

**REASON'S SELF-ACTUALIZATION: AN ESSAY ON SELF-CONSCIOUSNESS AND
RATIONAL AGENCY**

by

Joshua Stuchlik

B.A., Philosophy, University of Notre Dame, 2004

M.A., University of Pittsburgh, 2006

Submitted to the Graduate Faculty of
The School of Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH
SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Joshua Stuchlik

It was defended on

September 9, 2010

and approved by

Robert Brandom, Distinguished Professor of Philosophy

Kieran Setiya, Associate Professor of Philosophy

Peter Machamer, Professor of the History and Philosophy of Science

Dissertation Advisor: John McDowell, Distinguished University Professor of Philosophy

Copyright © by Joshua Stuchlik

2011

REASON'S SELF-ACTUALIZATION: AN ESSAY ON SELF-CONSCIOUSNESS AND RATIONAL AGENCY

Joshua Stuchlik, PhD

University of Pittsburgh, 2011

In my dissertation I show that we cannot conceive of ourselves as embodied beings unless we know some of our physical features without observation or inference. I also argue that we have the requisite sort of self-knowledge, and that it consists in our knowledge of ourselves as intentional *agents*.

Descartes claimed that when one is self-consciously aware of oneself, one is aware of oneself as a purely psychological being. In chapter two I argue that if his claim were correct, it would be unclear what it could mean for one to identify oneself with a human being. I then argue that self-conscious beliefs about oneself are beliefs about oneself that are not grounded on observation or inference.

In chapter three I take up the task of making it plausible that we do possess the required sort of self-knowledge. I offer a novel interpretation of Anscombe's thesis that we know what we are doing intentionally without observation or inference. The key lies in the Aristotelian doctrine that action itself can be the conclusion of practical reasoning.

In chapter four I reply to two objections to my account. The first is an argument for *volitionalism*, or the thesis that events that are describable as an agent's moving her body are acts of trying that occur prior to her bodily movements. In response, I argue for an alternative, according to which bodily action is a temporally extended process that is complete only when one's body has moved. The second argument begins from the premise that we can act intentionally without knowing

that we are succeeding. I argue that this shows only that our self-conscious capacity to act intentionally is fallible in a certain respect. Conditions which potentially inhibit the success of one's doing such-and-such intentionally also inhibit one's capacity to know that one is doing so when the action is successful.

Finally, in chapter five I defend a non-reductionistic account of intentional action in contrast to dominant reductionistic models. I conclude that an intentional action is simply an exercise of a rational agent's will, described as such.

TABLE OF CONTENTS

PREFACE.....	VIII
1.0 INTRODUCTION.....	1
2.0 THE PROBLEM OF BODILY SELF-KNOWLEDGE	20
2.1 TWO RIVAL WAYS OF CONCEIVING OF ONESELF	22
2.2 DESCARTES ON SELF-CONSCIOUSNESS	25
2.3 THE PROBLEM OF BODILY SELF-KNOWLEDGE.....	30
2.4 ANSCOMBE ON KNOWLEDGE IN INTENTION AND SELF- REFERENCE	36
3.0 PRACTICAL KNOWLEDGE AND PRACTICAL REASONING	43
3.1 GRICE'S OBJECTION, AND THE OUTLINE OF A REPLY	45
3.2 DAVIDSON ON PRACTICAL REASONING.....	56
3.3 ANSCOMBE AND THE FORM OF PRACTICAL REASONING	62
3.4 DE-MOTIVATING THE INFERENTIAL THEORY.....	71
3.5 COMPARISON WITH COGNITIVISM.....	75
4.0 TRYING, BODILY AGENCY, AND THE LIMITS OF PRACTICAL KNOWLEDGE	81
4.1 THE ARGUMENT FOR VOLITIONALISM.....	83
4.2 FAILURE AND FALLIBILITY	92

4.3	A DUAL ASPECT THEORY	98
4.4	FAILURE TO KNOW WHAT ONE IS DOING	101
4.5	THE WILL AS A FALLIBLE EPISTEMIC POWER	106
5.0	REDUCTIONISM AND THE METAPHYSICS OF AGENCY	112
5.1	WHERE'S THE AGENT?	114
5.2	VELLEMAN AND FRANKFURT	120
5.3	REDUCTIONISM AND NATURALISM	128
5.4	A NON-REDUCTIONISTIC ACCOUNT OF INTENTIONAL ACTION	135
APPENDIX A		140
A.1 SELF-CONSCIOUSNESS, PERCEPTION, AND RECEPTIVITY		142
A.2 IS BODILY AWARENESS RECEPTIVE KNOWLEDGE?		145
A.3 A FAMILY OF NON-RECEPTIVE EPISTEMIC POWERS		149
BIBLIOGRAPHY		152

PREFACE

I arrived in Pittsburgh in the belief I would be writing a dissertation on ethics. I quickly realized, though, that many of the great moral philosophers of the past either prefaced their ethical theories with an account of the will or else implicitly presupposed one. So Aquinas begins his ethical theory in the *Summa theologiae* with treatises on happiness, the object of the will, and on action; Hume only turns to moral philosophy in Book III of the *Treatise* after providing an account of the passions in Book II, under which heading he includes the will; and Kant derives the categorical imperative in *Groundwork* II from an understanding of rational agents as beings who act in accordance with their conception of a law. The very different shapes of these philosophers' moral theories can be traced back in large part to their varying accounts of the nature and function of the will. On reflection this is not so extraordinary, for if ethics seeks to guide us in deliberating about how to act, it must do so within constraints set by assumptions about what deliberate action *is*. Once I became interested in intentional action as a topic, however, I discovered that it comes with a host of metaphysical and epistemological questions of its own. At the same time, I came to believe that a satisfactory account of rational agency goes hand in hand with a certain conception of our embodied nature, which can function as an important corrective to distortions of it that are sometimes found in contemporary moral philosophy. This dissertation is the result of my work on this cluster of issues.

Along the way I have incurred debts to many people. In the first place I would like to express my sincere gratitude to my director, John McDowell. It was reading his work that motivated me to pursue my graduate studies at the University of Pittsburgh. *Mind and World* changed my conception of what philosophy can be. This dissertation would not have been possible without his guidance and kindness during seminars, directed studies, and countless conversations. His influence is no doubt discernable on every page. I cannot imagine having a more careful or generous advisor.

I am also grateful to the other members of my dissertation committee. All of them have read portions of drafts and provided helpful feedback. I give special thanks to Kieran Setiya, who read through earlier versions of every chapter, sometimes multiple times, and whose critical remarks pushed me to further think through and clarify my thoughts on many of their central arguments.

