in the world that is said to be known in this way', as Moran puts it (2004, p. 55). Thompson, on the other hand, embraces this consequence when he goes on to accept that it follows from this account that my knowledge of my intentional action in truth exists only and precisely when there is no action, but only something I am doing and thus something of which more is to come (2011). On Thompson's view our knowledge in action – or *qua* being an agent – thus pertains to 'that which becomes but never is', as Plato says (*T* 27d), since as soon as it becomes an action our knowledge of it will no longer be practical but something based on observation. This account will therefore never yield spontaneous knowledge of an accomplished action.

The account I have sketched out above, however, is in better position to provide spontaneous knowledge of such accomplished actions since it is natural to go on and reject the common-sense assumption that spontaneous or non-observational knowledge cannot yield knowledge of something external. Given that practical knowledge is caused by our participation in the events that are our intentional actions under favourable circumstances we can allow ourselves a notion of spontaneous knowledge whose objects are a full action. When we are Φ-ing intentionally we are either exercising our ability to directly Φ intentionally – viz. if Φ-ing is a basic intentional action[42] – or else one is Φ-ing by doing something else intentionally (where that something else is more than just trying to Φ or intending to Φ). What my framework allows for is that we have spontaneous knowledge of at least some of the things we did intentionally – where that something is more than just having tried or intended to Φ – insofar as the Φ-ing was intentional. This framework is therefore able to disarm the apparent tensions between (1), (2), (3) and (4) without relaxing either of them since it provides a model of knowledge on which the commonsense assumption can be rejected.

[41] By speaking of Theophrastus Anscombe (1957, pp. 2–3) is actually referring to a sentence in Aristotle's *Magna Moralia* (*MM* 1189b22). The explanation is that it is often assumed that, even though the *Magna Moralia* is numbered among Aristotle's works, it was his younger student and colleague, Theophrastus, who was its author. Also see Teichman (2008, p. 22n10).

[42] In chapter 2, §3, I introduced Setiya's notion of a basic intentional action as a development of Danto's notion of basic actions (Danto 1963; 1965; and 1973, ch. 2). This is something one can intentionally do without needing to do something else intentionally (Setiya 2008b, p. 390).

The framework also allows us to explain why 'Theophrastus' remark holds good,' as Anscombe says (1957, p. 82). What will impinge on the issue of whether one has practical knowledge or not is the performance or the agent's participation in her own actions. If something goes wrong so that her participation fails to cause knowledge – either because an action fails to come about, it happens in the wrong way, and so on – the agent will fail to acquire knowledge of what she is doing in a practical way. She may still know what she is doing or not – say by observation, testimony, and so on – but since this knowledge will fail to come about in a practical way it will no longer be practical knowledge or agent's knowledge. Given our definition of intentional actions in terms of the will's production of such knowledge, this failure means that it will no longer be the case that 'the event in question counts as an intentional action', as Moran says (2004, p. 67). A failure to gain agent's knowledge will thus find its explanation in features that pertain to the agent's performance and participation; and in this way, it will be true to say that her mistake is in the performance.

Towards the end of *Intention* Anscombe also says that 'it is the agent's knowledge of what he is doing that gives the descriptions under which what is going on is the execution of an intention' (1957, p. 87). She then sums up her considerations by quoting Aquinas who said of the nature of practical knowledge that it is 'the cause of what it understands' (*ST* 1a.2ae.3.5.1). When Anscombe proceeds to clarify these statements she says they mean more than that agent's knowledge is a necessary condition; rather, it means that 'without it [i.e. the agent's knowledge of what he is doing] what happens does not come under the description—execution of intentions—whose characteristics we have been investigating' (1957, p. 88). As Grunbaum points out, it seems as if she is making the 'constitutive claim that it is this knowledge that somehow makes the action intentional' (2009, §4.3). At several places Anscombe characterizes intentional descriptions as describing 'what *further*' an agent is doing '*in* doing something' and that the intentional description is an 'enlarged description' of what an agent is doing compared with the '*more* immediate, nearer to the merely physical' nature of the latter (1957, p. 86). She mentions for instance that a cat is stalking a bird '*in* crouching and slinking along with its eye fixed on the bird and its whiskers twitching' (1957, p. 86). Similarly we could say that I am baking bread in kneading the dough, in having my eyes fixed on a certain page in the cookbook, fingers running over the page, and so on. Such enlarged descriptions can be given for any event that has a describable effect. We could speak of trees that 'drop their leaves or their fruit', or of a meteor that loses particles the size of dust in its revolutions around the sun; so it is not that the event of stalking a bird allows an enlarged

description that would make it – if it is – an intentional action. Rather, Anscombe seems to think that those enlarged descriptions are parasitic on the intentional form and that when we combine enlarged descriptions with agent's knowledge we get what is 'quite characteristic of description of intention in acting' (1957, p. 86). As an illustration, she once again mentions the cat and says that it is truly stalking the bird – viz. in crouching, slinking, ... etcetera – because it has knowledge of these things as being organized with the bird in view. What seems to be going on here is that enlarged descriptions of what further the agent is doing apply in the primary case as an intentional form because the agent has agent's knowledge that corresponds to the form of the enlarged descriptions. Anscombe's suggestion thus appears to be that agent's knowledge is the ground for these enlarged descriptions by being the ground without which nothing that happens would come under an intentional description. Thus, agent's knowledge appears not only to be a necessary condition for acting intentionally but a state the possession of which underpins the whole edifice.

What Anscombe seems to have in mind with her remarks is thus that an agent's knowledge in action – viz. her spontaneously knowing what she is doing and why she is doing it – grounds the form of descriptions of events that allows us to see them as intentional actions. In other words, what makes it right to describe some Φ-ing as an intentional action – viz. by adding to its description that its agent is Φ-ing 'in order to *A*', 'because *P*' (where '*P*' is the agent's reason for Φ-ing), and so on – is that the agent has what I have defined as practical knowledge of the events so described. I think we are a step closer to understanding that by having defined the will as an epistemic faculty and an intentional action as the successful exercise of this epistemic faculty (viz. an act that causes its agent to have practical knowledge of what she is doing and why). On this framework it becomes clear that insofar as an agent is Φ-ing without this knowledge what happens does not come under the description *execution of intentions*, if I may paraphrase Anscombe, whose characteristics we have been investigating. In other words, it is the possession of this knowledge – which is a substantive fact – that, as Moran points out, 'determines which descriptions of "what happens" may count as descriptions of what the person is intentionally *doing*' (2004, p. 54).

To fully understand Anscombe's idea we need to take into account what Anscombe means when she says at a crucial turn in *Intention* that '[t]he notion of "practical knowledge" can only be understood if we first understand "practical reasoning"' (1957, p. 57). She then goes on to investigate the notion which she takes to be 'one of Aristotle's best discoveries' despite the fact that its 'true character has been obscured' (1957, p. 58): namely, the practical

syllogism, or reasoning with a view to action. Knowledge may be the crucial bit that puts an event of doing something up as an event of intentionally doing something; however, saying that is not enough to explain the intentional form that 'describes an order which is there whenever actions are done with intentions' (1957, p. 80). On our framework, however, that order may be explained as a result of the provenance of practical knowledge. As we recall, on Hossack's framework a knowledge-report reveals both the fact that its subject knows as well as its mode of knowing. It is thus at the level of the provenance of knowledge we find mental states, acts and processes and their corresponding descriptions in virtue of which their agent knows some fact, *f*. The fact in a case of practical knowledge is that one is doing something, Φ-ing, for a reason, *R*; and I have suggested that knowledge of this fact is practical insofar as its provenance of knowledge – viz. its mode or way of knowing – is distinctively practical. Reporting such knowledge must therefore make sure that the relevant mode of knowing – and its corresponding intentional form or description – gets properly ascribed to the agent. To use an example of Thompson's, if 'I am laying bricks because I am building a monument to the great works of Frege' (Thompson 2008, p. 87), the knowledge you ascribe to me by reporting what I know as an agent should reveal the relevant mode of my presentation of this fact. You would presumably get something wrong if you reported my practical knowledge of this fact via the description 'he is laying bricks because ... a monument to the author of *Über eine geometrische Darstellung der imaginären Gebilde in der Ebene*'.

Practical knowledge therefore has a formal or constitutive role with respect to the intentional description that portrays the event that is one's action as an intentional action. 'If the agent didn't *know* this happening under *this* description,' as Moran says, 'then as so specified it *would not be* "what he is intentionally doing"' (2004, p. 54). By defining practical knowledge as knowledge caused by a practical way of knowing we may explain this constitutive role of practical knowledge. This knowledge pertains to the happening that is our action; however, we only have this knowledge in virtue – viz. by being caused by – a specific way of knowing: namely, the exercise of our will. That exercise contains descriptions of one's action that match the intentional descriptions whose characteristics we have been investigating. This participation – the will's way of knowing – will therefore impose the intentional order that Anscombe talks about and it imposes it because it is the order under which this fact gets known by its agent in virtue of acting intentionally. The fact that our way of participating in our actions causes its agent to have practical knowledge of the fact that she is Φ-ing for some reason, *R*, is what gives ground for the intentional descriptions in the first

place. In a quite literal way this puts the whole institution of intentional agency and enlarged descriptions up, as it were. In this way, it is true to say that practical knowledge is the cause of what it understands only if by 'cause' one means the logical ground, reason (*ratio*) or formative cause of what it is knowledge of; and most readers take it that it was such a notion that Anscombe had in mind when she quoted Aquinas with sympathy.[43]

§6 IMPACT on ACTION THEORY and RATIONALISM

A key advantage of the model I have been favouring is, as I have emphasized, that the *Causal Thesis* allows us to conceive of agent participation and taking as one's reason as an epistemic relation without needing to impose alien cognitive elements onto the process by which we act intentionally. In fact, on this view there are no elements we need to impose on the nature of this process from the claim that we have practical knowledge of our own actions. What this claim involves is that the knowledge we have of our actions comes about in virtue of – viz. by being caused by – the distinctively practical mental states, attitudes, or acts that compose agent participation or choice (*prohairesis*) when the exercise of our will goes about under epistemically favourable circumstances. This model of practical knowledge is, however, compatible with a range of views about the nature of the will's way of knowing. Since the will's way of knowing just is the intentional activity, it follows that this model is compatible with a range of views about the nature of intentional actions. What I shall argue in this section is therefore that the issue of agent's knowledge is not something one can impose on action theory to settle the dispute that Setiya has with proponents of Ethical Rationalism.

According to O'Brien, the sort of view I am recommending – a view where agents know what they are doing through and by virtue of agent-participation – will require us to explain what 'way or mode of acting' makes an action knowable through agent-participation (O'Brien 2007, p. 183). If one operates with a constitutive view of knowledge, such a requirement will provide an impetus on how we should understand the nature of intentional actions. Setiya, for one, took it as counting in favour of 'reductive psychological accounts of acting for reasons' where we would need to 'break down what is involved in acting for a reason' in order to 'see how one of its elements is, or involves, the relevant belief' or other suitable knowledge-bearers (2007b, p. 28). Setiya's kind of inquiry makes sense if one

[43] As Moran says, 'the sense of the phrase from Aquinas is not about the efficient cause role of intention in producing movements, but rather concerns the formal or constitutive role of the description embedded in one's practical knowledge making it the case that *this* description counts as a description of the person's intentional action' (2004, p. 54). Also see Paul (2009a, p. 2). There is historical precedence for such a notion of 'cause'. Spinoza, for instance, has primarily such a notion in mind when he talks about a things cause and means sufficient condition when he says 'efficient cause'. The notion that pertains to billiard balls hitting one another is better rendered as 'mechanical causation'. See Spinoza (*E* p. 25 [*Translator's Preface* §12]). As I have used the notion elsewhere, I have meant mechanical causation.

remains within the framework of the *Constitutive Thesis*. For if one takes it to be essential an action to be intentional that its agent knows what she is doing, what the *Constitutive Thesis* takes to be basic building blocks of knowledge must also be components of acting intentionally. The *Constitutive Thesis* would therefore encourage us to go looking for enabling conditions of knowledge – viz. conditions under which it is possible to know (Cassam 2007, pp. 16–7) – among the enabling conditions for acting intentionally. In this respect, Setiya's strategy would be a natural way to deal with what is known in the literature as a *what-makes-it-possible* question, or the kind of explanation-seeking question that is not so much an effort to overcome an obstacle but rather asks for a positive explanation of the possibility of some philosophically interesting phenomenon in relationship to certain background assumptions we thereby allowing us to see 'how things fit together', as Nozick says (1981, p. 10).[44]

But, as Cassam points out, 'what strikes one as a worthwhile how-possible question is bound to be influenced by one's background assumptions' (2007, p. 23). Given our framework there is no reason to start looking for enabling conditions of knowledge in the process of acting intentionally and taking as one's reason. After all, if this framework is on to something knowledge has no interesting enabling conditions other than itself. Knowing something is 'a *bare* truth', as Dancy puts it (2000, p. 162). Moreover, we also found that the distance between the state of knowing and the ways of knowing allows for a plurality of ways of knowing as long as the way of knowing is an exercise of an epistemic faculty. In this respect, I have said more than just that 'knowledge is the state you get into when you have exercised some ability, whatever that might be, to acquire knowledge', which, as Millar rightly points out, 'would be thin beer indeed' (2007, p. 188). We provided more than just thin beer, however, by claiming that the will is an epistemic faculty and then defending the idea that the proper exercise of this ability causes its agent to know what she is doing. We then capitalized on the necessity of agent's knowledge in intentional agency, as articulated by *AK*, to favour the claim that intentional actions are definable as the successful exercise of this faculty. That is, we have said that insofar as the will's practical way of knowing causes its agent to know what she is doing and why, that activity is an intentional action.

Given the *Causal Thesis*, however, the flipside of this framework is nevertheless that there appear to be little – if any – substantial constraints for determining the nature of agent participation that can be derived from the truth of this account. I therefore think that

[44]See Cassam (2007, pp. 16, 23). Also see Nozick (1981) and McDowell (1998d, pp. 16–8). As McDowell points out elsewhere, 'one's first move, if someone tries to interest one in a "How is it possible?" question' – viz. in its obstacle-mood – 'should be to ask: why exactly does it look to you, and why should it look to me, as if such-and-such a thing (...) is *not* possible?' (McDowell 1998c, pp. 57–8)

Anscombe was clearly on to something when she pointed us to the notion of practical reasoning in order to understand the notion of practical knowledge (1957, p. 57). By so doing, Anscombe made the connection which I think any plausible understanding of human intentional activity must make: namely, to connect the nature of an intentional action – or the object of what one knows when one has practical knowledge – with the agent's practical thought to see the nature of what can be 'said to prompt an action' (Anscombe 1957, p. 65). By pointing to Aristotle's notion of practical reasoning, Anscombe pointed to the notion that she saw as part of an answer to what makes something 'operate as my reason' so that what I am doing is 'done *in pursuit* of the end and *on grounds* of the belief' (Anscombe 1995, p. 110 [original emphasis]). Within our framework Anscombe's move is natural. After all, what prompts an intentional action by virtue of making something operate as my reason for acting is the way of knowing that is causing its agent to have practical knowledge of what she is doing and why. This way of knowing is nothing other than the agent's participation in her own actions insofar as she is acting intentionally and for reasons. Understanding the nature of this participation may thus require an investigation of practical thought or reasoning.

The difference between Anscombe's move in *Intention* and what Setiya pursues in *Ethics without Rationalism* is that Anscombe inquires about the notion of practical thought in order to illuminate the notion of practical knowledge. Setiya, on the other hand, takes our knowledge in action as his springboard into an inquiry of the nature of practical thought and what he takes to be a 'central instance of practical thought', namely acting intentionally or for a reason (2007b, p. 8). What I want to point out here is not whether Anscombe's account of what prompts an action is right or not, but that we will have reason to go in Anscombe's direction given the framework we have adopted in the metaphysics of knowledge and our model for practical knowledge. The reason is that for agent's knowledge to provide a springboard into action theory we must have a clear idea of what one is looking for. You will have that within the framework or the *Constitutive Thesis*. The nature of knowledge will then provide some constraints on action theory. Within the framework of the *Causal Thesis*, however, there are no such obvious elements to look for other than the state of knowledge itself. It is more open what the nature of its provenance is given that the relationship between the provenance of knowledge and the knowledge-relation is causal rather than constitutive. There are therefore no clues to be had about the nature of the causal antecedents of practical knowledge – viz. our acting intentionally – from the nature of the knowledge-relation itself.

One thing that follows is therefore that our explanation of the truth of *AK* will make no impact on the dispute between Setiya and ethical rationalists. Rationalists, as you may recall, seek to vindicate morality from practical reason – a project that requires what I have referred to as a normative interpretation of intentional agency where the nature of intentional actions allows us to recognize what the good standards of acting are independently of what the good moral standards for such actions are. To fully understand the nature of practical knowledge we must, as Anscombe does, embark on an inquiry into the nature of practical thought. However, we are not aided in this inquiry by the fact that intentional actions are intimately connected with practical knowledge. For all we know we may therefore find that the success of our inquiry into practical thought will depend on invoking principles and constraints that are friendly to the rationalist project. This is not the time and place to execute and pursue such an inquiry; but it might be worth noting that Anscombe herself takes it that 'What's the good of it?' is something that can be asked of an intentional action 'until a desirability characterisation has been reached' about which this question 'does not arise' since the action has been made intelligible (1957, pp. 74–5). This 'good' that Anscombe talks about here, she goes on to say, is due to a conceptual connection between wanting and 'what the agent conceives to be good' (1957, p. 76). Although she emphasizes that this 'good' is a multiform notion that requires its agent to see what she wants under 'the aspect of *some* good', Anscombe is nevertheless stressing that it 'must *really* be one of the many forms of good' (1957, pp. 75, 77 [original emphasis]). I am not saying that she is right; but it seems as if the required inquiry into the nature of practical thought may give hope to rationalists.

Recently, Setiya has sought to block such moves by arguing that the kind of generalization that Anscombe makes here relies on a mistake of reasoning in deriving universal constraints of rational and intentional agency from general considerations connecting human action, choice and desire 'with appearances of the good' (Setiya 2010, p. 102). According to Setiya, this may show that '[t]he source of the guise of the good in action theory may be a familiar but tempting parochialism: the mistake of thinking that our characteristic form of agency shows us what agency essentially is' (2010, p. 104). In other words, Setiya is willing to grant the truth of such observations, but he nevertheless want to deny that they support the claim that 'any possible instance of acting for a reason is an instance of acting under the guise of the good' (Setiya 2010, p. 101). Apparently, Setiya does not think support for principles pertaining to rational agency can be found in generalizations that pertain to our way of being rational agents. As a methodological principle, I agree with

Velleman in thinking that the philosophy of action should in principle 'set about describing possible agents, without concern for whether they are real' (2007, p. xxxi). We may then find that a feature pertaining to our ways of being an agent is no more than an idiosyncrasy owing to our nature. It will definitely require more work to decide whether this is true or not for the principles invoked by proponents of Rationalism. What I have argued in this last section, however, is that an understanding of agent's knowledge does not provide a suitable explanatory principle for action theory. To make further progress, then, we may find that the rationalist's principles will recommend themselves. This requires further inquiry. We now know that it will not be easy to establish principles of rational agency rather than general truths about human beings and their behaviour. But that is as it should be; after all, 'preaching morals is easy; grounding morals hard,' as Schopenhauer says (*WN* p. 372).

§7 CONCLUSION

Schopenhauer once said that 'what as representation of perception I call my body, I call my will in so far as I am conscious of it in an entirely different way comparable with no other' (*WWR* §20). What our considerations suggest is something similar: namely, that what as the object of non-practical ways of knowing I call my merely motivated actions, I can call my intentional actions only insofar as I must know them in an entirely different way comparable with no other. I have argued that this way of knowing is knowledge by willing or through participation, which is a state of knowing caused by an agent's distinctively practical participation in the events that are her actions. In one of his letters, Spinoza compares our sense of participation, freedom and agency with a moving stone that has become conscious of its own endeavour. Scornfully, Spinoza derides that 'human freedom which all men boast of possessing, and which consists solely in this, that men are conscious of their desire and unaware of the causes by which they are determined' (1992b, p. 286). We are now in position to say that Spinoza was somewhat inaccurate at this point. What the stone would need to have to be anything like us is not merely knowledge of its own endeavour but practical knowledge. Human beings rise to the level of rational agency not merely by being conscious of their motives and actions but by being aware of them or knowing them in a way comparable with no other: by virtue of actively participating in their own actions. Human agency – whether it implies freedom or not – involves more than mere awareness: it involves practical knowledge.

7

The Causal Thesis of Knowledge

> *"I know" is supposed to express a relation not between me and the sense of a proposition (like "I believe") but between me and a fact. So that the fact is taken into my consciousness.*
>
> Ludwig Wittgenstein[1]

§1 INTRODUCTION

In the previous chapter, I proposed that the *Causal Thesis* and the metaphysics of knowledge suggested by Hossack (2007) could be used to understand the nature of knowledge in action as a distinctively practical way of knowing. There I emphasized that even though this framework has its advantages over its competitors – viz. different versions of the *Constitutive Thesis* – with respect to the issue of knowledge in action, it can also be supported by independent considerations in the metaphysics of knowledge. Given that one may be unaccustomed to Hossack's framework – or worried about some of its ontological commitments, such as its dependence on an ontology of facts – and given that it had some interesting consequences for action theory all by itself, I want to spend this final chapter providing support for the framework on independent ground. I shall argue that the framework recommends itself as the best way to conceive the metaphysics of knowledge. In particular, I shall argue that the *Causal Thesis* is favourable since there is reason to reject both the idea that knowledge is a metaphysical hybrid constituted by mental states and suitable circumstances (§2) and Williamson's claim that knowledge is a mental state or a factive stative attitude (§3). I will then show how the *Causal Thesis* avoids these problems and that it recommends itself as a suitable framework for conceiving the metaphysics of knowledge (§4).

[1] Wittgenstein (*OC* §90). I refer to the 1969 edition of *On Certainty*, which was reprinted in his *Major Works* (2009).