Beyond my committee I have benefited from discussions and debates with friends and colleagues. I mention in particular two works that appeared during the course of my dissertation writing that have exerted an enormous influence on the way I see the issues I herein address: Sebastian Rödl's *Self-Consciousness* and Michael Thompson's *Life and Action*. My main partners in philosophical dialogue have been Justin Shaddock, Nicholas Joshua Teh, Stephen Makin, Jennifer Frey, Stephen Puryear, and Michael Darcy, C.O. I would also like to thank my first philosophical mentor, Michael Loux, who encouraged me to pursue philosophy as a career, and as way of life.

Finally, I thank my family for their continuing love and support. I dedicate this dissertation to my parents, Mark and Roberta Stuchlik.

1.0 INTRODUCTION

Since one's day to day relationships are organized around one's participation in concrete biological and social practices, it is not surprising that in the course of his phenomenological analysis of the category "being-for-others", Sartre remarks on the relation of consciousness to the body. The mind-body problem is a familiar one, but his diagnosis of the source of its apparent intractability is novel. He writes:

The problem of the body and its relations with consciousness is often obscured by the fact that while the body is from the start posited as a certain *thing* having its own laws and capable of being defined from outside, consciousness is then reached by the type of inner intuition which is peculiar to it. Actually if after grasping "my" consciousness in its absolute interiority and by a series of reflective acts, I then seek to unite it with a certain living object composed of a nervous system, a brain, glands, digestive, respiratory, and circulatory organs whose very matter is capable of being analyzed chemically into atoms of hydrogen, carbon, nitrogen, phosphorus, *etc.*, then I am going to encounter insurmountable difficulties. But these difficulties all stem from the fact that I try to unite my consciousness not with *my* body but with the body *of others*. In fact the body which I have just described is not *my* body such as it is *for me*. (1984, 401)

A reliable method for provoking Cartesian intuitions is to juxtapose the properties that one attributes to oneself *qua* reflective, self-conscious knower with those that one attributes to one's body. When Descartes turns his mind's eye upon himself in the *Second Meditation* he finds powers for thought, will, and sensation, and he remarks how different these properties are from those that belong to the chunks of matter that compose the corporeal world (AT VII 27-8/CSM II 18-9). Sartre's response is that Descartes' method subtly prejudices the question

whether his thinking, conscious life is “unified” with a certain body by conceiving that body as that “of others” rather than as *his*. This accusation is cryptic to say the least, but Sartre goes on to elaborate what he intends by it:

Of course during a radioscopy I was able to see the picture of my vertebrae on a screen, but I was outside in the midst of the world. I was apprehending a wholly constituted object as a *this* among other *thises*, and it was only by a reasoning process that I referred it back to being *mine*... (1984, 402)

This passage makes clear that the distinction between my body *qua* body of *another* and my body *qua* my body is an epistemological one. These phrases pick out a single object under the aspect of two different ways in which it is knowable. The paradigmatic way in which I get to know about the properties of my body “as other” is through sense experience, such as seeing my vertebrae on a hospital monitor. When I observe my body, the way it presents itself to me is under a demonstrative mode of presentation, the same way that the body of another person would present itself to me (as “a this among thises”). So while it may be true that when I observe a certain bodily organism or a part thereof I observe what is in fact myself, my consciousness of myself on this basis is not essentially first personal.¹ If I know, say, that the vertebrae I see on a screen are mine, it is because I know that they belong to a certain man who is under examination and that that man is me.

This suggests that I have available to me another way or ways of knowing about my bodily attributes which is or are not grounded in observation. Sartre contends that I can know about my body not only as “a this” but as *mine*. And it would be natural to say that if I could

¹ Sebastian Rödl notes that Aristotle writes in the *Metaphysics* that when a physician heals himself he does so “as other” (1019a15-18). The reason is that it does not belong to the *nature* of healing that the healer be identical with the one healed. Rödl maintains that since it is not part of the nature of sense perception to represent oneself, we can say on similar grounds that perception of oneself represents oneself as other (2007, 8). Sartre appears to be making a similar point in the passages quoted above.

know about it in this way, I could know it not only from the outside, as just another thing “in the midst of the world”, but from within. It is because I know in this way that I am a certain embodied human individual that when I observe that individual’s body I can identify it as mine. I can even learn that it is composed of elements the concepts of which are completely foreign to my pre-scientific consciousness of myself. This would also mean that Descartes goes wrong in his meditations upon himself when he denies he has any essentially first personal of knowing about himself as a bodily creature.

On Sartre’s view we should expect that if one holds Descartes’ view one will not be able to make sense of the identity between oneself and a certain corporeal being. The problem will be that one has no mode of access to one’s bodily attributes that would make it possible for one to identify oneself, as the object that comes into view when one reflects on one’s own nature, with an object one encounters in experience.

But do we actually have first person knowledge of our bodily attributes? We are certainly not aware of all of our parts and organs. The only way we know we have, e.g., a liver or a brain is ultimately by relying on observations of them, whether our own or (more likely) those of doctors and physiologists. But it appears that we do have at least two other ways of knowing about some of our bodily states and activities that are not, or at least not obviously, based on observation. First, we are typically aware of the position and motions of our limbs, and we know when we are in certain somatic states such as being hot or cold, hungry or thirsty, aroused, or uncomfortable. Insofar as we have these powers of bodily awareness we are conscious both of bodily states of ourselves and of the condition of certain interior and exterior regions of our bodies.

Second, we also seem to have a distinctive knowledge of ourselves as intentional *agents*. In her monograph *Intention* G.E.M. Anscombe points out that we normally know what we are doing intentionally without having to observe ourselves acting. She calls such knowledge “knowledge in intention” or “practical knowledge”. One way of making the phenomenon of knowledge in intention salient is by bringing out how disconcerting it is when it is lost. J. David Velleman asks us to imagine such a scenario at the beginning of his book, *Practical Reflection*:

You are walking up Fifth Avenue. All of a sudden you realize that you don’t know what you’re doing. You can see that you’re walking up Fifth Avenue, of course: the surroundings are quite familiar. But the reason why you’re walking up Fifth Avenue escapes you, and so you still don’t know what you’re doing. Are you walking home from work? Trying to catch a downtown bus? Just taking a stroll? (15)

As Velleman acknowledges, it may be possible for us to observe our intentional actions under their more immediate descriptions, such as “walking”. But we normally also know what we are doing intentionally under descriptions like “going home from work” or “catching a bus downtown”. And these descriptions are not knowable solely on the basis of observation of oneself and one’s surroundings. Hence the feeling of alienation that often accompanies the loss of knowledge in intention.

The significance of the special epistemic access we have to our intentional actions for resisting Cartesian claims about one’s nature was recognized by Schopenhauer. He mockingly refers to the knowing subject as Descartes conceives it as a “winged cherub”. However, he does not object to the idea that first person knowledge of oneself is distinct from empirical knowledge. What he rejects is the claim that the only sort of knowledge we have of our bodies is empirical:

To the subject of knowing, who appears as an individual only through his identity with the body, this body is given in two entirely different ways. It is given in perception of the intellect as representation, as an object among other objects. But it is also given in quite a different way, namely as what is known immediately

to everyone and is denoted by the word *will*. The act of will and the action of the body are not two different states objectively known, connected by the bond of causality; they do not stand in the relation of cause and effect, but are one and the same thing, though given in two entirely different ways, first quite directly, and then in perception and understanding. (1966:2, 100)

In this remarkable passage, Schopenhauer does two things. First, he makes an assertion about two different ways we are aware of our bodily actions. One of these ways is through perception: although it is not always possible to tell on the basis of observation that someone is performing a certain bodily intentional action (e.g., catching a bus downtown), such actions are spatiotemporal processes and as they unfold they are in principle observable to well-placed perceivers. According to Schopenhauer rational agents also have another way of knowing what they are doing intentionally, namely, through having willed to do it. In this respect, he is in agreement with Anscombe. But he goes on to provide a metaphysical explanation of the nature of knowledge in intention. Its source lies in the fact that one's acts of will and one's intentional actions are not separate causally related events, but are identical. So Schopenhauer agrees with Descartes that we have a first personal way of knowing about the content of our will. His central move is rather to question the additional claim that an act of will is distinct from the action of the body. The will should not be regarded as a purely psychological power but as a *psychophysical* one.²

The goal of this dissertation is to fill out, more precisely and in detail, the line of reasoning I have just sketched by reference to the writings of Sartre and Schopenhauer. Both

² I believe this passage from Schopenhauer is the clearest expression of a conception of intentional action that was shared by Fichte and Hegel. It would be too much to defend this claim here, but the same idea seems to be present in the heading of Hegel's discussion of intentional action in the *Phenomenology*, after which this dissertation takes its title, and which reads: "The Actualization of Rational Self-Consciousness Through Its Own Activity" (1977, 211). That intentional action is a psychophysical unity has been held more recently by Hornsby (1977) and O'Shaughnessy (1980).

knowledge in intention and bodily awareness serve as starting points for objecting to Descartes' claims about our knowledge of ourselves. However, due to constraints of space my focus will be mainly on the former. My remarks about bodily awareness are confined to a brief appendix.

* * *

I remarked above that when Descartes seeks knowledge of his own nature he does not do so by listing properties he observes about himself, but by reflecting on what he can know about himself immediately and without observation. Why think this method is a good one for acquiring self-understanding? And why think that it somehow renders problematic the relation between oneself and one's body? The purpose of Chapter 2 is to answer these questions. There I show how reflection on one's own nature can seem to generate a problem about how it should be conceived. The difficulty arises when we consider self-consciousness, or the consciousness that one has of oneself *as oneself*. This contrasts with knowledge that one has of oneself under some other specification, such as my knowledge of myself under the name "Joshua Stuchlik". Descartes is clear that when I reflect on myself self-consciously, I am aware of my psychological states and activities but not of any bodily ones. I argue that if he is right about this, then it's hard to see how I could be identical with a certain human being. It is plausible that significant judgments of identity must involve, at least implicitly, a general concept containing a criterion of individuation. Assuming that J.S. is a thing of the kind "human being", then if I am J.S. I am a human being.³ Human beings are material particulars and it is a conceptual truth that such particulars are individuated by their kind and the space they occupy at a time. But if Descartes is right, when I think of myself self-consciously I don't think of myself as occupying any space.

³ It follows from this assumption that Descartes should conclude he is not the man Descartes but rather, e.g., his Ego.

So this truth isn't applicable to me, and it is not clear what it would even mean to say that I'm identical with a certain human being.

In order to evaluate whether Descartes' thesis about self-consciousness is correct, we need to know which of our epistemic capacities are self-conscious ones. A self-conscious capacity is one that yields knowledge of oneself as oneself. Therefore, when a self-conscious capacity provides one with knowledge that a certain property is instantiated, it provides one with knowledge that oneself has the property. In other words, thoughts about oneself formed on the basis of self-conscious capacities have the property Sydney Shoemaker calls "immunity to error through misidentification relative to the first person pronoun".

We can distinguish between "I"-thoughts that are immune to error through misidentification and those that are not. In the *Blue Book* Wittgenstein refers to "I"-thoughts of the first sort as thoughts about oneself in which "I" is used "as subject" and he calls "I"-thoughts of the second sort thoughts in which it is used "as object". I argue that knowledge about oneself acquired on the basis of observation and inference is knowledge of oneself "as object", whereas knowledge of oneself that is not so grounded is knowledge of oneself "as subject". Since in contemporary epistemology the latter sort of knowledge of oneself is typically referred to as "self-knowledge", we can say that our self-conscious capacities are identical with our capacities for self-knowledge.

Earlier I said that Descartes asserts that one's self-conscious knowledge of oneself is restricted to purely mental states and activities, and that this view raises a difficulty for the view that one is a human being. I have now also said that our self-conscious capacities are our capacities for self-knowledge. In Chapter 2 I call this difficulty "the problem of bodily self-

knowledge”. In order to avoid this problem we must show that we have self-knowledge of some of our bodily states or activities, sufficient to individuate us as human beings.

Self-knowledge of the required sort would be knowledge of one’s bodily attributes that one has “as subject” rather than “as object”. This corresponds to Sartre’s idea of knowledge of one’s body *as one’s own*, as opposed to knowledge of one’s body as other. In Chapter 3 I turn to the task of making it plausible that the epistemology of intentional action has the resources for resolving the problem of bodily self-knowledge, or more accurately, for dissolving it, if Anscombe is correct that we typically have knowledge of what we are doing intentionally without observation or inference. The problem therefore only arises if we fail to take into account all of our self-conscious capacities.

While Anscombe’s thesis is *prima facie* plausible it has also been found objectionable. One notable example is H.P. Grice’s objection that beliefs about what one is doing not based on sufficient evidence would be no better than cases of wishful thinking. Anscombe’s conception of knowledge in intention is admittedly such that if she is correct then the beliefs about action it picks out share some features with wishful thought, in particular, with wishful beliefs—beliefs in things we desire, not based on evidence—about spatiotemporal processes. Since Grice thinks that such beliefs could not possibly be warranted he puts forth a different account of knowledge in intention. On his view when we know what we are doing intentionally this knowledge is evidentially grounded in our knowledge of what we intend to do. Like many philosophers, Grice supposes that the will is a self-conscious power, such that we have immediate and non-observational knowledge of the content of our will. I argue that this is a legitimate assumption to make, but making it also opens up a possibility the defender of Anscombe can exploit. For if it is true, as Schopenhauer claims, that intentional bodily action is itself an act of will, then the sort

of knowledge we have of our intentional actions is simply a case of self-conscious knowledge. It is this feature of knowledge in intention that distinguishes it from cases of wishful thinking.