§2 The METAPHYSICS of KNOWLEDGE

Knowledge, as I have said before, is primarily a way to describe a mental condition of an individual. It is thus meant to denote the kind of personal epistemic relation that pertains to all and only mental individuals, minds, or agents and which relates the relevant individual to a fact or something that is the case. *S* knowing some fact, *f*, at a time, t_1, thus conveys an individual's epistemic relationship to *f* at t_1 as a relationship of knowing. When we take personal knowledge to describe someone's mental condition it is natural to think of cases where a mind knows some fact, *f*, as situations where there are some mental events or states in virtue of which the mind has this knowledge, or whereby the knowing subject partakes in a knowledge-relation to *f*. I may, for instance, have perceptual knowledge of the fact that my office wall is yellow, or remember how my late grandmother used to bake the most delicious honey cakes. It is natural to think that I know these things in virtue of processes and events that go on in my mind – or have gone on in my mind – which in turn are related to processes in my senses and memory that link me to the relevant facts or events. When we attribute knowledge to someone we therefore naturally associate '*S* knowing *P*' with attributing some mental state or another to *S*, just as other attributions, such as '*S* believes *P*' or '*S* desires *P*' would (Williamson 2000, p. 6).[2] According to Hossack, however, there is a fundamental disagreement in the metaphysics of knowledge between the majority of philosophers who take knowledge to be *composed* or *constituted* by the mental states or acts in virtue of which it obtains and those who, like Hossack himself, take the relationship to be *causal*.

In what follows, I shall argue that we have reasons from the metaphysics of knowledge to agree with Hossack on this issue since his framework allows us to preserve the unity of knowledge in the face of the vast plurality in ways of knowing. If you follow the majority and endorse the first view, you will commit to some version of the following:

> *Constitutive Thesis*: Knowledge is constituted by being in the right psychological or mental state in a favourable context.[3]

Following Hossack, I think we should reject the *Constitutive Thesis* and opt for the following *Causal Thesis* or the idea that knowledge stands to the subject's way of knowing – or its provenance of knowledge – as effect stands to its cause:

[2] Williamson, of course, takes this as *prima facie* evidence for the idea that knowledge is a mental state just like beliefs and desires. As we shall see (§3), I disagree with the idea that knowledge is or involves mental states. I shall say in what respect it is still right to think that knowledge-attributions do attribute mental states to the subject that knows when I discuss knowledge-reports in §4.
[3] Hossack (2007, p. 11).

> *Causal Thesis*: Knowledge is caused by being in the right psychological state in a favourable context.[4]

To render the truth of the *Causal Thesis* it will be crucial that the *Constitutive Thesis* is not a simple consequence of the idea that ascribing someone with knowledge is a way to describe their mental condition, or that it is entailed by the claim that knowledge obtains in virtue of mental processes. However, before we turn to this issue let me briefly say something about the background for this thesis and its role in contemporary epistemology.

The *Constitutive Thesis* is widely assumed in contemporary epistemology to be a sort of default metaphysics of knowledge. It usually serves as a starting point for analyzing or explaining the nature of knowledge in terms of something one takes to be conceptually prior or metaphysically more basic: namely, the psychological or mental states in virtue of which one knows something. The *Constitutive Thesis* may therefore suggests that we interpret such an 'in-virtue-of' relation as pointing to constitutive elements, components, or basic building blocks of the knowledge-relation. If a subject has knowledge of some *f* in virtue of having a certain epistemically qualified belief (or belief-like mental state) about *f* being true, it would suggest that knowledge is a complex or metaphysical hybrid that is to be analyzed, at least in part, as a qualified belief.[5] By combining the *Constitutive Thesis* with the claim that it is necessary to believe that-*A* in order to know it, philosophers have come up with what Hossack calls 'doxastic priority' theories in the metaphysics of knowledge where the aim is to analyze or explain knowledge as a peculiar kind of belief (2007, p. 11).[6] In turn, this has led philosophers to assume that we must explain knowledge as the conjunction of belief with whatever must in fact be added to this belief for it to constitute knowledge. The thesis may in this way lend support to a reductionist programme in the metaphysics of knowledge and the idea that knowledge is a metaphysical complex or hybrid crying out for analysis or reduction in terms of its basic building blocks.[7]

On the conceptual side we find a similar assumption in the idea that knowledge as a concept must be defined in terms of more basic concepts. This has led philosophers on a hunt for a non-trivial definition of knowledge that spells out necessary and sufficient conditions for it to be true that some subject, *S*, knows some *f*, where the right-hand side of the definition

[4] Hossack (2007, p. 11).
[5] I follow Williamson in treating the notion of a 'metaphysical hybrid' as characterizing views about the nature of knowledge that are entertained by reductionist programmes in the metaphysics of knowledge. On these views knowledge is accounted for as 'a mixture of mental states together with mind-independent conditions on the external world' and the aim of the reductionist programme is to break 'the mixture down into its elements' (Williamson 1995, p. 533). Also see, Williamson (2000, chs. 1, 2 and 3), O'Brien (2007, ch. 8).
[6] Also see Williamson (2000, *Introduction*).
[7] See Williamson (2000, *Introduction*, chs. 1, 2 and 3). Also see Hossack (2007, ch. 1).

does not involve the concept of knowledge or cognate terms. The *Tripartite Analysis*, or the claim that knowledge can be defined as a justified true belief, and its many descendants are all examples of this default strategy. Reliabilists follow suit by explaining knowledge as a reliable true belief; others have pursued what we might call a 'practical priority theory' trying to analyze knowledge in terms of practical abilities.[8]

Over the last few years these ideas have been criticized by several philosophers as resting on an unwarranted dogma. On the conceptual side philosophers have increasingly begun to question the assumption that philosophically important notions and concepts yield to such non-trivial analysis. As Millar says, 'there is an issue as to why we should even expect the concept of knowledge to yield to a traditional analysis' given that '[t]here seem to be relatively few concepts that do' (Millar 2007, p. 179). On pain of infinite regress we should not expect all important concepts to yield to such analysis and definition anyway. Moreover, the history of philosophy suggests that most concepts of philosophical interest are not so analysable. In this respect, 'bachelor' seems to be the exception rather than the rule, as Williamson says (2000, p. 31).[9] In particular, the last 50 years or so after Edmund Gettier (1963) refuted the Tripartite Analysis have seen attempt after attempt to define the concept of knowledge, where a succession of increasingly complex analyses and definitions have appeared only to be overturned by increasingly complex counterexamples, which more than suggests that knowledge is not so analysable.[10] A stubborn resistance to analysis is at least evidence for the claim that without the concept of knowledge and cognate terms we will not have sufficient conceptual resources for characterizing the set of necessary and sufficient conditions that is coextensive with the state of knowing.

There is also evidence from the last 30 years or so of research in cognitive science that supports the idea of the *analysandum* (viz. knowledge) as conceptually more basic than its alleged *analysans*. Child development data suggest, for instance, that the reverse order is in play in the growing child's acquisition of these concepts because children come to understand ignorance or the lack of knowledge long before they understand error and false beliefs.[11] This suggests that children acquire the concept of knowledge before the concept of belief. If

[8] Hyman defends something like the practical view of knowledge and suggests defining knowledge as 'the ability to do things, or refrain from doing things, or believe, or want, or doubt things, for reasons that are facts' (1999, p. 441). For a more recent defence of the idea that knowledge is justified true beliefs, see Haddock (2010). For reliabilist conceptions, see the classical expositions of Unger (1968), Dretske (1971; 1981), Armstrong (1973), Goldman (1975; 1986) and Nozick (1981). For a more recent defence, see Sherrilyn Roush (2005).

[9] Even Plato – whose philosophical interest to a large extent was determined by the Socratic pursuit for definition of what is in common between different instances of some philosophically interesting *X*, such as virtue, wisdom, or justice – seem to have found few definitions, if any, that could satisfy his philosophical criteria. Socrates appears to be happy only with the definition of 'clay' as 'earth mixed with moisture' in the *Theaetetus* (*Th* 147c2); but apart from this example and the rather triumphant definition of a 'sophist' at the end of Plato's *Sophist* (*S* 268d), satisfactory definitions are few and far between in the Platonic corpus.

[10] See Robert Shope (1983) for an overview of these debates as they have unfolded after Gettier's published his counterexamples.

[11] Josef Perner reports such data in his (1991, pp. 145–203).

concept-acquisition is anything to go by, then, knowledge/ignorance thus seems more conceptually basic than belief/error. In other words, these results could support the idea that the alleged *analysandum* (knowledge) is more basic than its suggested *analysans* (belief).

In more recent years cognitive scientists have also increasingly come to view our capacity for knowledge as quite fundamental and central to human cognition. Research in developmental cognitive psychology by Elisabeth Spelke and her associates have for instance provided evidence for the existence of a set of basic capacities which they call *core knowledge*. Core knowledge is a modular, domain-specific cognitive system geared to represent different sorts of entities and whose operation shapes and guides the life of an organism (not excluding the adult human being). These systems are primitive and perform central and indispensable roles in the cognitive system. The fact that these basic and central capacities develop long before there is evidence of a developed belief-psychology in children can thus be taken as further evidence of the conceptual priority of knowledge (Kinzler and Spelke 2007; Spelke 2000a; 2000b; 2004);[12] or, as Pascal Engel says:

> *One of the interesting facts provided by the huge literature on the acquisition of psychological concepts by children is that the concept of knowledge is acquired by children much before the concept of belief, and belief attributions seems to be supervenient upon the acquisition of the concept of knowledge. So naive psychology seems itself based on knowledge instead of belief.*[13]

A final point worth emphasizing is research on autistic subjects. Autism is a complicated phenomenon with different sub-categories ranging from the well-functioning (e.g. Asperger's Syndrome) to severe cognitive and social impairment. A common thread in much of the literature, however, is impairment in mind-reading capabilities and the attribution of mental states – and especially belief-states – to themselves and others. Research suggests, nevertheless, that autistic subjects with high verbal skills can reliably attribute knowledge/ignorance to themselves and others (Kazak et. al. 1997). The fact that they have trouble with belief-attributions but not knowledge-attributions is therefore further evidence that knowledge is conceptually more basic and psychologically more central than belief.

Perhaps the problem is the pursuit of a definition itself and that, as Williamson says, pursuing a definition is to participate in 'a degenerating research programme' (2000, p. 31).[14]

[12] Also see Susan Carey (1985; 2001) and Carey and Spelke (1996).
[13] Engel (*forthcoming*). For similar arguments, also see Engel (2003).
[14] We should bear in mind that this is a claim about the prospects of analyzing the concept of *knowledge* in particular. I need not defend giving up analysis or pursuit of definitions in general; however, if this were true in general, it would fit very well with the idea that knowledge is not analyzable. For support of the general demise of definition, see Fodor (1998).

After all, if it is false that knowledge yields to a non-trivial definition, pursuing the traditional project will, as John Cook Wilson remarked, itself be a source of errors:

> *Perhaps most fallacies in the theory of knowledge are reduced to the primary one of trying to explain the nature of knowing or apprehending. We cannot construct knowing – the act of apprehending – out of any elements.*[15]

Williamson therefore takes the history of unsuccessful definitions, the demise of the programme of definition and the data from cognitive science as inductive support for the working hypothesis that there is no definition of the concept of knowledge in terms of more basic concepts and that we should rather treat knowledge as a conceptual primitive.

My main focus will be on the metaphysical side of this issue rather than its conceptual counterpart. Even though it is plausible that the project of defining the concept of knowledge in more basic terms is stillborn, there may be something speaking in favour of a non-trivial account or metaphysical analysis of the knowledge-relation and for pursuing a reductionist programme in the metaphysics of knowledge. Of course, while failure to define the concept of knowledge will lend itself to scepticism with respect to such a programme, there might be independent metaphysical considerations supporting the idea that knowledge is a metaphysical hybrid constituted by more basic building blocks, considerations that our linguistic or conceptual primitives may fail to be responsive to. This may sound paradoxical, but we should, as Unger points out, always be prepared to entertain the possibility that there is 'something wrong with the language in which [these concepts] receive substance and expression' and that there might be an antiquated 'theory of things embodied in our language, inherited from an ancestor language, or languages' that represent the thinking 'of people a very long time ago' (1975, pp. 5–6).[16] Finding evidence for the idea that knowledge is conceptually basic and primitive may thus not translate directly into the metaphysics of knowledge. What the demise of the programme of definition brings out, however, is that we simply cannot assume that knowledge is a composite and a hybrid. Decisive metaphysical considerations – if there are any – must be brought to the table in order to settle the fate of the

[15] Wilson (1926, p. 803). As Charles Travis argues (2005), McDowell makes a similar move when he opposes what he calls 'the hybrid view', or the view that knowledge is decomposable into some condition for which we are completely responsible, and some surrounding conditions for which we are not (McDowell 1995a, p. 405).

[16] As Anscombe notes, Aristotle's 'philosophy has conferred more terms on ordinary language than any other' (1957, p. 88). If Aristotle's philosophy should turn out to be fundamentally wrong, that may arouse suspicion that 'all is not well with common sense', as Unger puts it (1975, p. 4), given that common sense would be largely determined by such antiquated concepts and terms. There are, of course, those who happily think that the notion we have of ourselves as knowers, agents, and active contributors to the world belongs to a dubious manifest image that should be discarded in favour of a more accurate scientific image. According to them, the notions we can find in the manifest image belong to the legacy of outdated metaphysics or the kind of systems which, in Thompson's words, must 'fall without remainder at the first breath of modernity' (2008, p. 13).

Constitutive Thesis and whether it supports the orthodox view that knowledge is a metaphysical hybrid or complex rather than a non-composite simple.

There are two ways to avoid the orthodox assumption that knowledge is a sort of belief: the first alternative is explored by Williamson who is in a position to accept the *Constitutive Thesis* because he thinks knowledge is a basic and *sui generis* mental or psychological state. According to Williamson, knowledge is a *factive stative attitude* (FSA), or a kind of propositional attitude state – as opposed to a process – that one has to all and only true propositions or facts. He therefore thinks there is a category or general kind of mental state, M_K, being in which is both necessary and sufficient for knowing some particular fact or proposition (2000, pp. 21, 37). This would render the *Constitutive Thesis* trivially true since it allows us to say that knowledge is constituted by this mental state of knowing (i.e. M_K):

> *Knowledge as Mental*: There is a state, M_K, such that M_K is a mental state and being in M_K is both necessary and sufficient for knowing some particular fact or proposition.
>
> *Trivial Constitutive Thesis*: Knowledge is constituted by being in M_K.

If defenders of the view that knowledge is a sort of belief cannot rule out the *Trivial Constitutive Thesis*, there is thus a reading of the *Constitutive Thesis* available that is compatible with the idea that the knowledge-relation is metaphysically simple as opposed to a complex or a hybrid. The orthodox view will therefore need to deny Williamson's framework in order for the truth of the *Constitutive Thesis* to support a reductionist programme in the metaphysics of knowledge. On the assumption that the *Constitutive Thesis* is true, Williamson can similarly take problems of its non-trivial versions as support for his own view.

An alternative way to avoid the orthodox view is to deny the truth of the *Constitutive Thesis* – both its trivial and non-trivial versions – which is what Hossack suggests. When we say that knowledge describes the cognitive state of an agent and that one knows some *f* in virtue of undergoing certain mental events, this need not entail that the 'in-virtue-of' relation holds true because there is a constitutive relation between knowing and those psychological states. An alternative interpretation is rather that the 'in-virtue-of' relation holds because there is a causal relationship between one's psychological states and the knowledge-relation. In other words, the following *Causal Thesis* is an alternative to the *Constitutive Thesis*:

> *Causal Thesis*: Knowledge is caused by being in the right psychological state in a favourable context.[17]

What these two alternatives to the orthodox views have in common is that they would allow us to treat knowledge as metaphysically simple and basic. They would thus allow us to reject the orthodox assumption that knowledge is a metaphysically analysable relation or metaphysical hybrid constituted by more basic building blocks and crying out for analysis and explanation. Where they differ, however, is in what they take the nature of knowledge to be. Williamson still thinks that knowledge is a kind of mental state: namely, the mental state, *M*, which is a factive stative attitude (FSA) towards some proposition that is necessary and sufficient for knowing. Whereas Hossack, as we shall see below, entertains the idea that knowledge is basic mental relation that relates minds with facts without *any* intermediaries such as mental states, propositions, contents, and so on.

There are therefore three main models on the table: (1) the non-trivial *Constitutive Thesis*; (2) the trivial *Constitutive Thesis*; and (3) the *Causal Thesis*. Let me now spend the rest of this section arguing against the first non-trivialist option, before turning to settle the dispute between options (2) and (3) below (§3).

A METAPHYSICAL ANALYSIS OF KNOWLEDGE would of course be appropriate if knowledge really were a complex composite or metaphysical hybrid that is constituted by more basic building blocks and components. However, this depends on there being a non-circular account of what it is to know something in terms of its alleged constitutive components. If knowledge is a kind of belief, for instance, a metaphysical explanation would require a story of what it is to know something that, at least partly, is given in terms of what it is to believe something. What both Williamson and Hossack deny is that we can have a story of knowledge as a complex combination of favourable circumstances and having psychological states other than knowledge itself. Beyond the failure to non-trivially define the concept of knowledge, there are reasons for thinking that the metaphysics of knowledge will not support the truth of a non-trivial *Constitutive Thesis*. For one thing, the fact that there are many ways of knowing – and thus apparently a plethora of different psychological states in virtue of which one may know things – would threaten the unity of such a theory. A sudden flash may reveal the truth of a mathematical theorem; we can see that the ball is red, or that little Todd is behaving in his best manner; we remember that Granny could bake the very best honey cakes; we hear and

[17] Hossack (2007, p. 11).

experience that Mahler's 9th symphony is an expression of sublime beauty; and we may even infer from this experience that Mahler was a genius. If knowledge can be constituted by 'any of such a miscellany of states', knowledge appears as a disunified relation rather than as one thing (Hossack 2007, p. 12). I shall call this the *Problem of Disunity*. As we shall see, it will show up in different versions along the way and in the end I will take it that the *Causal Thesis* recommends itself as the unique best way to retain the unity of knowledge in the face of a plethora of disunity in ways of knowing.

A way to recover the unity of knowledge for a non-trivialist – viz. a proponent of strategy (1) above – is if one could find a shared psychological component which all these instances had in common. One could then suggest that the knowledge-relation is constituted, at least in part, by this shared component, thereby retaining the unity of knowledge in the midst of disunity of ways of knowing. A candidate component is, of course, a belief. If beliefs must always be present in instances of knowing, the many different ways of knowing will all involve a knowledgeable belief. Knowledge could then be said to be a kind of belief. Non-trivialists would therefore avoid the problem of disunity if the following were true:

B: If *S* knows that *A*, then *S* believes that *A*.

As Williamson points out, the truth of *B* is crucial for someone who takes knowledge to be analysed as or constituted by a sort of belief (2000, p. 42). However, he goes on to say, it clearly does not sound '*trivially* valid' in the way that the truth of 'If *S* knows *P*, then *P* (is true)' appears to be.[18] Even if we accept that there is a corresponding belief for the different ways of knowing that I mentioned above – viz. the flash of mathematical discovery, the sweet memories of Granny and the musical experiences – it is a problem for this thesis that there appear to be cases where subjects know something even in the absence of relevant beliefs. We may, for instance, judge that *A* is true without believing it to be true. It then seems plausible that we may judge that *A* and thereby come to know it without this knowledge being accompanied by a relevant belief (Hossack 2007, p. 12). To see this we should first recognize that judgements and beliefs are different ontological categories. Making a judgement is probably best rendered as a mental action – or a *cognitive mental act* – which is the act of putting a proposition forward in one's mind as true. That operation is an act or a process rather than a state because it is an event and something we do for a reason. It is cognitive

[18] Most philosophers appear to accept the factivity condition on knowledge; however, there are also some exceptions. Nancy Cartwright (1980), for instance, argues that the laws of physics cannot be formulated without loss of explanatory power as something that could possibly correspond with the truths. Yet, these laws express our best scientific knowledge. Also see Helen Longino (2002, chs. 4 and 5).

because this event or action is to present something as being true. A belief, on the other hand, is a mental state that represents something as true, or what we may follow Shah and Velleman in characterizing as a *cognitive attitude* (2005, p. 503). Belief-states may normally occur as a result of judging; however, in a given case judging may fail to lead to the formation of the relevant belief. As Peacocke says, '[s]omeone can make a judgement, and for good reasons, but it [need] not have the effects that judgements normally do – in particular, it may not result in a stored belief' (1998, p. 90). Recall the case we discussed earlier:

> *Someone may judge that undergraduate degrees from countries other than their own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all.*[19]

As Cassam points out in his discussion of this case, belief-formation may be influenced by other forces, say non-rational forces such as 'self-deception, prejudice and phobias' (2010, p. 82).[20] An unshakeable prejudice may simply hinder me from taking my own reasonable judgements on this matter to heart, as it were. As Cassam says, I may 'mentally affirm that undergraduate degrees from countries other than my own are of an equal standard to my own and yet my attitude towards this proposition is not the attitude of belief, as evidenced by my hiring decisions and letters of recommendation' (2010, p. 82). But even if one's beliefs fail to track one's judgements in these cases, there is no need to think that one thereby fails to know what one reasonably judges to be true. This deviance between judgements and beliefs thus opens a way for knowing without believing.