This strategy sets us the task of making a case for the metaphysics of agency endorsed by Schopenhauer. He was not the only figure in the history of philosophy to hold that intentional action is the actualization of a psychophysical power, though. The will is a rational capacity, whose characteristic activity is manifested in reasoning directed at action. And the idea this activity does not stop short of bodily action is also contained in the well-known Aristotelian doctrine that action can be the *conclusion* to a piece of practical reasoning.⁴ My goal in the rest of the third chapter is therefore to elucidate the Aristotelian doctrine. The immediate motivation for doing so is the role it plays in helping resolve the problem of bodily self-knowledge. But I think that when developed the metaphysics of rational agency that emerges is attractive in its own right.

Aristotle holds that practical reasoning begins with a certain good or apparent good that the agent desires (*DA* 433a28-9). Anscombe captures this aspect of the Aristotelian view when she says that an action is intentional when the question “Why?”, in a special sense, is applicable to it. The question is special insofar as it asks for a reason for acting, and the agent provides a reason by giving a desirable characteristic she took the action to possess (1963, §37). Similarly, Joseph Raz describes what he calls the “the classical approach” to intentional action as follows:

[T]ypical intentional actions are actions about which their agents have a story to tell..., a story which explains why one acted as one did. Moreover, and this point is crucial, the explanation makes intelligible not only why the action happened; it

⁴ I am supposing that when Aristotle speaks of practical reasoning he has in his sights the activity of the power modern philosophers refer to as “the will”. It is still occasionally said that the Greeks had no concept of the will, and that it was invented by Augustine or the Judeo-Christian tradition. This does not square with the fact that later philosophers such as Aquinas and Anscombe, who make it explicit that their topic is the will, constantly refer back to Aristotle. For an analysis of the issue, see Irwin 1992.

makes it intelligible as an action chosen, or otherwise undertaken, by the agent. . . According to [the classical approach] the “story” is of what agents took to be facts which show the act to be good, and therefore constitute a reason for its performance, making it eligible. (1999, 24)

The classical approach is not uncontroversial, however, and as it is the starting point of the investigation of practical reasoning in Chapter 3, I will say a word in support of it.⁵ Opponents of it usually argue against it, in part, by claiming that it is subject to counterexamples such as cases of weakness of will, *accidie*, and malicious action (cf. Stocker 1979, Velleman 1992a). But the classical approach has the resources to explain the existence of these phenomena. Consider, for example, weakness of will. The weak-willed agent knows or believes that a certain course of action is the best or most reasonable course of action available to her, yet she forms the intention to perform an incompatible action instead (which can include simply omitting to perform the act she deems best). This would only pose a problem for the classical approach if it were also committed to the claim that rational agents always perform the action they believe to be the most desirable one available. But it is only committed to the weaker claim that when agents act intentionally they take it that their act has some desirable characteristic. So it can easily account for cases in which an agent intentionally performs an act she believes to be sub-optimal.

The opponent of the classical approach may concede that this response is all right as far as it goes but complain that it does not capture all cases of weakness of will. It seems there are cases in which one performs an intentional action one believes has nothing desirable, or even approximately desirable, about it (cf. Setiya, 2010). Suppose someone is deliberating whether to enter into an adulterous relationship with a co-worker. No doubt in certain situations the

⁵ This is not intended to be conclusive. For a fuller defense of the classical approach that is consonant with what I argue here see Boyle and Lavin, 2010.

property of being an act of intercourse is a desirable one. But, we may suppose, this situation is such that an act of intercourse with this co-worker would not be desirable. Again, I think it is a mistake to think that cases like this are incompatible with the classical approach. While on that view an agent must take it that it that an act she intentionally performs is desirable in some respect, the notion of “taking” in question covers more than just believing. It may be that one takes an act to be desirable insofar as it *appears* to have a desirable characteristic, even if one’s considered judgment is that it does not.⁶ Sexual intercourse is the kind of activity that can continue to appear to be desirable even though one knows that doing it in one’s particular situation would not be, and acting on the basis of a tempting appearance is irrational.⁷

I do not think that the truth or falsity of the classical approach can be established by appeal to examples. While I believe its account of the phenomenology of action is more insightful than that of its opponents, the question of interpretation is likely to lead to a stalemate. A better line of defense would proceed from first principles, showing how the classical approach is entailed by the nature of the will. The will has traditionally been defined as “rational appetite”. The concept of appetite is the concept of a power possessed by living things, and it is in that context that the classical approach must be understood. To be a living thing is to be a certain sort of teleologically organized system. It is what Kant calls a *Naturzweck*, a complex

⁶ Aristotle allows for this possibility, writing that when something would be immediately pleasant it triggers an appetite, in the light of which it appears both pleasant without qualification and good without qualification (*DA* 433b1-10). He is here following Plato, who in the *Republic* famously argues from the existence of motivational conflict to a division of motivational faculties in the soul. See also Tenenbaum 2007, ch.1.

⁷ I have concentrated on weakness of will because it is most often put forth as an alleged counterexample to the classical approach. Velleman also challenges that approach on the grounds that it cannot make sense of malicious action, but his objection does not strike me as very compelling. Velleman considers the proclamation of Milton’s Satan, “Evil be thou my good”, and remarks that if Satan takes evil to be desirable then he is no longer malicious but a well-intentioned fool (1992, 119). This is not at all fair to Milton. A well-intentioned fool is someone who wants to do the right thing but fails because he is confused or ignorant of what he means to do. But there is no hint that Satan is confused or ignorant: he perfectly well knows that his actions wrong others. The fact that his deliberately unjust acts are means to ends, such as liberty and glory, that he takes to be desirable does not alter this.

self-organizing system in which the parts, insofar as they possess vital functions, are present because of the way they contribute to the realization of certain ends or purposes (*CJ*, Ak 5:373). And insofar as something is alive, it will be equipped with tendencies towards these ends. These tendencies are what are picked out by the term “appetites”. At the same time for a creature to realize these ends is for it to achieve its good. For example, taking in nourishment from the environment is an end that is shared by plants and animals and when a plant or animal is nourished it achieves something good for it.

The function of appetite, then, is to direct a living thing to ends or goods constitutive of its nature, and the will is the specific shape this power takes in rational creatures. As rational agents, we do not simply pursue goods that are in fact appropriate objects of desire. We pursue things precisely insofar as we apprehend them *as* desirable. It is this conceptual connection between desire and desirability that is captured by the classical approach to intentional action.⁸

Practical reasoning begins with an end that the agent apprehends as desirable and takes him from it to action. In Chapter 3 I examine attempts by Davidson and Anscombe to spell out how this process works. I find Davidson’s theory unsatisfying. He speaks approvingly of the Aristotelian doctrine that practical reasoning can conclude in action, but his own account obscures rather than illuminates it. According to Davidson, practical reasoning is non-deductive reasoning that proceeds from desires to intentions, where desires are conceived as being or corresponding to judgments that certain acts are *prima facie* desirable, and intentions are or correspond to unconditional desirability judgments. The problem with this, at least for purposes

⁸ As Anscombe notes, one can act intentionally for no particular reason (1963, §17). But it is important that in any particular circumstance not just anything can be done idly or on a whim, and what sets boundaries to the intelligibility of such actions is precisely the good or apparent good. A person’s explanation for why, e.g., he is eating a saucer of mud would likely refer to some unusual dietary theory or to a mistaken belief about the nature of what he is consuming, but we would not have been given an adequate explanation of the action if the man reported that he was chomping away on such a disgusting substance “for no particular reason”.

of my project, is that if he is correct it seems practical reasoning actually concludes not in action but in a judgment *that* a certain sort of action is desirable.

What Anscombe says about practical reasoning in *Intention* fares much better on this score. Her key insight lies in her contention that for Aristotle the difference between practical reasoning and theoretical reasoning is not one of content but of *form*. The formal difference between these two species of reasoning is constituted by the different questions which they address and the way they respond to them. Theoretical reasoning addresses the question what to think. The deliberating subject answers this question by arriving at an affirmation of the truth or falsity of a proposition, and her affirmation consists in the formation of a belief. On the other hand, practical reasoning is not addressed to the question what to think but rather *what to do*. An agent practically deliberates when she has an end she finds desirable and must figure out what to do to realize it, and her reasoning concludes by affirming some act within her power. However, this affirmation consists not in judging that doing such-and-such is desirable; rather, at least when the time for acting is here and now, it consists in *doing* such-and-such. The conclusion is the action.

Philosophers often distinguish between two sorts of intention: intention for the future and intention in action (cf. Searle 1983, 84). If an agent forms an intention to perform a certain act at a later time, then when the time of acting arrives it becomes an intention in action. We may use this language if we like, but the present point is that the agent's intention in action is not distinct from the action that is the execution of it. This is simply another way of putting Schopenhauer's remark that in action "the act of the will and the movement of the body" are not related as cause

and effect, but are one.⁹ The whole process of acting is an activity of the self-conscious power of the will.

The Aristotelian doctrine that action can be the conclusion to, rather than merely the effect of, a piece of practical reasoning expresses the idea that for animals like us, rationality is not a feature that is simply tacked onto pre-rational powers for bodily movement that we share with other animals. These powers, which are found in some creatures in a non-rational form, are themselves transformed by the presence of rationality.¹⁰ There are bound to be questions about the viability of this approach. In Chapter 4 I further develop the Aristotelian conception of action I defend by responding to two objections.

It is an all too familiar fact that we sometimes fail to do what we intend. This might be thought to show that the relation between intending and acting is a causal one. For what is the difference between cases in which I fail to execute my intention and those in which I succeed if it is not that in the latter case but not the former my intention causes my body to move? Actually, even in cases of failure I do more than simply form an intention: I also *try* to perform an act of the ill-fated type. But this simply means that we must recast the objection in terms of trying. According to *volitionalism* practical reasoning properly concludes in a psychological act of trying that is distinct from bodily movement and when an agent's trying is successful it causes her body to move. An agent's active contribution to the course of events is her act of trying, and whether this act is successful depends on whether it is causally effective.

Volitionalism is troubling because it conceives of the relation between an agent and her body as a purely instrumental one. Fortunately, as I argue in Chapter 4, there is a way to refuse

⁹ Compare Anscombe's denial that "the relation of *being done in execution of a certain intention, or being done intentionally* is a causal relation between act and intention" (1983, 95; italics in original).

¹⁰ For a similar claim about human perceptual capacities see McDowell 1996, ch. 2.

to go along with the line of reasoning that leads to it. The volitionalist is right both that an agent's activity is constituted by her trying and that it is possible for an agent to try and completely fail to move her body. But the volitionalist goes wrong in thinking it follows that an act of trying and the movement of the body must be two causally related events in the successful case. We should instead conceive of an agent's activity as a single temporally extended process that begins prior to bodily movement but is complete only when the body has moved. It is possible that as this process unfolds it will be impeded even before the body moves, and if this occurs it will be describable only as a failed trying. However, when the process reaches its completion it is describable not only in terms of trying but also in terms of bodily movement.

The key point is that the human will is fallible insofar as it is a power for bodily action. Powers are defined by their proper acts—they are specified as powers to do A where "A" stands for some form of activity. An infallible power results in a completed A-ing each time it is exercised, whereas a fallible power can be prevented or impeded from reaching its full actualization. It issues in a completed A-ing when it is not so impeded.

The will is also a self-conscious power, and this fact might also be thought to render the account of intentional agency I defend problematic. The reason is that rational agents sometimes act intentionally without *knowing* whether their action is successful. All they know is what they intend to do and what they are trying to do. It is tempting to infer from this that acts of the will, which is essentially self-conscious, must be restricted to acts of intending and trying. An assumption along these lines often lies in the background of debates over Anscombe's thesis that it belongs to rational agents to have self-conscious knowledge that extends all the way to action. It explains why, for example, opponents try to refute it by pointing to cases where someone

cannot know she is successfully A-ing because she cannot observe an event whose occurrence would make it the case she is doing A.¹¹

Unlike some of Anscombe's defenders I do not attempt to deny that such cases are genuinely possible. Instead I question the assumption that underlies the thought that they constitute genuine counterexamples to the idea that the will is by nature a self-conscious power of action. Once again the notion of fallibility is crucial. Cases of the relevant sort show only that the will is fallible insofar as it is an epistemic power. In this respect the human will is like the human power of perception, which is also fallible insofar as it is a self-conscious capacity. That these powers as they are present in us are fallible is just what we should expect given that we are *finite* rational agents.¹²

In Chapter 4 I respond to two particular objections to the Aristotelian conception of action. But there are also more general features of contemporary philosophy of mind that make this conception difficult even to comprehend. According to currently dominant reductionistic models of action what it is to be an intentional action is to be a bodily movement that is produced in the right way by certain mental states or events, viz. desires, beliefs, and perhaps, intentions. I

¹¹ As in Paul 2009.