Similarly, an unconfident examinee may take herself to merely be making conjectures and guesses on a range of questions in an examination in, say, English history. She does not really believe the things she says are true; rather, she sincerely takes herself to be ignorant about English history. She is merely making a shot in the dark, as it were, in order to make the best out of a truly uncomfortable and perhaps a little desperate situation. In a word, she would be less of a believer and more like Pascal's famous wager, who was unable to believe in the existence of God but who could nevertheless make a reasonable bet and act as if he did believe in the existence of the Almighty.[21] However, suppose that the unconfident examinee reliably produces the correct dates and facts as a result of forgotten history lessons. Would it

[19] Peacocke (1998, p. 90). I discussed this case in connection with Paul's Inferential Theory. See chapter 5, §4.
[20] Shah and Velleman discuss similar cases (2005, p. 508).
[21] See Blaise Pascal (1670).

not then seem as if she knew those things after all? That is, if she can answer that the Battle of Hastings was in 1066, or that King Charles I was executed in 1649, would it not put pressure on us to say that she knew these things without believing them? After all, these answers do not appear to be the result of 'pure fluke', as Colin Radford says (1988, p. 498), so it would be extremely odd to flat-out deny that the examinee has knowledge in this case; yet, it appears odd to insist that she therefore must also believe these things.[22] It thus seems as if we can sometimes pick up, know and act upon information without believing or remembering it. You may, for instance, watch the evening news where they tell you that your commuter train will be out for a week. Next day you may nevertheless find yourself on your way to the train station. If you are like me in the morning, you may do this in a mindless, habitual manner. You will soon enough remember that your commuter train is out of service, though, but for the nonce, as you are absent-mindedly walking to the train station, it appears as if you know that the commuter train is out of service without this being a thought or belief of yours. That is, you appear to have knowledge that is not accompanied by a relevant belief. After all, when you remember, you will reproach yourself saying things like 'Strewth! That was stupid! I knew it all along!' and not 'I believed it all along'.[23]

In reply to such cases, defenders of the orthodox view have replied that subjects such as the unconfident examinee not only fails to believe that the Battle of Hastings was in 1066, but that she also fails to have knowledge. The reason is, they argue, that the subject fails to have justification (or evidence) for this fact and justification is arguably another necessary condition for knowing something.[24] It would follow that these cases are not counterexamples to *B* and thus that beliefs may still be a candidate unifying psychological state to substantiate a non-trivial *Constitutive Thesis*. The problem with this reply, however, is that it is possible for opponents of the non-trivial *Constitutive Thesis* to follow Williamson and take one's body of evidence or justifiers to include one's body of knowledge (2000, ch. 9). That knowledge serves a central role in justifying one's beliefs is exactly the kind of move that is available to views that departs from the orthodox view since by treating knowledge as a basic relation one may use this notion to define, explain, or illuminate other notions in epistemology such as justification and epistemic warrant. To use knowledge in this role may, for one thing, explain the traditional problem of the regress of justification since it can offer an answer to the

[22] See Radford (1966; 1988). Also see Mathias Steup (2006) and Williamson (2000, p. 42).
[23] Williamson discusses a similar case where one is simply incapable of having the belief that it rains even though one is clearly having a factive stative attitude (FSA) towards this fact. Suppose for instance that one is seeing that it is raining. Since seeing *P* is a factive stative attitude – and thus involves knowledge, according to Williamson, since knowledge is the most general FSA – he concludes that such cases put pressure on 'the link between knowing and believing' (2000, p. 38). For a different conclusion, see Cassam (2007, pp. 28–9).
[24] See Steup (2006).

question of what justifies the justifiers (or one's body of evidence). However, if one's knowledge is what justifies, we are basically operating with a 'very modest kind of foundationalism' on which all one's knowledge serves as a foundation for the propositions that are justified on one's body of evidence (Williamson 2000, p. 186). There will then be a sense in which knowing that the Battle of Hastings was in 1066 serves as its own evidence or justification. The justification condition is thus satisfied insofar as one knows this fact. To dispute the unconfident examinee's justification is thus be to dispute whether she knows this proposition. In a word, the defenders of the orthodox view will have begged the question.

Two other points are in favour of the idea that knowledge can be had without belief. On the one hand, it appears that perception may yield knowledge also in simple creatures that are not capable of having complex representational mental states such as beliefs. On the other hand, it also appears as if creatures can act intentionally – and thus knowingly – without the relevant conceptual resources for representing what they are doing. Let me deal with the case of perception first before turning to the case of acting intentionally. It appears as if certain creatures may be able to know things by perceptually noticing and representing entities, events and states of affairs in their environment – and thereby be aware of them and know them – despite the fact that their representational states seem to be non-conceptual. That means that they are capable of representing some states of affairs, *f*, despite not possessing the concepts required to specify the relevant content of these representational states. By representing the world non-conceptually the creature has experiences whose impressions are made by the world on the creature's senses – or are a product of its sensitivity – without thereby implicating that 'those impressions themselves already have conceptual content', to use a phrase of McDowell's (1994, p. 46). If this is possible, it seems natural to think that these creatures also have knowledge resulting from non-conceptual representations. Again, it thus appears that knowledge can obtain in the absence of beliefs.

Moreover, this might also carry over to human perception, which seems to involve perceptual states whose representations involve a richness, texture and fineness of grain that beliefs do not and cannot have because they are locked on to conceptual representations. Our experiences are, as Peacocke argues, not 'restricted in [their] range of possible contents to those points or ranges picked out by concepts – *red, square, straight ahead* – possessed by the perceiver' (1992, p. 68); at least, no more than the experiences of certain animals are so restricted. There may thus be non-conceptual components also to human perception that are nevertheless representational and which allow the perceiver to represent things that go beyond

her conceptual resources. In turn, it is plausible to think that such perceptual states give rise to a sort of immediate and detailed kind of knowledge of the environment that is non-conceptual. Since that would allow the perceiver to know something that goes beyond her conceptual resources it follows that one can have knowledge without corresponding beliefs since beliefs do have to represent their objects conceptually. Thus, we seem to have states yielding knowledge of what we do not – or cannot – have beliefs about.[25]

Finally, we find additional support for the idea that knowledge can come about without belief if we briefly turn our attention to actions and intentional agency. As Susan Hurley has argued, intentional agency seems to make 'normative space between a mere stimulus-response system and conceptual abilities' (2001, pp. 425–6). When some agent takes some *f* as her reason for acting only a 'minimal kind of recombinant structure' is required, Hurley goes on to say, to allow the agent to combine an intention to so act with different perceptions, and the given ends with different means. Given that it still seems to hold that such creatures must, as Aristotle insisted, know or be aware of 'the particular circumstances of the action' – such as what it is doing and for what purpose – such cases would seem to involve a kind of non-conceptual knowledge too (viz. whose objects of knowledge need not be conceptually available to the knower or agent). Thus, an agent may be Φ-ing with some end, *e*, in view without thereby being able to conceptually represent her action as a Φ-ing with a view to *e*. Non-conceptual representation – and a corresponding non-conceptual way of knowing what one is doing and why – may be all that is required. In action, we may thus employ non-conceptual knowledge to provide us with reasons for acting that may guide and steer our movements without requiring the corresponding conceptual resources for representing our reasons. In turn, we may also envisage agents doing something intentionally despite lacking conceptual resources for representing and describing what one is thereby doing. Again, it thus appears as if one may know some *P* – this time, knowing one's reasons and action – without being required to believe *P* since so believing would in turn also require the creature to be able to conceptually represent *P*.

[25] See Hossack (2007, pp. 12–3). Also see José Luis Bermúdez (1998, pp. 50–8). John McDowell has argued that one can have a sort of conceptual view of such perceptual states if the conceptual content can be characterized by the use of a demonstrative and a general pronoun (McDowell 1994, lecture 3). To take an example: perceivers typically do not have a rich range of concepts for different shades of colour, such as maroon, lilac and so on. However, one may nevertheless be able to recognize the particular shade of a maroon object by conceptually representing it as *that shade*. The disadvantage of this model is that it would have to portray a case where someone's behaviour is explained by her confusing shades of colours as the misapplication of the content *that shade*. However, the subject may mistake a lilac object for an object with maroon shade, in one case, while mixing a maroon object with a mauve object in another case. For McDowell, these cases are both explained as misapplication of the content *that shade* as it pertains to the shade maroon. If we invoke non-conceptual content, however, we are allowed to invoke and distinguish the perceiver's representations of mauve and lilac shades without the perceiver possessing these concepts. That model can thus explain the two cases of misapplication as different mistakes (Bermúdez 1998, pp. 54–7). Hossack also objects to McDowell's model and argues that one does not have knowledge of how some shade looks by believing it looks like *that*; rather, one is in position to believe that it looks like *that* because one already has knowledge of how it looks (Hossack 2007, p. 13).

The problem with these cases for a non-trivial reading of the *Constitutive Thesis* – viz. one that supports something like the orthodox idea that knowledge is a kind of belief – is that these knowledge-relations obtain without the subject having a corresponding belief. What they indicate is therefore that there are many different 'verdictive' mental states that give rise to knowledge other than belief-states. This will throw the non-trivial *Constitutive Thesis* back to the Problem of Disunity. These are serious blows to the orthodox view and the non-trivial reading of the *Constitutive Thesis*. In addition to the non-analyzability of knowledge, they therefore count against the idea that knowledge is a sort of belief or metaphysical hybrid constituted by psychological states as its more basic metaphysical building blocks.

§3 AGAINST WILLIAMSON

Hossack takes the above considerations to count in favour of the *Causal Thesis* (2007, p. 13); however, he thereby fails to acknowledge the kind of trivial version of the *Constitutive Thesis* that Williamson may endorse by treating knowledge as a *sui generis* mental state, M_K. On Williamson's account, knowledge is accounted for as the most general of the factive stative attitudes (FSAs), where an FSA is defined as a state attitude one necessarily has only to facts or truths. Examples are states such as seeing *P*, remembering *P*, or knowing *P*. If knowledge is the most general FSA, it means that insofar as one bears an FSA to some fact, *f*, one must also know *f* (2000, p. 34). For Williamson it is therefore possible to explain the unity of the above cases by saying that they all involve FSAs and thus, since knowledge is the most general FSA, that they all involve the mental state of knowing, M_K. M_K could then serve as the unifying psychological state for all these ways of knowing. We must therefore find a way to rule out Williamson's framework to recommend the *Causal Thesis*. Since I used the *Causal Thesis* to support a view about practical knowledge in the previous chapter, I shall therefore spend this section vindicating the *Causal Thesis* by rejecting Williamson's view.

I think it is safe to say that Williamson's account has met most resistance on the very idea that knowledge is a mental state, or the idea that knowing is constituted by a mental state, M_K, the occurrence of which is both necessary and sufficient for knowing some particular fact or proposition. Not that this view is obviously false. Aristotle, for instance, seems to accept that 'the intellectual states by which we grasp truth' come in two different kinds since 'some are always true and some admit falsehood' (*PA* 100b6). Moreover, as Williamson points out, it is also a 'natural assumption' that sentences of the form '*S* knows *p*' is attributing some subject, *S*, with mental states just as much as a sentence of the form '*S* believes *p*' or '*S*

desires *p*' (2000, p. 6). But although one attributes some mental state or another by attributing *S* with knowledge,[26] it is a further question whether the mental state one thereby attributes could itself be a state recognizable as the mental state of knowing. For one thing, such a mental state seems *prima facie* to run counter to certain features we take to be constraints on candidate mental states. Knowledge is, after all, factive and thus entails the truth of the object of knowledge. To bear the alleged mental state of knowing to some *P* – that is, to M_K some *P* – would entail the truth of *P*. Such a factive mental state will thus typically not supervene narrowly on the internal characteristics of its subject and will need more than just what goes on within the skull (or under the skin) of the knowing subject (Williamson 2000, p. 49).[27] Rather, the constitution of such a state will depend on parts of the world that are external to an agent's intrinsic features – viz. the features of the world on which the truth of the object of knowledge depends. In a word, such a state will be what we may call an *externalist* state as opposed to an *internalist* state of the subject.

A problem is therefore that since knowledge depends constitutively on features of the environment external to the subject it indicates that knowledge – whatever it is – is not a purely mental phenomenon or state. A related worry is that knowledge as a purely mental state would not be transparent to its bearer since the external dependence of knowledge entails that one will not always be in a position to know whether one knows something. One may know *P* and thereby be in the mental state, M_K, that constitutes knowing something, without knowing that one is in this state. Finally, one may object that since this state depends on – and thus in a sense is – more than the agent's internal characteristics, it cannot contribute causally to the agent's behaviour since causal relations must be local and due to casual powers that must be intrinsic (viz. supervene on what goes on inside the subject's head). Knowledge as a mental state will therefore fail to occur in the psychological explanation of actions in the way that other mental states do (that is, in the way states that supposedly supervene all and only on the internal characteristics of its subject may causally influence the subject's behaviour). On the contrary, only the elements of the knowledge-relation that do depend on the subject's internal features can seem to partake in psychological explanations. These worries thus count in favour of a view of knowledge as factorizing into its proper psychological constituents and some additional factors that are its non-mental elements.[28]

[26] This is a point that, among others, Tyler Burge insists on (1979, p. 599).
[27] The exception is if one knows something whose truth depends all and only on one's internal characteristics, such as for instance knowledge of one's own thoughts or mind.
[28] See Williamson (2000, ch. 1) for discussion of these *prima facie* objections. Also see Asbjørn Steglich-Petersen (2005, p. 16).

Against these pre-theoretic expectations, the claim that knowledge is a mental state is a surprising philosophical thesis. At this point, common sense may seem to support the orthodox view and to favour the idea that knowledge has some mental component or another in addition to other conditions that are non-psychological. In a word, our pre-theoretic conception could be taken as support for the claim that knowledge is a metaphysical hybrid and a complex of mental and non-mental states. Indeed, many philosophers have pursued something like the following argument. They start off from the idea that what warrants the operation of some state, M_K, in psychological explanations – viz. an explanation that cites mental states and acts of an agent to explain her behaviour – is that it makes a distinctive contribution to causal generalizations. They then argue that causal generalizations apply to individuals in virtue of the individual's causal powers and that these powers must supervene on intrinsic features of the individual. It follows that the properties and states that one may cite in psychological explanations must be intrinsic too; and this, in turn, will rule out knowledge as a mental state since knowledge is not intrinsic. So at most, only some of the constituents of this relation can figure in psychological explanations.[29]

A similar argument could, perhaps, be entertained by starting from considerations of transparency. We may fail to know that we know and thus fail to know that we are in the state, M_K, that allegedly constitutes knowing something. However, someone may argue that one cannot fail to know that they are in states, M_1, M_2,..., M_n, if these are genuine mental or psychological states. That is, one may treat the following as a universal truth:

> *Transparency*: For any mental state, *M*, and subject, *S*, whenever *S* is suitably alert and conceptually sophisticated, *S* is in a position to know whether she is in a state *M*.

Again, it would appear that there is at most some constituents of the knowledge-relation that are transparent in this way while the rest are non-transparent factors. We may thus perhaps conclude that the first correspond to the psychological states involved in knowing something whereas the other constituents of knowledge – viz. the ones that make knowing a non-transparent state – must be non-mental. If sound, such arguments will count in favour of the conclusion that, as Fodor puts it, there 'can't be a psychology of knowledge' (1981, p. 228).

The resistance to the idea that knowledge is a mental state is thus part of the fault line that goes between externalist and internalist conceptions of the mind. The internalism I have

[29] See Fodor (1981, p. 228). For Fodor, this consequence is given by his *formality condition*, according to which mental states and processes defined over representations apply in virtue of nothing but the syntax of the representations. Since knowledge involves truth and truth is a semantic notion – viz. irreducible to mere syntactic operations – it follows that knowledge cannot be a mental state or process.

in mind here is a position about the mental states or attitudes themselves (mental state internalism) rather than their contents (viz. content internalism). Internalists about mental attitudes hold that the nature of some psychological state, *M*, constitutively depends all and only on the total internal physical state of the agent, *S*, at some relevant time, t_1. Externalism about mental attitudes, on the other hand, is the denial of internalism and thus allows conceptual space for at least some state, M_{EX}, which is such that M_{EX} is a psychological state and constitutively depends on features of the total physical state of the environment external to *S* (viz. on features other than *S*'s total internal physical state). In this respect, internalists will restrict psychological states and properties to all and only *narrow* conditions of *S* – which are conditions that *S* will have over a range of different cases insofar as the cases are internally alike although their external total physical state may change – whereas externalists will allow at least some psychological states and properties to be *broad* (viz. not narrow) conditions. Knowledge is, of course, a notoriously broad condition since the relation depends on non-intrinsic features regardless of what the object of knowledge is, or whether the content representing the object known is externally individuated or not.[30] Armed with mental state internalism one would therefore typically treat a broad state or condition like knowledge as factorizing into its mental (viz. narrow) and non-mental (viz. broad) components. Again, the rationale behind such factorization may be that only the narrow components of knowing are transparent and thus mental; or one may push the idea that, as Williamson puts it, insofar as 'a causal explanation of action cites a broad mental condition [e.g. knowledge], an underlying narrow condition must do the real [causal] work' (2000, p. 65). A candidate mental component is, of course, the belief-state. Internalism about mental states may therefore serve as a crucial lemma towards the conclusion that, as Stephen Stich puts it, 'what knowledge adds to belief is psychologically irrelevant' (1978, p. 574); and in turn this conclusion will lend its weight to the reductionist programme in the metaphysics of knowledge.

In reply, Williamson argues that transparency conditions that knowledge fails to satisfy are also conditions that paradigmatic mental states fail to satisfy. He also argues that even the most uncontentious mental states, such as beliefs and desires, are broad since the

[30] As Steglich-Petersen points out (2005, p. 17), the broadness of an attitude may derive both from its attitude-component and its content-component. Some mental attitude, Φ, will be broad due to its content-component if Φ pertains to content that cannot be individuated without reference to features of the external environment of the subject. Hilary Putnam, for instance, has in a series of papers argued that the content of propositional attitudes constitutively depends on external features. In particular, Putnam argued, the content picked out by natural language terms, such as 'water', is externally individuated. In turn, similar arguments could be used to support widespread externalism about content and that contents in general must be individuated with reference to features of the subject's external environment (1973; 1975). '"[M]eanings" just ain't in the head!' as his slogan goes (1975, p. 587). Later he argued that externalism about content could be used to reject certain arguments for skepticism about knowledge (1981a). For similar arguments in favour of externalism about content, see Tyler Burge (1979) and Gregory McCulloch (1995, pp. 162–3). In the case of knowledge, however, the broadness derives from the attitude-component with the result that this attitude will be broad even if its content could be individuated all and only in terms of the intrinsic features of its subject. Knowledge and other factive attitudes will therefore be broad even if externalism about content should turn out to be false.

contents of these attitudes must be broad. From this it follows that there is no reason to rule out knowledge as a candidate mental state. On the contrary, there is reason, he argues, to invoke this candidate mental state and make it a part of our psychological repertoire since there are cases whose unique best explanation can only be had by attributing the agent with knowledge. The internalist's reply to these arguments, however, might be to argue that one can factorize the broad contents that our beliefs and desires relate to into their narrow and broad components, thereby separating the internal and external features of these states. If *S* believes some *p* and the content of *p* must be externally individuated, the internalist may factorize *p* into a proper mental representation or narrow condition, Ω, that has the content that *p* as its broad condition. The idea is to let Ω figure in psychological explanations given that Ω depends all and only on internal features of the subject, *S*, even if its content, *p*, depends on features beyond *S*'s intrinsic features.[31] If this strategy works, we can similarly factorize knowledge into its psychological and non-psychological components, and then in turn factorize the broad contents of these psychological states. The result is that knowledge may figure as a term in causal-psychological explanations; however, it is a belief and some narrow condition, Ω, that makes the explanation true and does the real causal work.

To combat this internalist reply, Williamson argues that efforts to factorize beliefs into their respective internal (narrow) and external (broad) components are bound to fail because beliefs, like knowledge, are prime conditions, or non-composite states. A condition, *C*, is *composite* if and only if it obtains whenever the conjunction of some narrow condition, *N*, and environmental or broad condition, *E*, obtains; whereas *C* would be *prime* if and only if it is not so composite.[32] Williamson's argument aims to show that beliefs are prime states because we could derive a contradiction in a range of plausible cases if we assumed that beliefs are composite of narrow conditions and broad content-conditions. Suppose for the purpose of deriving a *reduction ad absurdum* that (a) the belief that tigers growl, *Bp*, is a composite state, such that *Bp* can be factorised into a narrow condition, N_c, and a broad content-condition, B_c. Suppose further, that there are three cases, α, β and γ, and that (b) *Bp* obtains in α and β; (c) α and γ are internally alike (such that all narrow conditions in α obtain in γ and *vice versa*); (d) β and γ are externally alike; and (e) that *Bp* fails to obtain in γ. It then follows from (b) and (c) that N_c obtains in γ, and from (b) and (d) it follows that B_c obtains in γ, with the result that both N_c and B_c obtain in γ. Since, however, *Bp* is the composite of N_c and B_c it will follow via

[31] This is in essence how Fodor (1981) and Stich argue (1978, p. 574). Also see Steglich-Petersen (2005, p. 18).
[32] Note that a narrow mental state, *M*, will then be trivially composite since it will be the conjunction of itself with the environmental condition, *E**, that holds in all cases whatsoever (Williamson 2000, p. 66).

(a) *but* contrary to our assumption (e) that *Bp* also obtains in γ. In a word, we have reached a contradiction. A plausible combination of cases α, β and γ satisfying (b) through (e) will thus be inconsistent with the idea that *Bp* is a composite state (a).

Williamson demonstrates several such triad cases in order to show that certain states are prime rather than composite.[33] If we assume that the belief that tigers growl, *Bp*, is composite and that in addition to tigers there exist schmigers – which look just like tigers – the assumption that *Bp* is composite is inconsistent with the following case-descriptions:

α	External:	Tigers inhabit the mountains, schmigers the jungle.
	Internal:	*S* remembers encounter with tigers in the mountains but forgets encounter with schmigers in the jungle; *S* therefore believes that tigers growl but does not believe that schmigers growl.
β	External:	Schmigers inhabit mountains, tigers the jungle.
	Internal:	*S* remembers encounter with tigers in jungle but forgets encounter with schmigers in mountains; *S* therefore believes that tigers growl but does not believe that schmigers growls.
γ	External:	Schmigers inhabit mountains, tigers the jungle.
	Internal:	*S* remembers encounter with schmigers in mountains but forgets encounter with tigers in jungle; *S* believes therefore that schmigers growl and does not believe that tigers growl.

The internal conditions of α and γ are here the same while external conditions of β and γ are the same. The alleged narrow condition, N_c, of the belief that tigers growl, *Bp*, thus obtains both in α and γ; however, *Bp*'s broad content-condition, B_c, as it obtains in β should also obtain γ. If *Bp* really was a composite state, it should therefore follow that *Bp* obtains also in γ. Since *Bp* cannot obtain in γ it follows by *reductio ad absurdum* that *Bp* is a prime state and not a composite of narrow components and broad content-components.[34]

According to Williamson, a sound argument from primeness will thus establish that belief-states and knowledge are on a par as candidate mental states. If both are prime states, a belief will, after all, not lend itself to factorisation anymore than knowledge does. Now, as Frank Jackson says, '[h]ypotheses about what people believe and desire greatly facilitate the prediction of their behaviour' (1996, p. 377). And this is often taken to support the claim that these really are mental states. If the argument from primeness works, the claim that knowledge really is a mental state will therefore depend on whether attributing such a state is

[33] Williamson (2000, pp. 69–72). An assumption behind these triads is the principle of free recombination, or the idea that the external and internal components are independent variables. For a defence of this assumption, see Williamson (2000, pp. 73–5).
[34] For similar arguments, see Williamson (1995; 2006). Also, see Steglich-Petersen (2005, p. 19n7).

needed to yield the unique best explanation of someone's behaviour by facilitating predictive gains. As Steglich-Petersen has pointed out, however, a problem with Williamson's argument from primeness is that it fails to establish that the sources of broadness in beliefs are on a par with the sources of broadness for knowledge. The primeness argument concerns only the broadness that arises from the individuation of content. If it holds, it shows that this kind of broadness precludes the factorisation of beliefs because belief is a prime state. However, we should not forget that '[k]nowing that *p* involves features about the environment in a way that believing *p*, even if prime, clearly does not: believing *p* does not involve the truth of *p*' (Steglich-Petersen 2005, p. 20). That belief is, and has to be, a broad state that cannot be decomposed into its component due to its primeness is one thing; that knowledge in addition is broad in the sense that depends on the truth of its content is quite another. This leaves room for an opponent of 'the psychology of knowledge' to make a distinction and argue that truth-involving prime states are at an explanatory disadvantage in comparison to prime states that are not truth-involving. A related point is that by showing belief to be a prime state one only removes one reason – viz. that belief lends itself to successful factorisation – for thinking that knowledge is a hybrid consisting of a mental and some non-mental components. There might still be other reasons for thinking that knowledge is a hybrid; for instance, because knowledge is (notoriously) broad due to its attitude-component.

Finally, even if we were to grant Williamson's conclusion that knowledge is just as acceptable a candidate mental state as belief, and that both are on an equal footing, as it were, with regard to figuring in psychological explanations, it does not follow that knowledge is acceptable. All that follows is that knowledge is no worse off than belief in this respect. And being on an equal footing in regard to explanatory power is not being on a good footing in this respect, anyway. The claim that broad states are can be utilized in psychological explanation is not homed in by this conclusion unless we take it for granted that beliefs must figure in such explanation regardless of them being prime and broad. All Williamson's conclusion may show is that the explanatory power of belief is just as bad as the explanatory power of knowledge and that neither belief nor knowledge therefore can figure in a psychological explanation (Steglich-Petersen 2005, p. 20). There is therefore reason to be sceptical to Williamson's conclusion that knowledge is a candidate mental state.

The other half of Williamson's claim is that not only is knowledge a candidate mental state but that we also need to invoke this state in causal-psychological explanations since the fact that some agent, *A*, knows some *p* may be the unique best explanation for what he does.

To support this claim Williamson points to certain cases where it seems as if knowledge is the state yielding the highest predictive probability for *A*'s Φing in this situation:

> *A burglar spends all night ransacking a house, risking discovery by staying too long. We ask what features of the situation when he entered the house led to that result. A reasonable answer is that he knew that there was a diamond in the house. To say just that he believe truly that there was a diamond in the house would be to give a worse explanation, one whose explanans and explanandum are less closely connected*.[35]

The argument is that the best explanation of the burglar's behaviour is one that cites his knowledge as opposed to him merely having a true belief or other potential substitutes. What the argument takes advantage of is the fact that knowledge is a relation that has certain features that other combination of states will lack unless the combination provides the necessary and sufficient conditions for knowledge. In the absence of an analysis of knowledge Williamson's strategy will therefore work against potential substitutes. Williamson himself takes the feature in the burglar to be the incompatibility of knowledge with misleading evidence that the diamond is not in the house. If the burglar knows it is in the house, he cannot base his confidence on misleading evidence. However, a mere true belief might be so based. A mere true believer might therefore discover misleading evidence and, as a consequence, leave the house; whereas this possibility is ruled out on an explanation citing knowledge. The basic point, however, does not rely on the incompatibility between knowledge and misleading evidence, but on the idea that knowledge is more persistent than merely true belief because the subject relates to a truth which holds in a greater range of counterfactual circumstances in the case of knowledge in contrast to true beliefs that fall short of being known.[36] The problem is that, even if we grant that knowledge is more persistent and thus adds to the probability of an agent's behaviour in a way that is unique to the state of knowledge, it does not follow that there is a causal relationship between knowledge and the burglar's behaviour. The reason is that a condition may be essential for the causal explanation in a given case without itself being a cause or a causal condition.

John Heil points out that '[w]hen a term figures in a true causal explanation, the question remains: what is it about the world that makes the explanation true?' (2004, p. 233n12) For us, the question is even more poignant: even if some condition, *C*, figures as an essential part of a true causal explanation, it remains to find out what the role of *C* is. It may

[35] Williamson (2000, p. 62).
[36] This is incidentally also what Socrates is driving at in the *Meno* dialogue when he argues that 'knowledge differs from correct opinion in being tied down' (Plato *M* 98c). See Haddock (2010, p. 207) for a similar formulation.

for instance be the case that *C* is a condition that increases the probability of some other condition, *C**, being causally efficacious. Steglich-Petersen argues that knowledge thus relates to the agent's Φ-ing in a way that is structurally analogous to the way dry conditions relate to a fire. In the case of a fire, there is presumably a spark or something that is its causal condition whereas the dryness condition functions as a condition that dramatically increases the probability of the spark being causally efficacious. Similarly, the psychological events and processes in virtue of which the burglar knows that the diamond is in the house may be the causal condition for his behaviour, whereas the fact that he knows functions as a condition that increases the probability of these psychological events and processes being causally efficacious. An internalist may thus allow knowledge – as well as other broad conditions – to figure as essential elements of causal-psychological explanations as conditions facilitating the probability of the causal condition without invoking knowledge as a cause.

A final problem with Williamson's account becomes apparent when we look at the way he gives 'substance to the category' of factive attitudes. Williamson, as I said, takes knowledge to be the most general kind since knowledge accompanies any state, *F*, as long as *F* is a factive attitude. Williamson explicitly treats knowledge as a relation between a mind and a true proposition because he entertains what we may call a redundancy theory of facts that reduces facts to the class of true propositions (2000, pp. 34, 42–3). On this issue Williamson is on the same side as Davidson, who thought that '[t]alk about facts reduces to predication of truth' (1969b, p. 43), and with Peter Strawson who expressed similar opinions when he claimed that '[f]acts are what statements (when true) state; they are not what statements are about' (1950, p. 196). This gets carried over into Williamson's account of factive attitudes because he takes their characteristic expression in natural language – viz. as a *factive mental state operator* – as ascribing the subject with an attitude towards a proposition. It will then follow from saying that '*S* Φ-es that *f*'' (where 'Φ' denotes some candidate factive attitude and '*f*' some fact) that '*S* grasps the proposition that (corresponds to) *f*'' (Williamson 2000, p. 36). If that feature of their characteristic expression in natural language is taken to hold in general for factive stative attitudes – and it seems as if Williamson takes this to be the case – there can be no factive stative attitude obtaining in some of the cases I voiced as problematic cases for the non-trivial versions of the *Constitutive Thesis*.[37] Recall that some of those cases involved knowledge arising from or in virtue of states and processes that

[37] To be fair Williamson mentions that talk about facts as a *sui generis* metaphysical category would require only terminological changes to his position (2000, pp. 43–4). However, Williamson still wants to reject that belief and knowledge relate its bearer to different kinds of objects (viz. that one knows facts but believes propositions).

represented the objects of knowledge non-conceptually. These non-conceptual states would nevertheless relate a perceiver or an agent to certain facts; however, there will be no need – nor room for – invoking a relation between these subjects and propositions. Relating to propositions requires conceptual sophistication and that the individual satisfies a certain understanding condition which, I think, would extend beyond the capacities of what Brandom calls non-verbal and merely sentient animals: that is, creatures who share something with us in 'being *awake'* and thus relate to the environment in a different way than the mere 'reliable differential responsiveness we sentients share with artifacts such as thermostats and land mines' (1994, p. 5; 1995, p. 425). Rather than taking this to count against the idea that these animals are knowers, however, I think it counts against the idea that knowing must involve factive stative attitudes. Like Dretske I think that it would be favourable if we could have a characterization of knowledge 'that would at least allow for the possibility that animals (a frog, rat, or my dog) could know things without my having to suppose them capable of the more sophisticated intellectual operations involved in traditional analyses of knowledge' (Dretske 1985, p. 177). That can be done if we leave room for non-propositional factive relations and non-conceptual knowledge and that one may know something by virtue of non-conceptual ways of representing these facts. This requires knowing-relations that cannot be accounted for within a framework where knowing is defined as an attitude whose relata are minds and true propositions. Since propositions require too much sophistication and understanding, it follows that knowledge cannot be constituted by all and only the mental state of knowing. In a word, it gives us reason to dismiss Williamson's suggestion that knowledge is constituted by a factive mental attitude or state.

Williamson's account thus faces a problem similar to the aforementioned *Problem of Disunity* which we raised against non-trivial versions of the *Constitutive Thesis*. Just like the non-trivialists, Williamson's approach has failed to retain the unity of knowledge against the plurality of ways of knowing. The many ways of knowing goes wider than the range of cases where some knower has a factive stative attitude (FSA) towards some true proposition. In particular, one may know some fact, *f*, without having the conceptual resources required for representing its corresponding true proposition, p_f. In the previous chapter, I argued that Williamson's approach would fail to account for practical knowledge. If we add that challenge to the problems I have just discussed, I think we have enough reason to seriously consider the *Causal Thesis* and whether it fares any better. Now the previous chapter has already made the claim that the framework of the *Causal Thesis* is at an advantage with

respect to the metaphysics of practical knowledge; let me therefore also show that it is better with respect to the problems we have been discussing in this chapter.

§4 The CAUSAL THESIS

As I will argue, the *Causal Thesis* is at an advantage with respect to the lines of criticism I have voiced against the non-trivial *Constitutive Thesis* and Williamson's position. A proponent of the *Causal Thesis* is, on the one hand, free to treat the knowledge-relation as a basic relation, whose relata are minds and objects of knowledge that one may grasp or relate to both conceptually and non-conceptually. Following Hossack, I shall treat these objects of knowledge as *facts* and briefly discuss what I mean by that and what role they are meant to play in the theory. On the other hand, the *Causal Thesis* is also compatible with whatever the outcome may be of the internalist–externalist debate regarding the nature of psychological states and attitudes. In particular, it does not involve a commitment to the existence and causal-psychological efficacy of broad psychological states – whether they are content-broad or attitude-broad – without thereby severing the 'in-virtue-of' connection that holds between the state of knowledge and one's state of mind. I shall take that to be another virtue of the view. A third point is that, just like Williamson's model, the *Causal Thesis* allows us to diagnose and reply to some familiar ways of arguing for scepticism about knowledge. A fourth virtue of the view is that it allows us to explain the nature of practical knowledge, which, of course, is also my reason for being interested in this view in the first place and the immediate cause of our engaging with these discussions about the metaphysics of knowledge.

As we saw, the trouble for the other views is the problem of retaining the unity of the knowledge-relation in the face of the plurality of ways people and non-human animals can come to know things. The *Causal Thesis* is suggested as a way to retain the unity of knowledge in the face of such plurality. Since the relation between knowledge and the psychological states and non-mental states that obtain in virtue of which an agent knows something is causal rather than constitutive, the *Causal Thesis* would allow for a vast plurality of different ways a cognitive system may bring about this relation without this plurality threatening the unity of the knowledge-relation. Its main claim is therefore that knowledge is a basic, non-composite relation that is caused by a myriad of different psychological states of a knowing subject when these states obtain in the right way and in the right circumstances. To retain its unity, however, it is important that the *relata* of the relation are uniform as well. What Hossack suggests is that the relation takes all and only individual minds as its subject. I

will not have much more to say about this. What some agent must be like to qualify as a minded being is a controversial question; however, I will assume that we have enough of a working conception to see what kind of creature is meant to fill the subject slot of the relation. It is probably something that is cognitively sophisticated enough to rule out plants and amoeba, yet broad enough to include higher-order mammals, sea creatures, some of the birds, and so on. Let us therefore turn to the other *relatum*: the object of knowledge.

What we need as an object of knowledge is something that can preserve unity despite the plurality of ways one may know. Hossack suggests that we invoke facts, which he wants to treat as a *sui generis* non-conceptual entity. In his view, facts belong to the furniture of the world and are a *sui generis* category of particulars – a separate kind of entity in their own right, as it were – that are composed when individuals and universals partake in a certain combination-relation, something they do whenever the individual (or group of individuals) instantiates the property in question. Facts are thus not equated with the class of true propositions or contents, and it is therefore assumed that subjects may relate to some fact, *f*, without mastering the concepts needed to describe *f*. In this way, facts can explain the unity of knowledge since knowing will then be to partake in a basic relation that holds between minds and facts. Undergoing conceptual thought processes and inquiries will in fortunate cases cause a thinking creature, *S*, to know some fact, *f*, in virtue of conceptually representing *f* in the assertive mood by, for instance, believing some proposition *p* that corresponds to *f*. But it may also be the case that some animal, *A*, knows *f* in virtue of undergoing non-conceptual perceptual processes in some favourable context where the perceptual states represent *f* (that is, without requiring that *A* possesses the relevant concepts needed to describe *f*).

Of course, while this solution does indeed require a separate ontology of facts, Hossack argues that metaphysics provides a variety of good reasons for invoking an ontology of facts anyway. They are, for instance, needed as causal relata and to substantiate a theory of material beings. In addition, one may use a theory of facts to explain the nature of contents or propositions. Finally, facts allow us to ground logic and might provide the metaphysical glue that ties or combines particulars with universals and properties in a way that allows us to account for instantiation without throwing ourselves into what is known in the literature as *Bradley's Regress*.[38] A theory of facts thus offers ground for an account of objective

[38] Bradley's Regress is the well-known problem that, as Francis H. Bradley argued, pertains to versions of Realism that take instantiation of a property as the instantiation of a universal but which also treats instantiation as a universal (viz. as the relation that obtains between Socrates and the universal, *wisdom*, when and only when Socrates is wise). The problem with this Realist theory is that it treats some *S* as instantiating some *U* only if the instantiation itself is instantiated. Since instantiation is just another instantiation of a universal the first instantiation – viz. of *S* being *U* – will require a new universal, *U**, whose instantiation would instantiate that *S* instantiates *U*. However, it now becomes clear that we need another universal, *U***, whose instantiation would instantiates the instantiation that instantiates that *S* instantiates *U*. By now,

resemblance and to account for the objective articulation of natural differences. Facts may therefore earn their living as an essential supplement to the metaphysical doctrine of *Realism* about properties and universals since, as Hossack says, 'Realism is an incomplete theory unless it incorporates a theory of facts' (2007, p. 41).[39]

It will lead us too far astray to properly evaluate the arguments that favour an ontology of facts and whether these considerations favour Hossack's theory of facts. What we need to preserve the unity of knowledge, however, is some such notion or entity that knowers and thinkers can relate to without requiring the corresponding concept-possession that is required for adequately describing the state of affairs one knows. We already know that knowledge can be had in virtue of entertaining different kinds of contents that include non-conceptual kinds of representation. Facts, rather than propositions and contents, would therefore be the preferred candidate relata for knowledge if a theory of facts is satisfactory. Since the *Causal Thesis* recommends itself and it requires a category such as facts – viz. since we need some category than can be the second *relatum* of the knowledge-relation – I shall therefore assume that a theory of facts can be had. Following Hossack, I shall call this category facts and I will therefore assume that the *relata* of the knowledge-relation are minds and facts.

A consideration that favours facts as the relata of the knowledge-relation is that there is no immediate need to invoke a level of content or representation in knowledge since there is 'no need to make room for mistaken knowledge' (Hossack 2007, p. 6). Knowledge is, after all, a factive relation that pertains all and only to what is the case. The relation can therefore not be in error in the sense that it could fail to represent the world adequately. There is therefore no need for a level of content to account for mistaken representations of how things are. Indeed, there is little reason to invoke representation at all as a component of the knowing relation. The function of knowledge is simply to relate its bearer directly to the facts. Where we need representation is at the level of psychological states since they cause us to know when they are right and in the right circumstances, but may equally well fail to bring about knowledge by getting the facts wrong – viz. by representing some *f* as obtaining when it fails to do so. Of course, when we attribute someone with knowledge or report them as knowing we seem to invoke a level of content; that, however, can be explained inasmuch as the knowledge-report not only indicates the fact known: it also conveys the provenance of the

however, it is clear that we are on our way towards an infinite regress and thus that this version of Realism has got something wrong since it becomes clear that 'nothing further in the way of positing universals will help solve Bradley's Regress', as Hossack says (2007, p. 41). See Bradley (1897). Armstrong (1978a; 1978b; 2003) discusses a similar problem but calls it 'the relation regress'.

[39] See Hossack (2007, pp. 33–4). Mellor (1995, pp. 130–9) argues that facts are the causal relata and Hossack (2000) emphasizes their role in grounding plural logic. Content has also been sought explained in terms of facts: see Hossack (2007, ch. 3) and Jeff King (2007).

knowledge. In order to reveal the way or how someone knows some facts we need to invoke the contents or representations that were involved in the process in virtue of which the subject knows. It is therefore this feature of knowledge-reports that makes knowledge-attributions sensitive to the way in which the objects known are represented by its subject.

To take an example: I may, for instance, report to you that my colleague knows the tallest philosopher at the institute is an idiot, thereby conveying to you that he knows this fact, *f* – namely that some property, *F*, applies to some person, *P* – however I will then also convey to you that my colleague knows this fact under a certain description. For my colleague it is the description 'the tallest philosopher' that picks out *P* and she knows that whoever is picked out by this description instantiates the property of idiocy. Some other agent may know the same fact under a different description and that would be reflected in our report. Given the report's sensitivity to a description it does not necessarily imply that our subject knows that he himself, given that he is the tallest philosopher, is an idiot. To do that one would need to add that he also knows that he is the tallest philosopher. There will be need for content to make room for these distinctions in knowledge-reports; however, we should not infer from this that the knowledge-relation conveyed by the report is a relation between minds and contents. The need for content is due to the function of the reports – viz. that they convey the fact known and the provenance of knowledge – and does not necessarily pertain to the knowledge-relation itself. In fact, given our considerations about the metaphysics of knowledge it is reason to think that it only pertains to the provenance of knowledge: that is, to the psychological states in virtue of which one knows. A knowledge-report conveys not only that some *S* knows some *f* but also how they know *f* by providing the content or representation of the mental act in virtue of which *S* knows *f*. Hossack thus suggests that we analyze the logical form of knowledge-reports, such as '*S* know that *A*', as follows (2007, pp. 6–7):

> '*S* knows that *A*' $=_{df}$ $(\exists x)(\exists p)$ (*x* is a mental act & *f* is a fact & content(*x*) = *that-A* & *that-A* is a mode of presentation of *f* & *S* knows of *f* in virtue of *x*)

With this analysis in hand we can grant Wittgenstein that '"I know" is supposed to express a relation, not between me and the sense of a proposition (like "I believe") but between me and a fact so that the fact is taken into my consciousness' (*OC* §90); however, the knowledge-report also conveys the way in which the fact was taken into one's consciousness.

What these points touch on is the independence of the knowledge-relation with respect to the nature of the psychological states and events in virtue of which the knowledge-relation

obtains that is allowed by the framework of the *Causal Thesis*. This framework, as I have said, consists of the claims that the provenance of knowledge causes rather than constitutes the knowledge it brings about and that knowledge is a basic non-composite mental relation that connects minds without intermediaries with the facts. The points I just made is in recognition of the fact that there may be no direct theoretical needs to impute a range of features pertaining to the ways of knowing to the nature of knowledge itself. The *Causal Thesis* allows us to recognize that features needed for a theory of ways of knowing need not carry over as features of the knowledge-relation, unless, of course, there are specific theoretical needs for them also in the metaphysics of knowledge. Contents and representations are, as we just saw, features that are not needed in the metaphysics of knowledge; but they are needed in the process where a subject brings knowledge about, whether they be conceptual or non-conceptual. Likewise, mental states and acts are needed as a provenance of knowing – at least for creatures like us – but that should not lead to us to blindly conceive knowledge itself as such acts and states. As Hossack notes, while an omniscient mind will know all there is to know, due to its omniscience, there will be no theoretical need to attribute such a Being with judgements, beliefs, or similar psychological states and acts. Indeed, one may question whether such a Being – though it is a mind by virtue of knowing things – will have a need for a psychology at all (Hossack 2007, p. 267). As Spinoza wrote, 'God's intellect (...) differs from man's intellect both in respect of essence and existence, and cannot agree with it in any respect other than name' (*E* Pr.17, *Scholium*). What we would share with such a creature is thus not a psychology. Rather, we share an ability to know and thus partake in a relation that connects its bearer without intermediaries directly with the facts. The framework of the *Causal Thesis* thus allows us to recognize an important distinction, namely that 'having a mind (possessing knowledge) is conceptually distinct from having a psychology' and that 'there can perhaps be mental terms that are not psychological terms' (Hossack 2007, p. 267).