¹² There are other ways in which we can fail to know what we are doing which are not necessary to discuss at any length for the purposes of my project. One sort of case involves actions or aspects of actions that are automatic. Psychological studies have found that some human behavior can be unconsciously triggered by environmental stimuli. For instance, in one study subjects who solved scrambled sentences with words relating to the elderly were found to walk more slowly afterwards than subjects not exposed to such words (Bargh et al. 1996). There are also aspects of skillful actions that are automatic, e.g., people may not know how to describe what they are doing in driving a car at a very detailed level (cf. Vallacher and Wegner 1987). But what is significant about cases of automaticity is that the aspects of automatic behavior of which people are unaware are aspects under which it is not *intentionally* performed. They therefore do not fall under the scope of the kind of practical self-consciousness I discuss. There are also Freudian cases in which a patient is not fully aware of all the things she is intentionally doing because one or more of her motives are repressed. However, it is important that we recognize these cases as pathological ones, disorders of the will. My task is to get clearer about the nature of the power that they are disordered conditions of, and the way to do so is by focusing on its proper, non-disordered acts. The kinds of case I do focus on in Chapter 4 are relevant because they are ones in which agents who are not in a pathological state do not know what they are doing, under a description under which their action is intentional.

have claimed that intentional actions, which sometimes involve bodily happenings, are themselves acts of the will. If we accept a reductionistic model of action this claim is bound to appear unintelligible. On that model action falls squarely on the physical side of the mental/physical divide while acts of will fall on the mental side, which encompasses the causal antecedents of action. The possibility that intentional action is an irreducibly psychophysical process is ruled out from the start. The Aristotelian metaphysics of action goes along with a simple and powerful account of knowledge in intention, but it will not get a fair hearing unless one is willing to consider an alternative to reductionism. The purpose of the fifth and final chapter is therefore to motivate and develop the non-reductionistic framework presupposed in the rest of the dissertation.

I begin by pointing to a serious problem that reductionistic models of action face, namely, that they seem to leave no room for the agent to play a role in the causal order. In his recent paper “What Happens When Someone Acts?” Velleman is sensitive to this problem and attempts to develop a sophisticated version of reductionism to overcome it. He does so by drawing on the work of Harry Frankfurt, who seeks to elucidate what it is for an agent to actively participate, in a certain sense, in her action. However, I argue that Frankfurt’s project is importantly different from Velleman’s. Frankfurt is not concerned with reductionism and the conceptual materials he employs will not be of any help in responding to objections to it. The problem for reductionism seems to be an intractable one, and this raises the question whether it is well-motivated in the first place. Velleman contends that only reductionism is consistent with philosophical naturalism, but I explain why I think he is wrong about this. There is no inherent incompatibility between naturalism and non-reductionism.

The basic commitment of non-reductionism is that action is the exercise of a causal power of a substance, where the causal powers of a substance are metaphysically basic, not reducible to those of entities in other categories such as states and events. A substance is an agent insofar as it exercises one of its causal powers. What makes intentional agency distinctive is that when an agent acts intentionally it exercises a specific kind of causal power, namely, the will. We understand this power, in turn, when we understand practical reasoning, which is its characteristic activity.

On a non-reductionistic model, conceiving of intentional action as an irreducibly psychophysical process no longer seems so puzzling. If intentional bodily actions are themselves internal to the will's activity, then they are acts of that power. And there is no bar to saying that intentional actions are both physical happenings insofar as they involve changes in the spatiotemporal world, and psychological ones insofar as they are acts of the will.

* * *

The subject matter of this dissertation is metaphysical and epistemological. Its purpose is to show that the question of how we should think of ourselves cannot be answered independently of questions about self-knowledge, and to show how the epistemology of self-conscious agency is bound up with a certain metaphysics. But my motive in writing it has been, in part, practical. Many of the most difficult and contemporary ethical disputes concern who counts as one of "us" and these debates thereby raise the challenge of determining what our nature is. And even though the vast majority of contemporary moral philosophers are not substance dualists, many nonetheless work with an anthropology that is implicitly dualistic. Though the proposition that we are human beings is a commonsensical one, there is too often a tendency to overlook the bodily, animal aspects of our existence. The result is a one-sided emphasis on the virtues of

independence and autonomy to the exclusion of those virtues we need to confront vulnerability and disability.¹³ A dualistic anthropology is also manifested whenever it is assumed that we are essentially “selves” or “persons” (in one sense of that word)¹⁴ whose persistence conditions are different from those of human beings. While I do not directly argue against this self-conception, if the picture of self-conscious agency I put in place is correct then it is untenable. We are essentially embodied organisms and we are aware of this fact, if only implicitly, when we act intentionally.

¹³ The best attempt I know of to remedy this situation is MacIntyre 1999.

¹⁴ This is the sense in which it is essential to being a person that one has acquired or developed to some extent certain rational capacities. It is possible on this view for a particular individual to be a human being but not a person. As I note in Chapter 2.1, there is another sense of “person”, derived from Boethius, on which all human beings count as persons.

2.0 THE PROBLEM OF BODILY SELF-KNOWLEDGE

“To have an opinion is a state - a state of what? Of the soul? Of the mind? Well, of what object does one say that it has an opinion? Of Mr. N.N. for example. And that is the correct answer.” –Ludwig Wittgenstein, *Philosophical Investigations*, I § 573

In contemporary philosophy the term “self-knowledge” is often taken to denote the knowledge we have of ourselves that is not acquired on the basis of observation or inference. Whereas you would need some observational or evidential grounds in order to know, e.g., that I believe that the Steelers will win the Super Bowl or that I want to finish grading papers tonight, I need not listen to myself making football predictions or watch myself busily marking up my students’ papers in order to know that I so believe and desire.

However, “self-knowledge” might also naturally be taken to refer to the knowledge that one has of one’s own nature. Heard in this register, a philosophical inquiry into the nature of self-knowledge would be concerned with answering the question “What am I?” This question has had a long and distinguished career. Indeed, one way of characterizing the philosophical revolution inaugurated by Descartes would be to say that he made it central to epistemology, insofar as he argued that the question what one can know can be settled only after one has determined the nature of one’s epistemic capacities.

For Descartes, of course, the two senses of “self-knowledge” were intimately related. For he sought knowledge of himself, in the second sense, not by listing facts he had observed or

inferred about himself from sense perception, but by reflecting on those properties of himself that he claimed to know without having to observe them. It is therefore notable that today these two topics that fall under the heading of self-knowledge are usually treated independently of one another, and that while Descartes' claim about the our nature is widely rejected, a Cartesian conception of the province of self-knowledge in the first sense is still often taken for granted. This occurs when it is assumed that this sort of knowledge is restricted to psychological states and processes, described as such.¹⁵

My goal in this chapter is to argue that these two topics cannot be held apart. In particular I will argue that in order to be entitled to an adequate conception of ourselves we must also reject the Cartesian restriction on the scope of our non-observational and non-inferential self-knowledge. But I do not think that we should despair, for in the chapters that follow I will also defend the claim, famously made by G.E.M. Anscombe, that we have non-observational and non-inferential knowledge of what we are doing intentionally, where our intentional actions include some of our physically describable activities.

This chapter is organized as follows. I begin in Section I by contrasting the Cartesian view of oneself with the more commonsensical view that one is a human being. In Section II I clarify the content of Descartes' claim that when one is self-consciously aware of oneself, or aware of oneself as oneself, one is aware of oneself merely as a thinking thing. I argue that if he is correct, it is difficult to see what it would mean for one to be identical with a certain human being. I do not believe that this argument is merely of historical interest, and it is the task of Section III to explain why. Here I argue that there is an intimate relation between first person thought and self-knowledge in the first of the two senses above. The upshot is that if one holds

¹⁵ The qualification "as such" is important. Even if psychological states and processes are identical to physical states and processes, we do not have self-knowledge of them under physical descriptions.

that we have self-knowledge only of our psychological states and processes, one is committed to a problematic claim about the object of one's self-conscious self-awareness. In Section IV I show how Anscombe's claim that we have non-observational knowledge of our intentional actions is sufficient to overcome the Cartesian restriction on the scope of self-knowledge. I conclude by critiquing her argument that the first person pronoun is not a referring expression.

2.1 TWO RIVAL WAYS OF CONCEIVING OF ONESELF

What am I? The natural answer to this question is that I am an animal of a certain kind, a member of the species *homo sapiens*.¹⁶ In saying that I am an animal I do not mean to imply that I lack higher intellectual and moral capacities, as when people say in rebuke "Stop acting like an animal!" For as Aristotle famously remarked, human beings are distinguished among other kinds of animals in virtue of their rationality. The point of the rebuke is to castigate one for acting like a brute animal, an animal without the rational capacities characteristic of mature human beings.

It is crucial to note that the Aristotelian dictum "man is a rational animal" is not to be understood as a universal generalization but rather as a statement that rationality is a property of human nature. It is not impugned by the fact that some human individuals are not rational, just as the proposition that human beings have thirty-two teeth is not rendered false by the fact that many of us have had our wisdom teeth removed. This is not to say that there is no relation

¹⁶ For two recent defenses of this view see Olson 1997 and Wiggins 2001, ch. 7.

between individual humans and the property of being rational, however. That there is some relation between them was already implicit above when I said that the possession of rational capacities is “characteristic of mature human beings”. To say that human beings are rational animals means that, for any particular human being, it is part of his or her natural life-cycle to develop such capacities, and if this does not happen that fact requires a special explanation, such as lack of maturity, disease, or brain damage.¹⁷

The Aristotelian dictum also provides sense for one thing that might be intelligibly meant by the term “person.” According to the definition of Boethius, a person is an individual substance of a rational nature (*De persona et duabus naturis* c. 2), and human beings satisfy this definition insofar as they belong to a species whose essence is characterized by rationality.

I think that outside of philosophy this conception of what one is would be recognized as a piece of common sense. Yet many philosophers have denied it, including most of the giants of the early modern period. In this chapter I will be concerned with the route to one sort of denial of the natural view, namely, the one put forward by Descartes. Descartes famously concludes that he is not a human being—at least if we mean by “human being” a certain sort of living thing—but rather an immaterial substance whose essence lies solely in being a subject of thought and experience.¹⁸ My topic in the next section will be the central premise from which Descartes derives this conception of himself. But before turning to this I want to explain two related ways in which the Cartesian view contrasts with the natural one.

¹⁷ The Aristotelian dictum is thus an instance of a class of propositions Michael Thompson calls “natural historical judgments” (2008, Part I). See also Anscombe 1958, 38.

¹⁸ On Descartes’ view these are not separate characteristics, since according to him sense perception strictly speaking is itself a mode of thinking (AT VII 29/ CSM II 19).

First, as P.F. Strawson has stressed, we normally ascribe both mental and bodily characteristics to ourselves (1963, 83). But on the Cartesian view “I” must mean two very different things in the propositions “I believe that it is raining” and “I weigh 170 lbs.” There are two ways the Cartesian could posit an ambiguity. The first would be to say that in the proposition “I weigh 170 lbs.” the term “I” does not refer to oneself at all, but merely to the body with which one is somehow uniquely related, e.g., the body through whose eyes one sees and through whose movements one’s intentions get executed. The second would involve saying that while in the proposition “I believe it is raining” “I” refers to oneself, in the proposition “I weigh 170 lbs.” it refers to the composite of oneself and the appropriately related body. But either way the Cartesian is forced to posit an ambiguity where we are pre-philosophically not inclined to see any.¹⁹

Second, on the Cartesian picture one’s very nature consists in thinking, while normally we suppose that this is not so. I touched on this point already, when I noted that on the natural view the truth of the Aristotelian dictum is not impugned by the failure of some of its instances to possess the property that is ascribed by it to the kind they fall under. In this connection, it must be recognized that the Cartesian will accept that I am a person on the letter of the Boethian definition, but she will interpret this definition differently. For when the Cartesian says that I am a substance of a rational nature, she means that rationality characterizes my essence *qua individual self*. Thus, on the Cartesian view it is impossible for a person to persist without possessing capacities for thought and experience.

¹⁹ It is enough for my purposes that these remarks characterize at least one common understanding of the Cartesian position. The views of the historical Descartes on the topic of the unity of mind and body may be more complex. For instance, he is willing to adopt the scholastic language of the rational soul as a “substantial form” (e.g., AT VII 356/ CSM II 346, AT III 503,505/ CSM III 207-8) and an “incomplete substance” in relation to the whole human being (AT VII 222-3/ CSM 156-7). Whether those terms retained any of their traditional hylomorphic significance in his hands is another matter, but is one I need not go into here (but see Rozemond 1998, ch.5).

2.2 DESCARTES ON SELF-CONSCIOUSNESS

Descartes does not deny that we normally think of ourselves as human beings; he frankly admits that this is the view he took of himself prior to embarking on his meditations (AT VII 25/ CSM II 17). But he believes this view is naïve and cannot survive philosophical reflection. He derives his revisionary conclusion about his nature in part from an epistemological thesis about his awareness of himself, which may be put as follows:

I am aware of myself as something that possesses psychological properties and engages in psychological activities but not as something that possesses physical properties and engages in bodily activities.