Invoking facts as the *relata* of knowledge is thus a way to retain unity in the face of plurality and gives the *Causal Thesis* an advantage with respect to the Williamson view, in particular, but also with respect to the orthodox view. Knowledge, as it is conceived on the *Causal Thesis*, is thus a basic, non-composite mental relation holding between some subject or mind, *S*, and some fact, *f*, in virtue of *S* undergoing certain mental processes and events in favourable circumstances. A consequence of this is that it is open what kind of mental process that is. That takes me to a second respect in which the *Causal Thesis* has an advantage over Williamson's view. Proponents of the *Causal Thesis* are thus in a position where they can

concede to internalists – if internalism turns out to be right – that the properties in virtue of which a subject knows some *f* are the casual conditions of *S*'s behaviour, not the knowledge itself. These psychological properties are, after all, sufficient for these effects.[40] Given that there is no longer a need to defend a causal role for knowledge with respect to these effects – after all, we have given up the idea that knowledge is a mental state and need no longer defend that it is a cause of an agent's behaviour – one may thus accept that knowledge is not causally efficacious with respect to these events. We may still insist that knowledge is essential for a range of causal-psychological explanations given that this condition greatly facilitates the probability of the psychological states in question being causally efficacious; but its role will be more like dryness than the role of the spark with respect to the fire. In other words, knowledge is a background condition that may enhance and increase the probability of certain causal powers being effective. Thus, it may serve an irreducible role in psychological explanation, without contributing causally to our behaviour.

A thing that counts to the advantage of the Williamson view, of course, is that if knowledge is a mental state it allows us to undercut certain familiar arguments for scepticism about knowledge and justification by undermining one of its crucial lemmas. Sceptics often start out from the claim that a subject, *S*, can be a psychological duplicate of some counterpart subject, *S**, who happens to be placed in a sceptical scenario where very few – if any – of her beliefs are true even though *S**'s experiences and state of mind appear to be the same as *S*'s. If Williamson is right and knowledge is a mental state, however, your total mental state is not neutral between knowing and not knowing. To assume, as the sceptic does, that you are in the same total mental state, M_1, in a world or case, α, and the curious sceptical scenario, @ – say, where @ is a world in which you are the victim of an all-powerful evil demon bent on deceiving you – amounts to assuming from the outset that you do not know much in α. Given that this is what the sceptical argument was meant to prove, the argument will fail to establish the truth of scepticism in a non-question-begging manner. Williamson's model thus has the resources to show that 'any mental life in the sceptical scenario is of a radically impoverished kind' and that an anti-sceptic need not assume in the absence of decisive arguments that our mental life resembles the mental life of the sceptical scenario (Williamson 2000, p. 6, 26).[41]

Now the *Causal Thesis* may make a similar move by defining belief and other relevant mental states in terms of knowledge. Once the *Constitutive Thesis* is thoroughly rejected,

[40] The principle that supports this conclusion is that a property, *P*, is efficacious only if none of the properties in virtue of which it is instantiated would suffice for its effects (Steglich-Petersen 2005, p. 26). If the psychological processes in virtue of which *S* knows some *f* suffice to cause *S* Φ-ing, it follows that those properties rather than knowledge are causally efficacious with respect to the Φ-ing.
[41] For a similar argument, see Williamson (2005). Also, see Adam Leite on the advantage of rebutting scepticism (2005, p. 166).

there is room for thinking of knowledge as something that comes before belief and other mental states in the order of philosophical explanation. Knowledge as a basic mental relation can be treated as a relation that is the very ground for such mindedness and having mental states in the first place. It thus allows us to entertain what Hossack calls an 'epistemic priority theory' – or a *Faculty Theory* – which basically is to philosophically account for the nature of such mental states in terms of knowledge rather than the other way around (2007, p. 11, 266). One may then deny the sufficiency of the functionalist thesis that we can define a system of psychological states in terms of an interaction between inner states and behaviour without presupposing the concept of knowledge. By denying its sufficiency, however, one may retain what the functionalist theory in question has to say about the functional role of the mental state it accounts for – viz. what it has to say about the relationship between such a state and the subject's action and behaviour, the way it gets updated in light of experiences, and so on – however, one may add that the functionalist story lacks something crucial when, despite all its talk about behavioural and causal relationships, it omits mentioning the system's capacity for knowing things (Hossack 2007, pp. 266–7). On the Epistemic Priority/Faculty Theory one may thus argue that some cognitive system, *S*, can only be a system of beliefs – where a belief is conceived as a mental state – if at least some of *S*'s belief-states cause *S* to know. More generally, one may argue that group or category of states, M_1, M_2,...,M_n, that *S* has can only be mental states insofar as some of them cause *S* to know. It would follow, that we have resources also on this model to reject the crucial sceptical lemma of the traditional argument for scepticism since the mental life in the sceptical scenario is not recognizable as anything like us at all. Such a poor creature would not only be bereft of its knowledge but also of its mind. And that adds significantly to the sceptic's burden of proof.[42]

A final respect in which the *Causal Thesis* has an advantage over its competitors is, as I argued in the previous chapter, that it can be used to account for practical knowledge, or the knowledge an agent has in virtue of acting intentionally. By reading the 'in-virtue-of' relation as a causal claim rather than a claim about constitution allows a proponent of the *Causal Thesis* to avoid the conclusion that knowledge is composed of the properties, states, or entities in virtue of which it is instantiated (viz. the psychological states that figure as causal conditions in causal-psychological explanations). Thus, the causal factors for behaviour whose explanation requires attributing the agent with knowledge – viz. the burglar remaining

[42] Importantly, this move only works against sceptical arguments that capitalize on more or less exotic contrast scenarios. It will not work as a response to arguments such as the Unger's argument that knowing would involve an unwarranted dogmatic state of mind (1974; 1975, ch. 3), or the argument from lottery cases that has more recently been explored by, among others, John Hawthorne (2004, ch. 1).

in the house because he knows there is a diamond there – need not be constitutive components of the burglar's knowledge-relation. Similarly, if some agent may know what she is doing by virtue of acting intentionally or willing her action, there is no need to think that so acting/willing is a constitutive component of her knowledge. It is this feature of the *Causal Thesis* and its framework that allows knowledge to be an essential factor of different causal-psychological explanations that do not share any psychological states, acts, or components. In a word, all that the many ways of knowing must share is the ability to cause and bring about knowledge. What I argued in the previous chapter is that distinctively practical thinking can do that too and is therefore a way of knowing that is distinctively practical without thereby requiring us to invoke a knowledge-relation that is practical in this sense. That this framework allows this is, of course, in turn a point in favour of the *Causal Thesis* and its framework. In my view, it is sufficient all by its own to recommend accepting its framework in the metaphysics of knowledge. However, given that there are many sceptics out there with doubts and worries concerning the nature of the phenomenon we call practical knowledge or knowledge in action, it was important to show that the framework derives strong motivation from the metaphysics of knowledge independently of the issue of practical knowledge.

§5 CONCLUSION

The aim of this chapter has been relatively independent of the previous discussions in that it has sought to provide independent support for a view about the metaphysics of knowledge and for a view about how the state of knowledge relates to our ways of knowing, or the provenance of knowledge. I have argued that it is best to conceive this relationship as a causal relationship rather than a constitutive relation – whether trivial or non-trivial – and that we should conceive knowledge as a basic non-composite mental relation whose existence is caused by its provenance rather than constituted by it. In a word, I have favoured the framework offered by Hossack and his *Causal Thesis*. The virtue of this framework is basically the distance it allows between the nature of knowledge and our ways of knowing, a distance that allows us to retain unity to knowledge in the face of a great variety of ways we may come to know things and 'the great variety of mental acts and mental state in virtue of which we have knowledge', as Hossack says (2007, p. 12). As I argued in the previous chapter, that distance is also crucial with respect to the issue of knowledge in action since it makes Hossack's framework well-suited as basis for a conception of practical knowledge as the knowledge that results from a distinctively practical way of knowing.

Bibliography

Abrahamsen, A., W. Bechtel, R. Mathews and R. Stainton (eds.) 2006, *Contemporary Debates in Cognitive Science*, Blackwell Publishing, Oxford.

Ackrill, J.L. 1978, 'Aristotle on Action', originally published in *Mind*, Vol. 87/1978, pp. 165–79, reprinted in Rorty, A.O. (ed.) 1980, *Essays on Aristotle's Ethics*, University of California Press, Berkeley (C), pp. 93–101.

Aguilar, J.H. and A.A. Buckareff (eds.) forthcoming, *Causing Human Action: New Perspectives on the Causal Theory of Action*, Bradford Book/The MIT Press, Cambridge.

Alston, W.P. 1991, *Perceiving God: The Epistemology of Religious Experience*, Cornell University Press, Ithaca (NY).

Altham, J.E.J. and R. Harrison (eds.) 1995, *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, Cambridge University Press, Cambridge.

Alvarez, M. 2009, 'Acting Intentionally and Acting for a Reason', *Inquiry*, Vol. 52, No. 3/2009, pp. 293–305.

Anscombe, G.E.M. 1957/2000, *Intention*, Harvard University Press, Cambridge (MS).

------. 1958, 'Modern Moral Philosophy', originally published in *Philosophy* 53/1958, Cambridge University Press, pp. 1–19, reprinted in Geach, M. and L. Gormally (eds.) 2005, *Human Life, Action and Ethics: Essays by G.E.M. Anscombe*, Imprint Academic, Exeter, pp. 169–94.

------. 1962, 'On Sensations of Position', originally published in *Analysis*, Vol. 22, 3/1962, reprinted in her 1981, *Metaphysics and the Philosophy of Mind*, Blackwell Publishing Company, Oxford, pp. 71–4.

------. 1979, 'Under a Description', *Nous*, Vol. 13/1979, pp. 219–33.

------. 1981, *Metaphysics and the Philosophy of Mind*, Blackwell Publishing Company, Oxford.

------. 1983, 'The Causation of Action', in Ginet, C. and S. Shoemaker (eds.) 1983, *Knowledge and Mind*, Oxford University Press, Oxford, pp. 174–90.

------. 1995, 'Practical Inference', originally published in Hursthouse, R., G. Lawrence and W. Quinn (eds.) 1995, *Virtues and Reasons: Philippa Foot and Moral Theory*, Clarendon Press, Oxford, pp. 1–34, reprinted in Geach, M. and L. Gormally (eds.) 2005, *Human Life, Action and Ethics: Essays by G.E.M. Anscombe*, Imprint Academic, Exeter, pp. 109–47.

Aquinas, T. 2008, *The Summa Theologiæ of Thomas Aquinas: Volume 1*, Novantiqua, Scotts Valley (C).

Ariew, R. and E. Watkins (eds.) 1998, *Modern Philosophy: An Anthology of Primary Sources*, Hackett Publishing Company, Indianapolis/Cambridge.

Aristotle 1984a, *Eudemian Ethics*, in Barnes, J. (ed.) 1984b, *The Complete Works of Aristotle – Volume Two*, Princeton University Press, Princeton, pp. 1922–1981.

------. 1984b, *Magna Moralia*, in Barnes, J. (ed.) 1984b, *The Complete Works of Aristotle – Volume Two*, Princeton University Press, Princeton, pp. 1868–1921.

------. 1984c, *Nicomachean Ethics*, in Barnes, J. (ed.) 1984b, *The Complete Works of Aristotle – Volume Two*, Princeton University Press, Princeton, pp. 1729–1867.

------. 1984d, *On the Soul*, in Barnes, J. (ed.) 1984a, *The Complete Works of Aristotle – Volume One*, Princeton University Press, Princeton, pp. 641–92.

------. 1984e, *Posterior Analytics*, in Barnes, J. (ed.) 1984a, *The Complete Works of Aristotle – Volume One*, Princeton University Press, Princeton, pp. 114–66.

Armstrong, D.M. 1973, *Belief, Truth and Knowledge*, Cambridge University Press, Cambridge.

------. 1978a, *Nominalism and Realism: Universals and Scientific Realism – Volume I*, Cambridge University Press, Cambridge.

------. 1978b, *A Theory of Universals: Universals and Scientific Realism – Volume II*, Cambridge University Press, Cambridge.

------. 2003, 'How do Particulars stand to Universals?', in Zimmerman, D. (ed.) 2004, *Oxford Studies in Metaphysics: Volume 1*, Oxford University Press, Oxford, pp. 139–54.

Arpaly, N. 2000, 'On Acting Rationally against One's Best Judgment', *Ethics*, Vol. 110/2000, The University of Chicago, pp. 488–513.

------. 2003, *Unprincipled Virtue*, Oxford University Press, New York.

------. 2006, *Merit, Meaning, and Human Bondage: An Essay on Free Will*, Princeton University Press, Princeton.

Atkinson, R.L., R.C. Atkinson, D. Bem, S. Nolen–Hoeksema and E.E. Smith 1996, *Hilgard's Introduction to Psychology*, 12th ed., Harcourt Brace College Publishers, Orlando.

Audi, R. 1973, 'Intending', *Journal of Philosophy*, Vol. 70/1973, pp. 387–403.

Augustine 2008, *Confessions*, (tr. Chadwick, H.), Oxford University Press, Oxford.

Austin, J.L. 1953, 'How to Talk. Some Simple Ways', *Proceedings of the Aristotelian Society*, Vol. 53, pp. 227–46.

Baier, A.C. 1970, 'Act and Intent', *The Journal of Philosophy*, Vol. 67, 19/1970, pp. 648–58.

Barber, A. (ed.) 2003, *Epistemology of Language*, Oxford University Press, Oxford.

Barel, F.O. 2009, *Self–Knowledge and Annti–Individualism: A Defence of Neo–Rationalist Compatibilism*, Ph.D. Thesis, University of Oslo.

Barnes, J. (ed.) 1984a, *The Complete Works of Aristotle – Volume One*, Princeton University Press, Princeton.

------. 1984b, *The Complete Works of Aristotle – Volume Two*, Princeton University Press, Princeton.

Baron-Cohen, S. 2001, 'Theory of Mind in Normal Development and Autism', *Prisme*, Vol. 34/2001, pp. 174–83.

Bealer, G. 1998, 'A Theory of Concepts and Concept Possession', *Philosophical Issues*, Vol. 9, Rigdeview Publishing Company, Atascadero (CA).

------. 2004, 'The Origins of Modal Error', *Dialectica*, Vol. 58, pp. 11–42.

Beaney, M. (ed.) 1997, *The Frege Reader*, Blackwell Publishing, Oxford.

Beckett, S. 1955, *Molloy*, Grove Press, New York.

Bell, D. and N. Cooper (eds.) 1990, *The Analytic Tradition*, Blackwell Publishing, Oxford.

Berlin, I. 1969, *Four Essays on Liberty*, Oxford University Press, Oxford.

Bergman, M. 2006, *Justification without Awareness: A Defence of Epistemic Externalism*, Oxford University Press, Oxford.

Bermúdez, J.L. 1998, *The Paradox of Self–Consciousness*, MIT Press, Cambridge (MA).

Binkley, R., R. Bronaugh and A. Marras (eds.) 1971, *Agent, Action, and Reason*, University of Toronto Press, Toronto.

Bishop, J. 2002, 'Faith as Doxastic Venture', *Religious Studies*, Vol. 38, No. 4/2002, pp. 471–87.

------. 2005, 'On the Possibility of Doxastic Venture: A Reply to Buckareff', *Religious Studies*, Vol. 41, pp. 447–51.

------. 2010, 'Faith', *Stanford Encyclopedia of Philosophy* [online], Stanford University. Available at: http://plato.stanford.edu/entries/faith/ [20.08.2010]

Blackburn, S. 1993, *Essays In Quasi–Realism*, Oxford University Press, Oxford.

Bradley, F.H. 1897, *Appearance and Reality*, 2nd ed., Oxford University Press, Oxford, reprinted in his 1994, *Writings on Logic and Metaphysics*, (Allard, J.W. and G. Stock (eds.)), Oxford University Press, Oxford.

------. 1994, *Writings on Logic and Metaphysics*, (Allard, J.W. and G. Stock (eds.)), Oxford University Press, Oxford.

Brandom, R. 1994, *Making it Explicit: Reasoning, Representing, and Discursive Commitment*, Harvard University Press, London/Cambridge (MA).

------. 1995, 'Knowledge and the Social Articulation of the Space of Reasons', originally published in *Philosophy and Phenomenological Research*, Vol. 55/1995, pp. 895–908, reprinted in Kim, J. and E. Sosa (eds.) 2000, *Epistemology: An Anthology*, Blackwell Publishing, Oxford/Malden (MA), pp. 424–32.

Branquinho, J. (ed.) 2001, *The Foundations of Cognitive Science*, Oxford University Press, Oxford.

Bratman, M. 1991, 'Cognitivism about Practical Reason', originally published in *Ethics*, Vol. 102/1991, pp. 117–28, reprinted in his 1999, *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge University Press, Cambridge, pp. 250–64.

------. 1996, 'Identification, Decision, and Treating as a Reason', originally published in *Philosophical Topics*, Vol. 24/1996, pp. 1–18, reprinted in his 1999a, *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge University Press, Cambridge, pp. 185–206.

------. 1999a, *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge University Press, Cambridge.

------. 1999b, *Intention, Plans and Practical Reason*, CSLI Publications, Stanford (C).

------. 2001, 'Two Problems about Human Agency', originally published in *Proceedings of the Aristotelian Society*, 101/2001, pp. 309–26, reprinted in his 2007, *Structures of Agency*, Oxford University Press, Oxford, pp. 89–105.

------. 2007, *Structures of Agency*, Oxford University Press, Oxford.

------. 2009, 'Setiya on Intention, Rationality and Reasons', *Analysis Reviews*, Vol. 69, 3/2009, pp. 510–21.

Brill, A.A. (ed.) 1995, *The Basic Writings of Sigmund Freud*, The Modern Library, New York.

Broadie, S. 1991, *Ethics with Aristotle*, Oxford University Press, Oxford.

Brown, S.C. (ed.) 1974, *Philosophy of Psychology*, The Macmillian Press/Barnes & Noble Books, London/New York.

Brunero, J. 2009, 'Against Cognitivism about Practical Rationality', *Philosophical Studies*, Vol. 146/2009, pp. 311–25.

Burge, T. 1979, 'Individualism and the Mental', originally published in *Midwest Studies in Philosophy*, 4/1979, pp. 73–121, reprinted in his 2007, *Foundations of Mind: Philosophical Essays, Volume 2*, Clarendon Press, Oxford, pp. 100–50.

------. 1986, 'Intellectual Norms and Foundations of Mind', originally published in *Journal of Philosophy*, Vol. 83, pp. 697–720, reprinted in his 2007, *Foundations of Mind: Philosophical Essays, Volume 2*, Clarendon Press, Oxford, pp. 254–74.

------. 1990, 'Frege on Sense and Linguistic Meaning', originally published in Bell, D. and N. Cooper (eds.) 1990, *The Analytic Tradition*, Blackwell Publishing, Oxford, pp. 30–60, reprinted in his 2005, *Truth, Thought, Reason: Essays on Frege*, Oxford University Press, Oxford, pp. 242–70.

------. 2005, *Truth, Thought, Reason: Essays on Frege*, Oxford University Press, Oxford.

------. 2007, *Foundations of Mind: Philosophical Essays, Volume 2*, Clarendon Press, Oxford.

Burnyeat, M. 1980, 'Aristotle on Learning to be Good', in Rorty, A.O. (ed.) 1980, *Essays on Aristotle's Ethics*, University of California Press, Berkeley (C), pp. 69–92.

Calvin, J. 1559, *Institutes of the Christian Religion*, (tr. Lewis, F.B.), in McNewill, J.T. (ed.) 1960, *Library of Christina Classics*, The Westminster Press, Philadelphia.

Camus, A. 1955, *The Myth of Sisyphus*, Hamish Hamilton, London.

Carey, S. 1985, *Conceptual Change in Childhood*, Bradford Books/MIT Press, Cambridge (MA).

------. 2001, 'Evolutionary and Ontogenetic Foundations of Arithmetic', *Mind and Language*, Vol. 16/2001, pp. 37–55.

Carey, S. and E. Spelke 1996, 'Science and Core Knowledge', *Philosophy of Science*, Vol. 63, 4/1996, pp. 515–33.

Cartwright, N. 1980, 'Do the Laws of Physics State the Facts?', originally published in *Pacific Philosophical Quarterly*, Vol. 61/1980, pp. 75–84, reprinted in Curd, M. and J.A. Cover (eds.) 1998, *Philosophy of Science: The Central Issues*, W.W. Norton & Company, New York, pp. 865–77.

Cassam, Q. 2007, *The Possibility of Knowledge*, Oxford University Press, Oxford.

------. 2010, 'Judging, Believing, Thinking', *Philosophical Issues*, Vol. 20/2010, pp. 80–95.

Chalmers, D.J. (ed.) 2002, *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, Oxford.

Chappel, T. 2009, *Ethics and Experience: Life Beyond Moral Theory*, McGill–Queen's University Press, Ithaca.

Chignell, A. 2010, 'The Ethics of Belief', *Stanford Encyclopedia of Philosophy* [online], Stanford University. Available at: http://plato.stanford.edu/entries/ethics-belief/ [14.06.2010]

Chisholm, R. 1957, *Perceiving: A Philosophical Study*, Cornell University Press, Ithaca (NY).

Clifford, W.K. 1877, 'The Ethics of Belief', in *Contemporary Review*.

Cohon, R. 2010, 'Hume's Moral Philosophy', *Stanford Encyclopedia of Philosophy* [online], Stanford University. Available at: http://plato.stanford.edu/entries/hume–moral/ [27.08.2010]

Comrie, B. 1976, *Aspect*, Cambridge University Press, Cambridge.

Conee, E. and R. Feldman 2004, *Evidentialism: Essays in Epistemology*, Oxford University Press, New York.

Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge.

Crisp, R. and M. Slote (eds.) 1997, *Virtue Ethics*, Oxford University Press, Oxford.

Cullity, G. and B. Gaut (eds.) 1997, *Ethics and Practical Reason*, Clarendon Press, Oxford.

Curd, M. and J.A. Cover (eds.) 1998, *Philosophy of Science: The Central Issues*, W.W. Norton & Company, New York.

Dancy, J. 1993, *Moral Reasons*, Blackwell Publishing Company, Oxford.

------. 2000, *Practical Reality*, Oxford University Press, Oxford.

Danto, A. 1963, 'What We Can Do', *Journal of Philosophy*, Vol. 60/1963, pp. 434–45.

------. 1965, 'Basic Actions', *American Philosophical Quarterly*, Vol. 2/1965, pp. 141–8.

------. 1968, *Analytical Philosophy of Knowledge*, Cambridge University Press, Cambridge.

------. 1973, *Analytical Philosophy of Action*, Cambridge University Press, Cambridge.

Davidson. D. 1963, 'Actions, Reasons and Causes', originally published in *Journal of Philosophy*, Vol. 60/1963, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 3–19.

------. 1967, 'The Logical Form of Action Sentences', originally published in Rescher, N. (ed.) 1967, *The Logic of Decision and Action*, University of Pittsburgh Press, Pittsburgh, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 105–22.

------. 1969a, 'The Individuation of Events', originally published in Rescher, N. and D. Reidel (eds.) 1969, *Essays in Honor of Carl G. Hempel*, Reidel Publishing Company, Dordrecht, pp. 216–34, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 163–80.

------. 1969b, 'True to the Facts', originally published in *Journal of Philosophy*, Vol. 66/1969, pp. 748–64, reprinted in his 2001b, *Inquiries into Truth and Interpretation*, 2nd ed., Oxford University Press, Oxford, pp. 37–54.

------. 1970a, 'How is Weakness of Will Possible?', originally published in Feinberg, J. (ed.) 1970, *Moral Concepts* reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 21–42.

------. 1970b, 'Mental Events', originally published in Foster, L. and J.W. Swanson (eds.) 1970, *Experience and Theory*, The University of Massachusetts Press/Duckworth, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 207–27.

------. 1971, 'Agency', originally published in Binkley, R., R. Bronaugh and A. Marras (eds.) 1971, *Agent, Action, and Reason*, University of Toronto Press, Toronto, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 43–61.

------. 1973, 'The Material Mind', in Davidson 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 245–59.

------. 1974, 'Psychology as Philosophy', originally published in Brown, S.C. (ed.) 1974, *Philosophy of Psychology*, The Macmillian Press/Barnes & Noble Books, London/New York, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 229–44.

------. 1976a, 'Hempel on Explaining Action', originally published in *Erkentnis*, Vol. 10/1976, pp. 239–53, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 261–75.

------. 1976b, 'Hume's Cognitive Theory of Pride', originally published in *Journal of Philosophy*, Vol. 73/1976, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 277–90.

------. 1978, ‘Intending’, originally published in Yovel, Y. And D. Reidel (eds.) 1978, *Philosophy of History and Action*, The Magnes Press, The Hebrew University, reprinted in his 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, pp. 83–102.

------. 1980, ‘A Unified Theory of Thought, Meaning, and Action’, originally published as ‘Toward a Unified Theory of Meaning and Action’, in *Grazer Philosophische Studien*, Vol. 11/1980, pp. 1–12, reprinted in his 2004, *Problems of Rationality*, Oxford University Press, Oxford, pp. 151–66.

------. 1982a, ‘Paradoxes of Rationality’, originally published in Wollheim, R. and J. Hopkins (eds.) 1982, *Philosophical Essays on Freud*, Cambridge University Press, Cambridge, pp. 289–305, reprinted in his 2004, *Problems of Rationality*, Oxford University Press, Oxford, pp. 169–87.

------. 1982b, ‘Rational Animals’, originally published in *Dialectica*, vol. 36/1982, pp. 317–27, reprinted in his 2001c, *Subjective, Intersubjective, Objective*, Oxford University Press, Oxford, pp. 95–105.

------. 1985, ‘Reply to Quine on Events’, in Lepore, E. and B. McLaughlin (eds.) 1985, *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford, pp. 172–6.

------. 1986, ‘A Coherence Theory of Truth and Knowledge’, originally published in Lepore, E. (ed.) 1986, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, Oxford, pp. 307–19, reprinted in his 2001c, *Subjective, Intersubjective, Objective*, Oxford University Press, Oxford, pp. 137–53.

------. 2001a, *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford.

------. 2001b, *Inquiries into Truth and Interpretation*, 2nd ed., Oxford University Press, Oxford.

------. 2001c, *Subjective, Intersubjective, Objective*, Oxford University Press, Oxford.

------. 2001d, ‘What Thought Requires’, originally published in Branquinho, J. (ed.) 2001, *The Foundations of Cognitive Science*, Oxford University Press, Oxford, pp. 121–32, reprinted in his 2004, *Problems of Rationality*, Oxford University Press, Oxford, pp. 135–49.

------. 2004, *Problems of Rationality*, Oxford University Press, Oxford.

Davis, W. 1984, ‘A Causal Theory of Intending’, *American Philosophical Quarterly*, Vol. 21, pp. 43–54.

De Caro, M. and D. Macarthur (eds.) 2004, *Naturalism in Question*, Harvard University Press, Cambridge (MA).

DeRose, K. and T.A. Warfield (eds.) 1999, *Scepticism: A Contemporary Reader*, Oxford University Press, Oxford.

Descartes, R. 1641, *Meditations on First Philosophy*, reprinted in Ariew, R. and E. Watkins (eds.) 1998, *Modern Philosophy: An Anthology of Primary Sources*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 22–55.

Donnellan, K.S. 1963, ‘Knowing What I am Doing’, *The Journal of Philosophy*, Vol. 60, No. 14/1963, pp. 401–9.

Dorr, C.S. 2002, ‘Non–Cognitivism and Wishful Thinking’, *Noûs*, Vol. 36/2002, pp. 97–103.

------. 2004, ‘Non–symmetric relations’, in Zimmerman, D. (ed.) 2004, *Oxford Studies in Metaphysics: Volume 1*, Oxford University Press, Oxford, pp. 155–92.

Dostoevsky, F. 1868-9/2001, *The Idiot*, Vintage Publishing, New York.

Dreier, J. 1997, ‘Humean Doubts about the Practical Justification of Morality’, in Cullity, G. and B. Gaut (eds.) 1997, *Ethics and Practical Reason*, Clarendon Press, Oxford, pp. 81–100.

Dretske, F. 1971, ‘Conclusive Reasons’, *Australasian Journal of Philosophy*, Vol. 49/1971, pp. 1–22.

------. 1981, *Knowledge and the Flow of Information*, MIT Press, Cambridge (MA).

------. 1985, 'Précis of Knowledge and the Flow of Information', in Kornblith, H. (ed.), 1985, *Naturalizing Epistemology*, MIT Press, Cambridge (MA).

Dreyfus, H.L. 2005, 'Overcoming the Myth of the Mental: How Philosophers can Profit from the Phenomenology of Everyday Expertise', *Proceedings and Addresses of the American Philosophical Association*, vol. 79, 2/2005, pp. 47–65.

------. 2007a, 'Detachement, Involvement, and Rationality: Are We Essentially Rational Animals?', *Human Affairs*, Vol. 17/2007, pp. 101–9.

------. 2007b, 'The Return of the Myth of the Mental', *Inquiry*, Vol. 50, No. 4/2007, pp. 352–65.

Dummett, M. 1978, 'What do I Know when I Know when I Know a Language?', originally presented at *the Centenary Celebrations*, Stockholm University, reprinted in his 1993, *The Seas of Language*, Oxford University Press, Oxford, pp. 94–105.

------. 1991, *The Logical Basis of Metaphysics*, Harvard University Press, Cambridge (MA).

------. 1993, *The Seas of Language*, Oxford University Press, Oxford.

Dutton, D.G. and A.P. Aron 1974, 'Some Evidence for Heightened Sexual Attraction under Conditions of High Anxiety', *Journal of Personality and Social Psychology*, Vol. 30/1974, pp. 510–17.

Engel, P. 2003, 'What can we learn from Psychology about the Nature of Knowledge?', *Rivista di Filisofia Neoscolatica*.

------. *forthcoming*, 'Taking Seriously Knowledge as a Mental State'.

Everson, S. 2009, 'What are Reasons for Action?', in Sandis, C. (ed.) 2009, *New Essays on the Explanation of Action*, Palgrave Macmillan, London, pp. 22–47.

Falvey, K. 2000, 'Knowledge in Intention', *Philosophical Studies*, Vol. 99/2000, pp. 21–44.

Fara, M. 2008, 'Masked Abilities and Compatibilism', *Mind*, Vol. 117, pp. 843–65.

Fine, K. 1994, 'Essence and Modality', *Philosophical Perspectives*, Vol. 8/1994, pp. 1–16.

------. 2000, 'Semantics for the Logic of Essence', *Journal of Philosophical Logic*, Vol. 29/2000, pp. 543–84.

Firth, R. 1964, 'Coherence, Certainty, and Epistemic Priority', *The Journal of Philosophy*, Vol. LXI, No. 19/1964, pp. 545–57.

Fodor, J. 1981, *Representations: Philosophical Essays on the Foundation of Cognitive Science*, The MIT Press, Cambridge (MA).

------. 1998, *Concepts: Where Cognitive Science Went Wrong*, Clarendon Press, Oxford.

Foot, P. 1958–9, 'Moral Beliefs', originally published in *Proceedings of the Aristotelian Society*, Vol. 59/1958-9, pp. 83–104, reprinted in her 1978, *Virtues and Vices*, Basil Blackwell, Oxford, pp. 110–31.

------. (ed.) 1967, *Theories of Ethics*, Oxford University Press, Oxford.

------. 1978, *Virtues and Vices*, Basil Blackwell, Oxford.

------. 1994, 'Rationality and Virtue', originally published in Pauer–Studer, H. (ed.) 1994, *Norms, Value and Society*, Vienna Circle Institute Yearbook, Kluwer Academic Publishers, reprinted in her 2002, *Moral Dilemmas*, Oxford University Press, Oxford, pp. 159–74.

------. 2001, *Natural Goodness*, Oxford University Press, Oxford.

------. 2002, *Moral Dilemmas*, Oxford University Press, Oxford.

Ford, A., J. Hornsby and F. Stoutland (eds.) 2011, *Essays on Anscombe's 'Intention'*, Harvard University Press, Cambridge (MA).

Foster, L. and J.W. Swanson (eds.) 1970, *Experience and Theory*, The University of Massachusetts Press/Duckworth.

Frankfurt, H.G. 1971, 'Freedom of the Will and the Concept of a Person', originally published in *Journal of Philosophy*, Vol. LXVIII, No. 1/1971, reprinted in his 1998, *The Importance of What We Care About*, Cambridge University Press, New York, pp. 11–25.

------. 1976, 'Identification and Externality', originally published in Rorty, A. (ed.) 1977, *The Identities of Persons*, University of California Press, Los Angeles/Berkeley, reprinted in his 1998, *The Importance of What We Care About*, Cambridge University Press, New York, pp. 58–68.

------. 1978, 'The Problem of Action', originally published in *American Philosophical Quarterly*, 15/1978, reprinted in his 1998b, *The Importance of What We Care About*, Cambridge University Press, New York, pp. 69–79.

------. 1982, 'The Importance of What We Care About', originally published in *Synthese*, Vol. 53, No. 2/1982, pp. 257–72, reprinted in his 1998b, *The Importance of What We Care About*, Cambridge University Press, New York, pp. 80–94.

------. 1987, 'Identification and Wholeheartedness', originally published in Schoeman, F.D. (ed.) 1987, *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, Cambridge University Press, Cambridge, pp. 27–45, reprinted in his 1998b, *The Importance of What We Care About*, Cambridge University Press, New York, pp. 159–76.

------. 1992, 'The Faintest Passion', *Proceedings and Addresses of the American Philosophical Association*, Vol. 66, No. 3/1992, pp. 5–16.

------. 1998a, 'Rationality and the Unthinkable', in his 1998b, *The Importance of What We Care About*, Cambridge University Press, New York, pp. 177–90.

------. 1998b, *The Importance of What We Care About*, Cambridge University Press, New York.

Frege, G. 1892a, 'Comments on *Sinn and Bedeutung*', reprinted in Beaney, M. (ed.) 1997, *The Frege Reader*, Blackwell Publishing, Oxford, pp.172–80.

------. 1892b, 'On Concept and Object', originally published in *Viereljahrsschrift für wissenschaftliche Philosophi*, Vol. 16/1892, pp. 192–205, reprinted in Beaney, M. (ed.) 1997, *The Frege Reader*, Blackwell Publishing, Oxford, pp.181–93.

Freud, S. 1904, *Psychopathology of Everyday Life*, originally published in 1904, reprinted in Brill, A.A. (ed.) 1995, *The Basic Writings of Sigmund Freud*, The Modern Library, New York, pp. 1–146.

Fricker, E. 1998, 'Self–Knowledge: Special Access versus Artefact of Grammar – A Dichotomy Rejected', in Macdonald, C., B. Smith and C. Wright (eds.) 1998, *Knowing Our Own Minds*, Oxford University Press, Oxford, pp. 155–206.

------. 2003, 'Understanding and Knowledge of What is Said', in Barber, A. (ed.) 2003, *Epistemology of Language*, Oxford University Press, Oxford, pp. 324–66.

Gauthier, D. 1986, *Morals by Agreement*, Oxford University Press, Oxford.

Geach, M. and L. Gormally (eds.) 2005, *Human Life, Action and Ethics: Essays by G.E.M. Anscombe*, Imprint Academic, Exeter.

Geach, P. 1956, 'Good and Evil', originally published in *Analysis*, Vol. 17/1956, pp. 32–42, reprinted in Foot, P. (ed.) 1967, *Theories of Ethics*, Oxford University Press, Oxford, pp. 64–73.

------. 1957, *Mental Acts: Their Content and Their Objects*, Routledge and Kegan Paul, London.

Gettier, E. 1963, 'Is Justified True Belief Knowledge?', originally published in *Analysis*, 1963, pp. 121–3, reprinted in Kim, J. and E. Sosa (eds.) 2000, *Epistemology: An Anthology*, Blackwell Publishing, Oxford/Malden (MA), pp. 58–9.

Gibbard, A. 1990, *Wise Choices, Apt Feelings: A Theory of Normative Judgement*, Harvard University Press, Cambridge (MA).

Gibbons, J. 2001, 'Knowledge in Action', *Philosophy and Phenomenological Research*, Vol. LXII, No. 3/2001, pp. 579–600.

Ginet, C. and S. Shoemaker (eds.) 1983, *Knowledge and Mind*, Oxford University Press, Oxford

Ginet, C. 1990, *On Action*, Cambridge University Press, New York.

Goldman, A.I. 1967, 'A Causal Theory of Knowing', *Journal of Philosophy*, Vol. 64, No. 12/1967, pp. 357–72.

------. 1970, *A Theory of Human Action*, Princeton University Press, Princeton.

------. 1975, 'Innate Knowledge', in Stich, S. (ed.) 1975, *Innate Ideas*, University or California Press, Berkeley (CA).

------. 1986, *Epistemology and Cognition*, Harvard University Press, Cambridge (MA).

Grice, P. 1961, 'The Causal Theory of Perception', *Proceedings of the Aristotelian Society*, Vol. 35/1961, pp. 121–53.

------. 1971, 'Intention and Uncertainty', *Proceedings of the British Academy*, Vol. 57/1971, pp. 263–79.

------. 1991, *Studies in the Way of Words*, Harvard University Press, Cambridge (MA).

------. 2001, *Aspects of Reason*, Oxford University Press, Oxford.

Griffin, J. 1996, *Value Judgement: Improving Our Ethical Beliefs*, Clarendon Press, Oxford.

Grunbaum, T. 2009, 'Anscombe and Practical Knowledge of What is Happening', *Philosophische Studien*, Vol. 78/2009, pp. 41–68.

Gunderson, K. (ed.) 1975, *Language, Mind, and Knowledge*, University of Minnesota Press.

Habermas, J. 1994, *Justification and Application*, MIT Press, Cambridge (MA).

Haddock, A. 2005, 'At One with our Actions, but at Two with our Bodies: Hornsby's Account of Action', *Philosophical Explorations*, Vol. 8, 2/2005, pp. 157–72.

------. 2010, 'Knowledge and Action', in Haddock, A., A. Millar and D. Pritchard 2010, *The Nature and Value of Knowledge: Three Investigations*, Oxford University Press, Oxford, pp. 191–259.

Haddock, A., A. Millar and D. Pritchard 2010, *The Nature and Value of Knowledge: Three Investigations*, Oxford University Press, Oxford.

Haddock, A. and F. Macpherson (eds.) 2008, *Disjunctivism: Perception, Action and Knowledge*, Oxford University Press, Oxford.

Hampshire, S. 1960, *Thought and Action*, Chatto and Windus, London.

Hamsphire, S. and H.L.A. Hart 1958, 'Decision, Intention and Certainty', *Mind – New Series*, Vol. 67, No. 265/1958, pp. 1–12.

Harman, G. 1976, 'Practical Reasoning', *Review of Metaphysics*, Vol. XXIX/1976.

------. 1986, *Change in View*, MIT Press, Cambridge (MA).

Harrison, R. (ed.) 1980, *Rational Action*, Cambridge University Press, Cambridge.

Hawkins, J. 2008, 'Desiring the Bad Under the Guise of the Good', *The Philosophical Quarterly*, Vol. 58, No. 231/2008, Blackwell Publishing, Oxford, pp. 244–64.

Hawthorne, J. 2004, *Knowledge and Lotteries*, Oxford University Press, Oxford.

Heidegger, M. 1947, 'Letter on Humanism', reprinted in his 2008, *Basic Writings*, revised and expanded ed. (ed. Krell, D.F.), Harper Perennial, New York, pp. 217–65.

------. 2008, *Basic Writings*, revised and expanded ed. (ed. Krell, D.F.), Harper Perennial, New York.

Heil, J. 1983, 'Doxastic Agency', *Philosophical Studies*, Vol. 43, pp. 355–64.

------. 2004, 'Properties and Powers', in Zimmerman, D. (ed.) 2004, *Oxford Studies in Metaphysics: Volume 1*, Oxford University Press, Oxford, pp. 223–54.

Heinaman, R. (ed.) 1995, *Aristotle and Moral Realism*, UCL Press, London.

Helm, P. 1998, 'John Calvin, the *Sensus Divinitatis*, and the Noetic Effects of Sin', *International Journal for Philosophy of Religion*, Vol. 43, pp. 87–107.

Hieronymi, P. 2006, 'Controlling Attitudes', *Pacific Philosophical Quarterly*, Vol. 87/2006, pp. 45–74.

Hilgard, E.R. 1965, *Hypnotic Susceptibility*, Harcourt Brace Jovanovich, New York.

Hintikka, M.B. and B. Vermazen (eds.) 1985, *Essays on Davidson: Actions and Events*, Oxford University Press, Oxford.

Hobbes, T. 1651/1985, *Leviathan*, (Macpherson, C.B. (ed.)), Penguin Books, London.

Holton, R. 2008, 'Partial Belief, Partial Intention', *Mind*, Vol. 117/2008, pp. 27–58.

Horgan, T. and M. Timmons 2000, 'Nondescriptivist Cognitivism: Framework for a New Metaethic', *Philosophical Papers*, Vol. 29/2000, pp. 121–53.

Hornsby, J. 1980, *Action*, Routledge & Kegan Paul, London/Boston/Henley.

------. 2004a, 'Agency and Action', in Hyman, J. and H. Steward (eds.) 2004, *Agency and Action*, Cambridge University Press, Cambridge, pp. 1–23.

------. 2004b, 'Agency and Alienation', in De Caro, M. and D. Macarthur (eds.) 2004, *Naturalism in Question*, Harvard University Press, Cambridge (MA), pp. 173–184.

------. 2005, 'Semantic Knowledge and Practical Knowledge', *Aristotelian Society Supplementary*, Vol. 79/1, pp. 107–30.

------. 2008, 'A Disjunctive Conception of Acting for Reasons', in Haddock, A. and F. Macpherson (eds.) 2008, *Disjunctivism: Perception, Action and Knowledge*, Oxford University Press, Oxford, pp. 244–61.

------. *forthcoming*, 'The Standard Story of Action: An Exchange 2', in Aguilar, J.H. and A.A. Buckareff (eds.) forthcoming, *Causing Human Action: New Perspectives on the Causal Theory of Action*, Bradford Book/The MIT Press, Cambridge.

Hossack, K. 2000, 'Plurals and Complexes', *British Journal for Philosophy of Science*, Vol. 51, pp. 411–43.

------. 2003, 'Consciousness in Act and Action', *Phenomenology and the Cognitive Science*, Vol. 2/2003, pp. 187–203.

------. 2007, *The Metaphysics of Knowledge*, Oxford University Press, Oxford.

Humberstone, I.L. 1992, *Mind*, Vol. 101, No. 401/1992, pp. 59–83.

Hume, D. 1739–40/2000, *A Treatise of Human Nature*, (Norton, D.F. and M.J. Norton (eds.)), Oxford University Press, Oxford.

------. 1751/1998, *An Enquiry Concerning the Principles of Morals*, (Beauchamp, T.L. (ed.)), Oxford University Press, Oxford.

Hurley, S.L. 2001, 'Overintellectualizing the Mind', *Philosophy and Phenomenological Research*, Vol. LXIII, No. 2/2001, pp. 423–31.

Hursthouse, R. 1991, 'Arational Actions', *The Journal of Philosophy*, Vol. 88, No. 2/1991, pp. 57–68.

------. 1995, 'The Virtuous Agent's Reasons: A Reply to Bernard Williams', in Heinaman, R. (ed.) 1995, *Aristotle and Moral Realism*, UCL Press, London, pp. 24–34.

------. 1999, *On Virtue Ethics*, Oxford University Press, Oxford.

Hursthouse, R., G. Lawrence and W. Quinn (eds.) 1995, *Virtues and Reasons: Philippa Foot and Moral Theory*, Clarendon Press.

Hyman, J. 1999, 'How Knowledge Works', *The Philosophical Quarterly*, Vol. 49, No. 197/1999, pp. 433–51.

Hyman, J. and H. Steward (eds.) 2004, *Agency and Action*, Cambridge University Press, Cambridge.

Irwin, T.H. 1980, 'Reason and Responsibility in Aristotle', in Rorty, A.O. (ed.) 1980, *Essays on Aristotle's Ethics*, University of California Press, Berkeley (C), pp. 117–55.

Jackson, F. 1996, 'Mental Causation', *Mind*, Vol. 105, 419/1996, pp. 377–413.

Jackson, F. and M. Smith (eds.) 2005, *The Oxford Handbook of Contemporary Philosophy*, Oxford University Press, Oxford.

James, H. 1907-9/2009, *The Portrait of a Lady*, revised ed., Oxford University Press, Oxford.

James, W. 1892, *The Will to Believe and Other Essays in Popular Philosophy*, reprinted in Myers, G.E. (ed.) 1992, *William James: Writings 1878–1899*, The Library of America, Cambridge (MA), pp. 445–704.

Johnson, W.E. 1921, *Logic: Part I*, Cambridge University Press, Cambridge.

Johnston, M. 1992, 'How to Speak of the Colors', *Philosophical Studies*, Vol. 68, pp. 221–63.

Jones, O.R. 1983, 'Trying', *Mind* , Vol. 92/1983, pp. 368–85.

Kalderon, M.E. 2005, *Moral Fictionalism*, Clarendon Press, Oxford.

Kant, I. 1781/1787/1998, *Critique of Pure Reason*, (Guyer, P. and A.W. Wood (eds. and trs.)), Cambridge University Press, Cambridge.

------. 1785/1996, *Groundwork in the Metaphysics of Morals*, in his 1996, *Practical Philosophy*, (Guyer, P. and A.W. Wood (eds.)), Cambridge University Press, Cambridge, pp. 37–108.

------. 1996, *Practical Philosophy*, (Gregor, M.J. (eds. and tr.)), Cambridge University Press, Cambridge.

Kanwisher, N. and J. Duncan (eds.) 2004, *Attention and Performance, Vol. 20: Functional Neuroimaging of Visual Cognition*, Oxford University Press, Oxford.

Kazak, S., G.M. Collis and V. Lewis 1997, 'Can Young People with Autism Refer to Knowledge States? Evidence from Their Understanding of 'Know' and 'Guess'', *Journal of Child Psychology and Psychiatry*, Vol. 38/1997, pp. 1001–9.

Kiefer, H. and M. Munitz (eds.) 1970, *Language, Belief and Metaphysics*, State University of New York Press, Albany.

Kierkegaard, S. 1968, *Concluding Unscientific Postscripts*, (D.F. Swenson and W. Lowrie (tr.)), Princeton University Press, Princeton.

Kim, J. and E. Sosa (eds.) 2000, *Epistemology: An Anthology*, Blackwell Publishing, Oxford/Malden (MA).

King, J. 2007, *The Nature and Structure of Content*, Oxford University Press, Oxford.

Kinzler, K.D. and E. Spelke 2007, 'Core Knowledge', *Developmental Science*, Vol. 10, No. 1/2007, pp. 89–96.

Knobe, J. 2003, 'Intentional Action in Folk Psychology: An Experimental Investigation', *Philosophical Psychology*, Vol. 16/2003, pp. 309–24.

------. 2007, 'Acting Intentionally and Acting for a Reason', *Journal of Theoretical and Philosophical Pscyhology*, Vol. 27, No. 1/2007, pp. 119–22.

Knobe, J. and S. Kelly 2009, 'Act for a Reason without Acting Intentionally?', in Sandis, C. (ed.) 2009, *New Essays on the Explanation of Action*, Palgrave Macmillan, London, pp. 169–202.

Kornblith, H. 1982, 'The Psychological Turn', *Australasian Journal of Philosophy*, Vol. 60, No. 3, pp. 238–53.

------. (ed.) 1985, *Naturalizing Epistemology*, MIT Press, Cambridge (MA).

Korsgaard, C.M. 1986, 'Scepticism about Practical Reason', *The Journal of Philosophy*, Vol. 83, No. 1/1986, pp. 5–25.

------. 1996, *The Sources of Normativity*, Cambridge University Press, Cambridge.

------. 1997, 'The Normativity of Instrumental Reason', originally published in Cullity, G. and B. Gaut (eds.) 1997, *Ethics and Practical Reason*, Clarendon Press, Oxford, pp. 213–54, reprinted in her 2008, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford University Press, Oxford, pp. 27–68.

------. 2003, 'Realism and Constructivism in Twentieth–Century Moral Philosophy', originally published in *The Journal of Philosophical Research*, APA Centennial Suppl. ed. 2003, Charlottesvielle (V), pp. 99–122, reprinted in her 2008b, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford University Press, Oxford, pp. 302–26.

------. 2008a, 'Acting for a Reason', in her 2008b, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford University Press, Oxford, pp. 207–29.

------. 2008b, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford University Press, Oxford.

------. 2009, *Self-Constitution: Agency, Identity, and Integrity*, Oxford University Press, Oxford.

Langton, R. 2003, 'Intention as Faith', in Hyman, J. and H. Steward (eds.) 2004, *Agency and Action*, Cambridge University Press, Cambridge, pp. 243–58.

Leite, A. 2005, 'On Williamson's Argument that Knowledge is a Mental State', *Ratio*, Vol. 18, No. 2/2005, pp. 165–75.

Lehrer, K. (ed.) 1966, *Freedom and Determinism*, Random House, New York.

Lepore, E. (ed.) 1986, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, Oxford.

Lepore, E. and B. McLaughlin (eds.) 1985, *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford.

Lepore, E. and B.C. Smith (eds.) 2006, *The Oxford Handbook of Philosophy of Language*, Oxford University Press, Oxford.

Lewis, D. 1979, 'Attitudes *de dicto* and *de se*', *Philosophical Review*, Vol. 88/1979, pp. 513–43.

Lindgaard, J. (ed.) 2008, *John McDowell: Experience, Norm, and Nature*, Blackwell Publishing, Oxford.

Longino, H. 2002, *The Fate of* Knowledge, Princeton University Press, Princeton and Oxford.

Lovibond, S. and S. Williams (eds.) 1996, *Identity, Truth, and Value: Essays for David Wiggins*, Blackwell Publishing, Oxford.

Lowe, E.J. 2008, *Personal Agency: The Metaphysics of Mind and Action*, Oxford University Press, Oxford.

Ludwig, K., 2006, 'Donald Davidson: *Essays on Actions and Events*', in *Classics of Western Philosophy: The Twentieth Century: Quine and After*, Vol. 5/2006, Acumen Press, pp. 146–65.

Macdonald, C. 1998, 'Externalism and Authoritative Self-Knowledge', in Macdonald, C., B. Smith and C. Wright (eds.) 1998, *Knowing Our Own Minds*, Oxford University Press, Oxford, pp. 123–54.

Macdonald, C., B. Smith and C. Wright (eds.) 1998, *Knowing Our Own Minds*, Oxford University Press, Oxford.

MacFarlane, J. 2007, 'Relativism and Disagreement', *Philosophical Studies*, Vol. 132, Springer Netherlands, pp. 17–31.

Mackie, J.L. 1977, *Ethics: Inventing Right and Wrong*, Penguin Books, London.

Martinich, A.P. (ed.) 2001, *The Philosophy of Language*, 4th ed., Oxford University Press, Oxford.

McCann, H. 1998a, *The Works of Agency: On Human Action, Mind, and Freedom*, Cornell University Press, Ithaca (NY).

———. 1998b, 'Trying, Paralysis and Volition', reprinted in his *The Works of Agency: On Human Action, Mind, and Freedom*, Cornell University Press, Ithaca (NY), pp. 94–109.

McCulloch, G. 1995, *The Mind and its World*, Routledge, London.

McDowell, J. 1979, 'Virtue and Reason', originally published in *The Monist* 62, 1979, pp. 331–50, reprinted in his 1998b, *Mind, Value, and Reality*, Harvard University Press, Cambridge (MA), pp. 50–73.

------. 1980, 'The Role of *Eudaimonia* in Aristotle's Ethics', originally published in *Proceedings of the African Classical Associations* 15, 1980, pp. 1–14, reprinted in his 1998b, *Mind, Value, and Reality*, Harvard University Press, Cambridge (MA), pp. 3–22.

------. 1983, 'Aesthetic Value, Objectivity, and the Fabric of the World', originally published in Schaper, E. (ed.) 1983, *Pleasure, Preference, and Value*, Cambridge University Press, Cambridge, pp. 1–16, reprinted in his 1998b, *Mind, Value, and Reality*, Harvard University Press, Cambridge (MA), pp. 112–30.

------. 1994, *Mind and World*, Harvard University Press, London/Cambridge (MA).

------. 1995a, 'Knowledge and the Internal', originally published in, *Philosophy and Phenomenological Research*, Vol. 55/1995, pp. 877–93, reprinted in his 1998a, *Meaning, Knowledge, and Reality*, Harvard University Press, Cambridge (MA), pp. 395–413.

------. 1995b, 'Might there be External Reasons?', in Altham, J.E.J. and R. Harrison (eds.) 1995, *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, Cambridge University Press, Cambridge, pp. 68–85.

------. 1996, 'Incontinence and Practical Wisdom in Aristotle', originally published in Lovibond, S. and S. Williams (eds.) 1996, *Identity, Truth, and Value: Essays for David Wiggins*, Blackwell Publishing, Oxford, pp. 95–112, reprinted in his 2009b, *The Engaged Intellect: Philosophical Essays*, Harvard University Press, Cambridge (MA), pp. 59–76.

------. 1998a, *Meaning, Knowledge, and Reality*, Harvard University Press, Cambridge (MA).

------. 1998b, *Mind, Value, and Reality*, Harvard University Press, Cambridge (MA).

------. 1998c, ‘Responses to Crispin Wright’, in Macdonald, C., B. Smith and C. Wright (eds.) 1998, *Knowing Our Own Minds*, Oxford University Press, Oxford, pp. 47–62.

------. 1998d, ‘Sellars on Perceptual Experience’, originally published in *Journal of Philosophy*, Vol. 95/1998, pp. 431–91, reprinted in his 2009a, *Having the World in View: Essays on Kant, Hegel, and Sellars*, Harvard University Press, Cambridge (MA), pp. 3–22.

------. 1998e, ‘Some Issues in Aristotle’s Moral Psychology’, in his 1998a, *Mind, Value, and Reality*, Harvard University Press, Cambridge (MA), pp. 23–49.

------. 2007a, ‘Responses to Dreyfus’, originally published in *Inquiry*, Vol. 50/2007, pp. 366–70, reprinted in his 2009b, *The Engaged Intellect: Philosophical Essays*, Harvard University Press, Cambridge (MA), pp. 324–8.

------. 2007b, ‘What Myth?’, originally published in *Inquiry*, Vol. 50/2007, pp. 338–51, reprinted in his 2009b, *The Engaged Intellect: Philosophical Essays*, Harvard University Press, Cambridge (MA), pp. 308–23.

------. 2008a, ‘Avoiding the Myth of the Given’, originally published in Lindgaard, J. (ed.) 2008, *John McDowell: Experience, Norm, and Nature*, Blackwell Publishing, Oxford, pp. 1–14, reprinted in his 2009a, *Having the World in View: Essays on Kant, Hegel, and Sellars*, Harvard University Press, Cambridge (MA), pp. 256–72.

------. 2008b, ‘Intention in Action’, Howison Lecture Series, [online] University of California, Berkeley. Available at: http://grad.berkeley.edu/lectures/event.php?id=490&lecturer=280

------. 2008c, ‘Towards a Reading on Hegel on Action in the “Reason” Chapter of the *Phenomenology*’, originally published (in German) in Vieweg, K. and W. Welsch (eds.) 2008, *Hegels Phänomenologie des Geistes—Ein kooperativer Kommentar zu einem Schlüsselwerk der Moderne*, Suhrkamp Verlag, Frankfurt a.m., reprinted in his 2009a, *Having the World in View: Essays on Kant, Hegel, and Sellars*, Harvard University Press, Cambridge (MA), pp. 166–84.

------. 2009a, *Having the World in View: Essays on Kant, Hegel, and Sellars*, Harvard University Press, Cambridge (MA).

------. 2009b, *The Engaged Intellect: Philosophical Essays*, Harvard University Press, Cambridge (MA).

------. 2010, ‘What is the Content of an Intention in Action?’, *Ratio*, Vol. XXIII/2010, pp. 415–32.

Mele, A. 1992, *Springs of Action*, Oxford University Press, Oxford.

------. 2003, *Motivation and Agency*, Oxford University Press, Oxford.

Mellor, D.H. (ed.) 1990, *Philosophical Papers: F.P. Ramsey*, Cambridge University Press, Cambridge.

------. 1995, *The Facts of Causation*, Routledge, London.

Millar, A. 2007, ‘The State of Knowing’, *Philosophical Issues*, Vol. 17/2007, pp. 179–96.

Milligan, D.E. 1974, ‘Reasons as Explanations’, *Mind*, Vol. 83, pp. 180–93.

Moore, G.E. 1962, *Commonplace Books*, George Allen & Unwin Ltd., London.

------. 1966, *Ethics*, 2nd ed., Oxford University Press, London.

Moran, R. 2004, ‘Anscombe on “Practical Knowledge”’, in Hyman, J. and H. Steward (eds.) 2004, *Agency and Action*, Cambridge University Press, Cambridge, pp. 43–68.

Murdoch, I. 1970, *The Sovereignty of Good*, Routledge, New York.

Myers, G.E. (ed.) 1992, *William James: Writings 1878–1899*, The Library of America, Cambridge (MA).

Nagel, T. 1970, *The Possibility of Altruism*, Princeton University Press, Princeton.

------. 1986, *The View from Nowhere*, Oxford University Press, Oxford.

Newman, L. 2010, 'Descartes' Epistemology', *Stanford Encyclopedia of Philosophy* [online], Stanford University. Available at: http://plato.stanford.edu/entries/descartes–epistemology/ [20.06.2010]

Nisbett, R. and T.D. Wilson 1977, 'Telling More than We Know: Verbal Reports on Mental Processes', *Psychological Review*, Vol. 84, pp. 231–59.

Nisbett, R. and L. Ross 1980, *Human Inference: Strategies and Shortcomings of Social Judgement*, Englewood Cliffs (NJ).

Nozick, R. 1981, *Philosophical Explanations*, Harvard University Press, Cambridge (MA).

Nussbaum, M.C. 2001, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, revised ed., Cambridge University Press, Cambridge.

O'Brien, L. 2003, 'On Knowing One's Own Actions', in Roessler, J. and N. Eilan (eds.) 2003, *Agency and Self–Awareness*, Oxford University Press, Oxford, pp. 358–82.

------. 2007, *Self–Knowing Agents*, Oxford University Press, Oxford.

O'Shaughnessy, B. 1963, 'Observation and The Will', *The Journal of Philosophy*, Vol. 60, No. 14/1963, pp. 367–92.

------. 2003, 'The Epistemology of Physical Action', in Roessler, J. and N. Eilan (eds.) 2003, *Agency and Self–Awareness*, Oxford University Press, Oxford, pp. 345–57.

------. 2008a, *The Will: A Dual Aspect Theory – Volume 1*, 2nd ed., Cambridge University Press, Cambridge.

------. 2008b, *The Will: A Dual Aspect Theory – Volume 2*, 2nd ed., Cambridge University Press, Cambridge.

Parfit, D. 1984, *Reasons and Persons*, Oxford University Press, Oxford.

Pascal, B. 1670/1998, 'The Wager', reprinted in Ariew, R. and E. Watkins (eds.) 1998, *Modern Philosophy: An Anthology of Primary Sources*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 94–6.

Pauer–Studer, H. (ed.) 1994, *Norms, Value and Society*, Vienna Circle Institute Yearbook, Kluwer Academic Publishers.

Paul, S.K. 2009a, 'How We Know What We're Doing', *Philosopher's Imprint*, Vol. 9, No. 4/2009.

------. 2009b, 'Intention, Belief and Wishful Thinking: Setiya on "Practical Knowledge"', *Ethics*, Vol. 119/2009, pp. 546–57.

Peacocke, C. 1992, *A Study of Concepts*, MIT Press, Cambridge (MA).

------. 1998, 'Conscious Attitudes and Self–Knowledge', in Macdonald, C., B. Smith and C. Wright (eds.) 1998, *Knowing Our Own Minds*, Oxford University Press, Oxford, pp. 63–98.

Pears, D.F. 1985, 'Intention and Belief', in Hintikka, M.B. and B. Vermazen (eds.) 1985, *Essays on Davidson: Actions and Events*, Oxford University Press, Oxford, pp. 75–88.

Perner, J. 1991, *Understanding the Representational Mind*, MIT Press, Cambridge (MA).

Pieper, J. 1966, *The Four Cardinal Virtues*, University of Notre Dame Press, Notre Dame (IN).

Plantinga, A. 1983, 'Reason and Belief in God', in his and N. Wolterstorff (eds.) 1983, *Faith and Philosophy*, University of Notre Dame Press, London.

------. 2000, *Warranted Christian Belief*, Oxford University Press, Oxford.

Plantinga, A. and N. Wolterstorff (eds.) 1983, *Faith and Philosophy*, University of Notre Dame Press, London.

Plato, 1997a, *Gorgias*, in Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 791–869.

------. 1997b, *Laws*, in Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 1318–1616.

------. 1997c, *Meno*, in Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 870–897.

------. 1997d, *Republic*, in Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 971–1223.

------. 1997e, *Sophist*, in Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 235–93.

------. 1997f, *Statesman*, in Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 294–358.

------. 1997g, *Theaetetus*, in Cooper, J.M. (ed.) 1997, *Plato: Complete Works*, Hackett Publishing Company, Indianapolis/Cambridge, pp. 157–234.

Pryor, J. 2005, 'There is Immediate Justification', in Steup, M. and E. Sosa (eds.) 2005, *Contemporary Debates in Epistemology*, Blackwell, Oxford, pp. 181–202.

Putnam, H. 1973, 'Meaning and Reference', originally published in *Journal of Philosophy*, Vol. 70/1973, pp. 699–711, reprinted in Martinich, A.P. (ed.) 2001, *The Philosophy of Language*, 4th ed., Oxford University Press, Oxford, pp. 288–95.

------. 1975, 'The Meaning of "Meaning"', originally published in Gunderson, K. (ed.) 1975, *Language, Mind, and Knowledge*, University of Minnesota Press, pp. 131–93, reprinted in Chalmers, D.J. (ed.) 2002, *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, Oxford, pp. 581–96.

------. 1981a, 'Brains in a Vat', originally published in his 1981b, *Reason, Truth and History*, Cambridge University Press, Cambridge, pp. 1–21, reprinted in DeRose, K. and T.A. Warfield (eds.) 1999, *Scepticism: A Contemporary Reader*, Oxford University Press, Oxford, pp. 27–42.

------. 1981b, *Reason, Truth and History*, Cambridge University Press, Cambridge.

Quine, W.V.O. 1960, *Word and Object*, MIT Press, Cambridge (MA).

------. 1985, 'Events and Reification', in Lepore, E. and B. McLaughlin (eds.) 1985, *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford, pp. 162–71.

Quinn, W. 1992, 'Rationality and the Human Good', originally published in *Social Philosophy and Policy* Vol. 9, No.2/1992, pp. 81–95, reprinted in his 1993a, *Morality and Action*, Cambridge University Press, Cambridge, pp. 210–27.

------. 1993a, *Morality and Action*, Cambridge University Press, Cambridge.

------. 1993b, 'Putting Rationality in Its Place', in his 1993a, *Morality and Action*, Cambridge University Press, Cambridge, pp. 228–55.

Rachels, J. 2003, *The Elements of Moral Philosophy*, McGraw Hill, Boston.

Radford, C. 1966, 'Knowledge – by Examples', *Analysis*, Vol. 27/1966, pp. 1–11.

------. 1988, 'Radford Revisiting', *The Philosophical Quarterly*, Vol. 38, No. 153/1988, pp. 496–99.

Ramsey, F.P. 1925, 'Universals', originally published in *Mind*, Vol. 34/1925, pp. 8–30, reprinted in Mellor, D.H. (ed.) 1990, *Philosophical Papers: F.P. Ramsey*, Cambridge University Press, Cambridge, pp. 8–30.

Raz, J. 1999, *Engaging Reason: On the Theory of Value and Action*, Oxford University Press, Oxford.

-------. 2010, ‘On the Guise of the Good’, in Tenenbaum, S. (ed.) 2010, *Desire, Practical Reason, and the Good*, Oxford University Press, Oxford, pp. 111–37.

Rescher, N. (ed.) 1967, *The Logic of Decision and Action*, University of Pittsburgh Press, Pittsburgh.

Rescher, N. and D. Reidel (eds.) 1969, *Essays in Honor of Carl G. Hempel*, Reidel Publishing Company, Dordrecht.

Reisner, A. 2007, ‘Evidentialism and the Numbers Game’, *Theoria*, Vol. 73, No. 4/2007, pp. 304–16.

------. *forthcoming*, ‘Leaps of Knowledge’.

Roessler, J. and N. Eilan (eds.) 2003, *Agency and Self–Awareness*, Oxford University Press, Oxford.

Rorty, A.O. (ed.) 1977, *The Identities of Persons*, University of California Press, Berkeley (C).

------. (ed.) 1980, *Essays on Aristotle’s Ethics*, University of California Press, Berkeley (C).

------. 1980, ‘Where Does the Akratic Break Take Place?’, originally published in *The Australian Journal of Philosophy*, Vol. 58/1980, pp. 333–36, reprinted in her 1988, *Mind in Action: Essays in the Philosophy of Mind*, Beacon Press, Boston, pp. 229–45.

------. 1981, ‘Akrasia and Conflict’, originally published in *Inquiry*, Vol. 23/1981, pp. 1–22, reprinted in her 1988, *Mind in Action: Essays in the Philosophy of Mind*, Beacon Press, Boston, pp. 246–65..

------. 1988, *Mind in Action: Essays in the Philosophy of Mind*, Beacon Press, Boston.

Rorty, R. 1999, *Philosophy and Social Hope*, Penguin Books, London.

Roush, S. 2005, *Tracking Truth: Knowledge, Evidence and Science*, Oxford University Press, Oxford.

Rödl, S. 2010, ‘The Form of the Will’, in Tenenbaum, S. (ed.) 2010, *Desire, Practical Reason, and the Good*, Oxford University Press, Oxford, pp. 138–60.

Russell, B. 1918, ‘The Philosophy of Logical Atomism’, reprinted in his 1956, *Logic and Knowledge*, (ed. R. Marsh), Georg Allen & Unwin, London, pp. 177–281.

------. 1956, *Logic and Knowledge*, (ed. R. Marsh), Georg Allen & Unwin, London.

Ryle, G. 1949/2002, *The Concept of Mind*, New University of Chicago Press, Chicago.

Sandis, C. (ed.) 2009, *New Essays on the Explanation of Action*, Palgrave Macmillan, London.

Sax, G. 2010, ‘Having Know-How: Intellect, Action, and Recent Work on Ryle’s Distinction between Knowledge-How and Knowledge-That’, *Pacific Philosophical Quarterly*, Vol. 91, No. 4, pp. 507–30.

Scanlon, T.M. 1998, *What We Owe to Each Other*, The Belknap Press/Harvard University Press, Lodnon/Cambridge (MA).

Schaper, E. (ed.) 1983, *Pleasure, Preference, and Value*, Cambridge University Press, Cambridge.

Schoeman, F.D. (ed.) 1987, *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, Cambridge University Press, Cambridge.

Schopenhauer, A. 1859/1969, *The World as Will and Representation: Volume 1*, (Payne, E.F.J. (tr.)), Dover Publications, Mineola (NY).

------. 1840/2010, *The Two Fundamental Problems of Ethics*, (Cartwright, D.E. and E. Erdman (tr.)), Oxford University Press, Oxford.

------. 1877a/1903, *On the Fourfold Root of the Principle of Sufficient Reason*, 4th ed., reprinted in his 1903, *Two Essays by Arthur Schopenhauer*, revised ed., (Hillebrand, K. (tr.)), George Bell and Sons, London, pp. 1–189.

------. 1877a/1903, *The Will in Nature*, 4th ed., reprinted in his 1903, *Two Essays by Arthur Schopenhauer*, revised ed., (Hillebrand, K. (tr.)), George Bell and Sons, London, pp. 215–380.

------. 1903, *Two Essays by Arthur Schopenhauer*, revised ed., (Hillebrand, K. (tr.)), George Bell and Sons, London.

Schueler, G.F. 2003, *Reasons & Purposes: Human Rationality and the Teleological Explanation of Action*, Oxford University Press, Oxford.

Schwitzgebel, E. 2010, 'Belief', *Stanford Encyclopedia of Philosophy* [online], Stanford University, Berkeley. Available at: http://www.seop.leeds.ac.uk/entries/belief/#4 [21.11.2010]

Searle, J. 1983, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge.

Sellars, W. 1966, 'Thought and Action', in Lehrer, K. (ed.) 1966, *Freedom and Determinism*, Random House, New York, pp. 105–39.

------. 1991, *Science, Perception and Reality*, Ridgeview Publishing Company, Atascadero (C).

Setiya, K. 2003, 'Explaining Action', *The Philosophical Review* Vol. 112, No. 3/2003, pp. 339–93.

------. 2004, 'Hume on Practical Reason', *Philosophical Perspectives*, Vol. 18/2004, pp. 365–89.

------. 2007a, 'Cognitivism about Instrumental Reason', *Ethics*, Vol. 117/2007, pp. 649–73.

------. 2007b, *Reasons without Rationalism*, Princeton University Press, Princeton.

------. 2008a, 'Believing at Will', *Midwest Studies in Philosophy*, Vol. 32/2008, pp. 36–52.

------. 2008b, 'Practical Knowledge', *Ethics*, Vol. 118/2008, The University of Chicago Press, Chicago, pp. 388–409.

------. 2009a, 'Practical Knowledge Revisited', *Ethics*, Vol. 120/2009, pp. 128–37.

------. 2009b, 'Reply to Bratman and Smith', *Analysis Reviews*, Vol. 69, No. 3/2009, Oxford University Press, Oxford, pp. 531–40.

------. 2010a, 'Intention', *Stanford Encyclopedia of Philosophy* [online], Stanford University. Available at: http://plato.stanford.edu/entries/intention/ [05.01.2010]

------. 2010b, 'Sympathy for the Devil', in Tenenbaum, S. (ed.) 2010, *Desire, Practical Reason, and the Good*, Oxford University Press, Oxford, pp. 82–110.

------. 2011, 'Knowledge of Intention', in Ford, A., J. Hornsby and F. Stoutland (eds.) 2011, *Essays on Anscombe's 'Intention'*, Harvard University Press.

Shah, N. 2008, 'How Action Governs Intention', *Philosophers' Imprint*, Vol. 8, No. 5/2008, pp. 1–19.

Shah, N. and D.J. Velleman 2005, 'Doxastic Deliberation', *The Philosophical Review*, Vol. 114, No. 4/2005, pp. 497–534.

Shakespeare, W. 1975, *Williams Shakespeare: The Complete Works*, Random House/Grammery Books, New York.

Shope, R.K. 1983, *The Analysis of Knowing: A Decade of Research*, Princeton University of Press, Princeton.

Shoemaker, S. 1980, 'Causality and Properties', originally published in van Inwagen, P. (ed.) 1980, *Time and Cause*, D. Reidel Publishing Company, Dordrecht, pp. 109–35, reprinted in his 2003, *Identity, Cause, and Mind: Philosophical Essays*, expanded ed., Oxford University Press, Oxford, pp. 206–33.

------. 1998, 'Causal and Metaphysical Necessity', originally published in *Pacific Philosophical Quarterly*, Vol. 79/1998, pp. 59–77, reprinted in his , *Identity, Cause, and Mind: Philosophical Essays*, expanded ed., Oxford University Press, Oxford, pp. 405–26.

------. 2003, *Identity, Cause, and Mind: Philosophical Essays*, expanded ed., Oxford University Press, Oxford.

Smith, B.C. 2006, 'What I Know when I Know a Language', in Lepore, E. and B.C. Smith (eds.) 2006, *The Oxford Handbook of Philosophy of Language*, Oxford University Press, Oxford, pp. 941–982.

Smith, M. 1987, 'The Humean Theory of Motivation', *Mind*, Vol. 96/1987, pp. 36–61.

------. 1994, *The Moral Problem*, Blackwell Publishing, Malden (MA).

------. 2009, 'The Explanatory Role of Being Rational', in Sobel, D. and S. Wall (eds.) 2009, *Reasons for Action*, Cambridge University Press, Cambridge, pp. 58–80.

------. *forthcoming* a, 'The Standard Story of Action: An Exchange', in Aguilar, J.H. and A.A. Buckareff (eds.) forthcoming, *Causing Human Action: New Perspectives on the Causal Theory of Action*, Bradford Book/The MIT Press, Cambridge.

------. *forthcoming* b, 'Four Objections to the Standard Story of Action (and Four Replies)', *Philosophical Issues*.

Snowdon, P. 2003, 'Knowing How and Knowing That: A Distinction Reconsidered', *Proceedings of the Aristotelian Society*, Vol. 104, pp. 1–29.

Sobel, D. and S. Wall (eds.) 2009, *Reasons for Action*, Cambridge University Press, Cambridge.

Spelke, E.S. 2000a, 'Core Knowledge', *American Psychologist*, Vol. 55/2000, pp. 1233–43.

------. 2000b, 'Core Knowledge and Cognitive Development', *The Pinkel Lectures 2000: Second Annual Lecture* [online], Institute for Research in Cognitive Science: University of Pennsylvania. Available at: http://www.ircs.upenn.edu/pinkel/lectures/spelke/index.shtml [21.01.2000]

------. 2004, 'Core Knowledge', in Kanwisher, N. and J. Duncan (eds.) 2004, *Attention and Performance, Vol. 20: Functional Neuroimaging of Visual Cognition*, Oxford University Press, Oxford.

Spinoza, B. 1992a, *Ethics*, reprinted in Feldman, S. (ed.) 1992, *Baruch Spinoza: Ethics, Treatise on the Emendation of the Intellect and Selected Letters*, (tr. Shirley, S.), Hackett Publishing, Indianapolis/Cambridge, pp. 31–223.

------. 1992b, 'Letter 58', reprinted in Feldman, S. (ed.) 1992, *Baruch Spinoza: Ethics, Treatise on the Emendation of the Intellect and Selected Letters*, (tr. Shirley, S.), Hackett Publishing, Indianapolis/Cambridge, pp. 285–87.

Stampe, D. 1987, 'The Authority of Desire', *The Philosophical Review*, Vol. 96, pp. 335–81.

Stanley, J. 2010, 'Knowing (How)', *Noûs*, pp. 1–32.

Stanley, J. and T. Williamson 2001, 'Knowing How', *The Journal of Philosophy*, Vol. 98, No. 8/2001, pp. 411–44.

Steglich-Petersen, A. 2005, 'Williamson on Knowledge, Action, and Causation', *Nordic Journal of Philosophy*, Vol. 6, 1/2005, Philosophia Press, pp. 15–28.

Steup, M. 2006, 'The Analysis of Knowledge', *Stanford Encyclopedia of Philosophy* [online], Stanford University. Available at: http://plato.stanford.edu/entries/knowledge–analysis [16.01.2006]

Steup, M. and E. Sosa (eds.) 2005, *Contemporary Debates in Epistemology*, Blackwell, Oxford.

Stich, S. (ed.) 1975, *Innate Ideas*, University or California Press, Berkeley (CA).

------. 1978, 'Autonomous Psychology and the Belief–Desire Thesis', *The Monist*, Vol. 61/1978, pp. 573–91.

Stocker, M. 1976, 'The Schizophrenia of Modern Ethical Theories', originally published in *Journal of Philosophy*, Vol. 73/1976, pp. 453–66, reprinted in Crisp, R. and M. Slote (eds.) 1997, *Virtue Ethics*, Oxford University Press, Oxford, pp. 66–78.

.

------. 1979, ‘Desiring the Bad: An Essay in Moral Psychology’, *The Journal of Philosophy* Vol. 76, No. 12/1979, pp. 738–53.

------. 2004, ‘Raz on the Intelligibility of Bad Acts’, in Wallace, R.J., P. Pettit, S. Scheffler, and M. Smith (eds.) 2004, *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Clarendon Press, Oxford, pp. 303–32.

Strawson, P.F. 1950, ‘Truth’, originally published in *Proceedings of the Aristotelian Society*, supplementary Vol. 24, reprinted in his 1971, *Logico–Linguistic Papers*, Methuen, London, pp. 191–213.

------. 1954, ‘Wittgentstein’s *Philosophical Investigations*’, originally published in *Mind*, Vol. LXIII/1954, reprinted in his 1974/2008, *Freedom and Resentment and Other Essays*, Routledge, London/New York, pp. 147–85.

------. 1971, *Logico–Linguistic Papers*, Methuen, London.

------. 1974/2008, *Freedom and Resentment and Other Essays*, Routledge, London/New York.

Stroud, S. and C. Tappolet (eds.) 2003, *Weakness of Will and Practical Irrationality*, Oxford University Press, Oxford.

Szabó, Z.G. 2004, ‘On the Progressive and the Perfective’, *Noûs*, Vol. 38, No. 1/2004, pp. 29–59.

Teichman, R. 2008, *The Philosophy of Elizabeth Anscombe*, Oxford University Press, Oxford.

Tenenbaum, S. 1999, ‘The Judgement of a Weak Will’, *Philosophy and Phenomenological Research* Vol. LIX, No. 4/1999, pp. 875–911.

------. 2007, *Appearances of the Good: An Essay on the Nature of Practical Reason*, Cambridge University Press, Cambridge.

------. (ed.) 2010, *Desire, Practical Reason, and the Good*, Oxford University Press, Oxford.

Travis, C. 2005, ‘A Sense of Occasion’, originally published in *Philosophical Quarterly*, Vol. 55, No. 219/2005, pp. 286–314, reprinted in his 2008, *Occasion-Sensitivity*, Oxford University Press, Oxford, pp. 290–315.

------. 2008, *Occasion-Sensitivity*, Oxford University Press, Oxford.

Thompson, M. 2008, *Life and Action: Elementary Structures of Practice and Practical Thought*, Harvard University Press, Cambridge (MA).

------. 2011, ‘Anscombe’s *Intention* and Practical Knowledge’, in Ford, A., J. Hornsby and F. Stoutland (eds.) 2011, *Essays on Anscombe’s* Intention, Harvard University Press, Cambridge (MA).

Unger, P. 1968, ‘An Analysis of Factual Knowledge’, *Journal of Philosophy*, Vol. 65, pp. 157–70.

------. 1974, ‘An Argument for Skepticism’, originally published in *Philosophical Exchange*, Vol. 1, No. 4/1974, reprinted in Kim, J. and E. Sosa (eds.) 2000, *Epistemology: An Anthology*, Blackwell Publishing, Oxford/Malden (MA), pp. 42–52.

------. 1975/2002, *Ignorance: A Case for Scepticism*, Oxford University Press, Oxford.

Van Inwagen, P. (ed.) 1980, *Time and Cause*, D. Reidel Publishing Company, Dordrecht.

Velleman, D.J. 1985, ‘Practical Reflection’, *Philosophical Review*, Vol. 94/1985, pp. 33–61.

------. 1992a, ‘The Guise of the Good’, originally published in *Nous* 26, 1992, pp. 3–26, reprinted in his 2000b, *The Possibility of Practical Reason*, The Scholarly Publishing Office, Ann Arbor, pp. 99–122.

------. 1992b, ‘What Happens When Someone Acts?’, originally published in *Mind*, 1992, pp. 461–81, reprinted in his 2000b, *The Possibility of Practical Reason*, The Scholarly Publishing Office, Ann Arbor, pp. 123–43.

------. 1996, ‘The Possibility of Practical Reason’, originally published in *Ethics* 106, 1996, pp. 694–726, reprinted in his 2000b, *The Possibility of Practical Reason*, The Scholarly Publishing Office, Ann Arbor, pp. 170–99.

------. 2000a, ‘From Self Psychology to Moral Philosophy’, originally published in *Philosophical Perspectives*, Vol. 14/2000, pp. 349–77, reprinted in his 2006, *From Self to Self*, Cambridge University Press, New York, pp. 224–52.

------. 2000b, *The Possibility of Practical Reason*, The Scholarly Publishing Office, Ann Arbor.

------. 2006, *Self to Self: Selected Essays*, Cambridge University Press, New York.

------. 2007, *Practical Reflection*, CSLI Publications, Stanford (CA).

------. 2009, *How We Get Along*, Cambridge University Press, Cambridge.

Vieweg, K. and W. Welsch (eds.) 2008, *Hegels Phänomenologie des Geistes—Ein kooperativer Kommentar zu einem Schlüsselwerk der Moderne*, Suhrkamp Verlag, Frankfurt (a.m.).

Wallace, R.J. 1999, ‘Three Conceptions of Rational Agency’, originally published in *Ethical Theory and Moral Practice*, Vol. 2/1999, pp. 217–42, reprinted in his 2006, *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford University Press, Oxford, pp. 43–62.

------. 2001, ‘Normativity, Commitment, and Instrumental Reason’, originally published in *Philosopher’s Imprint*, Vol. 1, No.3/2001, reprinted in his 2006, *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford University Press, Oxford, pp. 82–111.

------. 2004, ‘Normativity and the Will’, originally published in Hyman, J. and H. Steward (eds.) 2004, *Agency and Action*, Cambridge University Press, Cambridge, pp. 195–216, reprinted in his 2006, *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford University Press, Oxford, pp. 71–81.

------. 2006a, *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford University Press, Oxford.

------. 2006b, ‘Postscript to Normativity, Commitment, and Instrumental Reason’, in his 2006, *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford University Press, Oxford, pp. 111–20.

------. 2008, ‘Practical Reason’, *Stanford Encyclopedia of Philosophy* [online], Stanford University. Available at: http://plato.stanford.edu/entries/practical–reason/ [06.11.2008]

Wallace, R.J., P. Pettit, S. Scheffler, and M. Smith (eds.) 2004, *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Clarendon Press, Oxford.

Watson, G. 1975, ‘Free Agency’, *The Journal of Philosophy* Vol. 72, No. 8/1975, pp. 205–20.

------. 1977, ‘Skepticism about Weakness of Will’, *The Philosophical Review* Vol. 86, No. 3/1977, pp. 316–39.

------. 1982, ‘Review of ‘Actions’ by Jennifer Hornsby’, *Journal of Philosophy*, Vol. 79, No. 8/1982, pp. 464–69.

------. 1987, ‘Free Action and Free Will’, *Mind*, Vol. 96/1987, pp. 154–72.

------. 2003, 'The Work of the Will', in Stroud, S. and C. Tappolet (eds.) 2003, *Weakness of Will and Practical Irrationality*, Oxford University Press, Oxford, pp. 201–29.

Wedgwood, R. 2009, *The Nature of Normativity*, Clarendon Press, Oxford.

Wellman, H.M., D. Cross, and J. Watson 2001, 'Meta–analysis of theory of mind development: The truth about false belief', *Child Development*, Vol. 72, pp. 655–84.

Wiggins, D. 1978-9, 'Weakness, Commensurability, Desire', originally published in *Proceedings of the Aristotelian Society*, Vol. 79/1978-9, reprinted in his 1998, *Needs, Values, Truth*, 3rd ed., Oxford University Press, Oxford, pp. 239–67.

------. 1998, *Needs, Values, Truth*, 3rd ed., Oxford University Press, Oxford.

Wikforss, Å.M. 2004, 'Direct Knowledge and Other Minds', *Theoria*, Vol. LXX, part 2–3/2004, pp. 271–93.

Wiland, E. 2007, 'Intentional Action and "In Order To"', *Journal of Theoretical and Philosophical Psychology*, Vol. 27, pp. 113–8.

Williams, B. 1970, 'Deciding to Believe', originally published in Kiefer, H. and M. Munitz (eds.) 1970, *Language, Belief and Metaphysics*, State University of New York Press, Albany, reprinted in his 1973, *Problems of the Self*, Cambridge University Press, Cambridge, pp. 136–51.

------. 1973, *Problems of the Self*, Cambridge University Press, Cambridge.

------. 1980, 'Internal and External Reasons', originally published in Harrison, R. (ed.) 1980, *Rational Action*, Cambridge University Press, Cambridge, pp. 17–28, reprinted in his 1981, *Moral Luck: Philosophical Papers 1973–1980*, Cambridge University Press, Cambridge, pp. 101–13.

------. 1981, *Moral Luck: Philosophical Papers 1973–1980*, Cambridge University Press, Cambridge.

------. 1985, *Ethics and the Limits of Philosophy*, Harvard University Press, Cambridge (MA).

------. 1989, 'Internal Reasons and the Obscurity of Blame', originally published in *Logos*, Vol. 10/1989, Santa Clare University, CA, reprinted in his 1995b, *Making Sense of Humanity and Other Philosophical Papers 1982–1993*, Cambridge University Press, Cambridge, pp. 35–45.

------. 1995a, 'Acting as the Virtuous Person Acts', in Heinaman, R. (ed.) 1995, *Aristotle and Moral Realism*, UCL Press, London, pp. 13–23.

------. 1995b, *Making Sense of Humanity and Other Philosophical Papers 1982–1993*, Cambridge University Press, Cambridge.

------. 1995c, 'Replies', in Altham, J.E.J. and R. Harrison (eds.) 1995, *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, Cambridge University Press, Cambridge, pp.185–224.

------, 2008, *Shame and Necessity*, University of California Press, Berkeley/Los Angeles.

Williamson, T. 1995, 'Is Knowing a Mental State?', *Mind*, Vol. 101, pp. 533–65.

------. 2000, *Knowledge and Its Limits*, Oxford University Press, Oxford.

------. 2005, 'Knowledge and Scepticism', in Jackson, F. and M. Smith (eds.) 2005, *The Oxford Handbook of Contemporary Philosophy*, Oxford University Press, Oxford, pp. 681–700.

------. 2006, 'Can Cognition be Factorised into Internal and External Components?' in Abrahamsen, A. et.al. (eds.) 2006, *Contemporary Debates in Cognitive Science*, Blackwell Publishing, Oxford, pp. 291–306.

------, 2007, *The Philosophy of Philosophy*, Blackwell Publishing, Malden (MA).

Wilson, J.C. 1926, *Statement and Inference*, Oxford University Press, Oxford.

Wilson, J.M. 2010, 'What is Hume's Dictum and Why Believe it?', *Philosophy and Phenomenological Research*, Vol. 80, 3/2010, pp. 595–637.

Wittgenstein, L. 1953/2001, *Philosophical Investigations*, 3rd ed., Anscombe, E. (tr.), Blackwell Publishing, Oxford.

------. 1969/2009, *On Certainty*, reprinted in his 2009, *Major Works: Selected Philosophical Writings*, HarperCollins Publishers, New York, pp. 319–439.

------. 2009, *Major Works: Selected Philosophical Writings*, HarperCollins Publishers, New York.

Wolf, S. 1982, 'Moral Saints', originally published in *Journal of Philosophy*, Vol. 79/1982, pp. 419–39, reprinted in Crisp, R. and M. Slote (eds.) 1997, *Virtue Ethics*, Oxford University Press, Oxford, pp. 79–98.

------. 1990, *Freedom within Reason*, Oxford University Press, Oxford.

Woods, M. 1972, 'Reasons for Acting and Desire', *Proceedings of the Aristotelian Society* Supplementary Volume, pp. 189–201.

Wollheim, R. and J. Hopkins (eds.) 1982, *Philosophical Essays on Freud*, Cambridge University Press, Cambridge.

Yovel, Y. And D. Reidel (eds.) 1978, *Philosophy of History and Action*, The Magnes Press, The Hebrew University.

Zimmerman, D. (ed.) 2004, *Oxford Studies in Metaphysics: Volume 1*, Oxford University Press, Oxford.