This claim is one about the content of *self*-consciousness. There may be various ways of being aware of what is *in fact* oneself, but not all of these qualify as self-consciousness, or awareness of oneself *as oneself*. For example, Descartes might suffer from amnesia and read that René Descartes wrote the *Meditations on First Philosophy*. Here he would recognize the truth of the thought “Descartes is the author of the *Meditations*” without thereby recognizing something he could express by “I am the author of the *Meditations*”.

Thinking of oneself self-consciously, then, is not the same as thinking of oneself as the bearer of a proper name. Similar considerations show that thinking of oneself self-consciously is also not equivalent to thinking of oneself under any sort of definite description or demonstrative expression. For any such description or expression I can imagine some circumstance in which I would fail to realize that I am the person picked out in such a way (cf. Castañeda 1967; Perry 1979). We can even imagine scenarios in which I know that “this self” is F but fail to realize that I am this self. As Sydney Shoemaker says, “Why should the belief that *this self*’s pants are on fire, together with the desire not to be burnt, gird me into fire-dousing behavior? It will, of

course, if I know that this self is myself; but given what turns on it, knowing that cannot be just a matter of knowing that this self is this self” (1986, 132).

The fact that thinking of oneself as oneself is essentially indexical explains why, when Descartes concludes that he is not a human being, he does not commit himself to the self-contradictory proposition that Descartes is not Descartes, even if he acknowledges that Descartes is a human being. The word “he” in the sentence “Descartes concludes that he is not a human being” is functioning as an indirect reflexive (cf. Anscombe 1975, 141). What this means is that his conclusion is one that would be properly formulated in direct speech as “I am not a human being”. Thus, if he acknowledges that Descartes is a human being, Descartes is also committed to denying the truth of the proposition he expresses by “I am Descartes”.²⁰ This is not a manifest contradiction, however, since “I”—even when thought by Descartes—does not have the same sense as “Descartes”. This is so even if his conclusion is false and, as I shall continue to assume, when thought or said by him they are co-referential.

A traditional criticism of Descartes is that he tries to derive a metaphysical conclusion about his nature solely from an epistemological premise about how he is self-consciously aware of himself. But, the complaint runs, even if we grant the thesis that I am self-consciously aware of myself only as the bearer of psychological properties, this does not rule out the possibility that I am a bodily thing and I merely do not think of myself as possessing any physical attributes when I self-consciously think of myself. Now, it should be noted that this is not fair to Descartes, for he takes pains to make it clear that he does not derive his dualism from the claim about self-consciousness alone. He also argues that this knowledge is clear and distinct

²⁰ On this reading Descartes is committed to thinking either that he is the Ego that is attached to the man Descartes or to postulating an ambiguity in the word “Descartes” according to which it sometimes refers to the Ego and other times to the body of that Ego or their conjunction. Gassendi was therefore not merely being facetious when he referred to the subject of the *Meditations* as “Mind” (AT VII 260/ CSM 181).

knowledge and that God has the power to arrange things so that they correspond exactly to his clear and distinct ideas of them. Understanding how Descartes arrives at his dualism would therefore take us deep into his philosophy. I do not intend to undertake such exegesis here, but I do want to argue that even aside from any further premises, his claim about the nature of self-conscious self-awareness would in fact make it problematic to hold the view that one is a human being.

We can bring the problem into view by reflecting on the way we normally establish that a material particular picked out in one way is identical with a particular picked out in another. For example, suppose we wished to know whether a heavenly body visible during one period of the night was identical with a heavenly body visible during a different period. We would do so by identifying the location of the heavenly body during the first period and tracking its movements over time, perhaps calculating them based on our astronomical knowledge if it were not visible during the entire interval. If we found that the heavenly body we originally picked out had moved to the location where we expected to find the heavenly body identified in the second way at a certain time, we would conclude that there is but a single heavenly body visible at two different locations at different times. Indeed, it was through following some such procedure that early astronomers did identify the Morning Star with the Evening Star.

The procedure I have here described is one for determining when an object picked out under one specification is identical with an object picked out under another. But the reason the procedure is legitimate is that it is a conceptual truth that spatiotemporal particulars are individuated by their kind and their location. Thus, e.g., what differentiates a certain star from every other object that exists at a time is the fact that it is a star that occupies a certain place at that time (Evans 1982, 107).

The trouble that arises with the first person, though, is that if Descartes is correct then it becomes difficult to see what it would mean to say that I am identical with a particular human being. A human being is a certain sort of animal and what differentiates a human being from other objects at a time is the fact that it is the thing living that sort of life at a particular spatial location. But if Descartes is correct then when I conceive of myself self-consciously, I do not conceive of myself as having any physical states or engaging in any bodily activities as such; rather I conceive of myself in entirely non-bodily terms. Therefore, though I may conceive of myself as a subjective point of view, I do not conceive of myself as a material object that occupies a certain region of space at or around that point. And it follows that the conceptual truth that objects like human beings are individuated by their kind and spatial location will not be applicable to me. In essence, it isn't clear what it would *be* to say of some individual human being that he is identical with me.²¹

It might be replied that even if I am not self-conscious of myself as having any physical properties, still the existence of certain psychophysical correlations could make it reasonable that I am a particular human being. I might notice, for instance, that a certain human being's head is at the origin of my subjective point of view and that when I will to raise my arm his arm rises. And the best explanation of these psychophysical correlations is the supposition that I am that human being.

I have three responses to this line of thought. First, that one is a human being is something knowable by ordinary people and young children. But it is simply not credible that a person ordinarily knows that he is a certain human being by making this sort of elaborate

²¹ This argument presupposes a conception of identity such that what gives statements of identity a determinate sense is the presence in them of a sortal concept containing a principle of individuation. It would take me too far afield to argue for this conception of identity here. For a defense of it see Wiggins 2001, especially chapter 2.

theoretical conjecture. Second, an inference to the best explanation of the imagined kind could not provide one with knowledge that one is a certain human being, but at best only with defeasible evidence that that is likely the case. The most important response, however, is that the reply misunderstands the nature of the problem that Descartes' thesis about self-consciousness poses, which is not epistemological but semantic and metaphysical. It is not that his view, if correct, would make it difficult for one to know which human being one is or to have evidence that one is this or that particular human being. The problem is the stronger one that the thought that one is a human being would not even be a candidate for knowledge or justified belief. And this in turn holds because one's understanding of oneself would have been cut loose from the conceptual criteria for the individuation of things like human beings.

Even if this is correct, Descartes' dualism does not follow merely from the premise that one is self-conscious of oneself as having only psychological properties. All I have claimed is that it would not be clear what it would mean to think that one is identical with a human being. But it may be that the very idea of thinking of something as a particular is inseparable from thinking of it as a body with a spatiotemporal location. If that is so, the first person pronoun would have merely the "form" of a referring expression without actually picking out an individual.²² But this conclusion is still sufficiently problematic to make Descartes' thesis one that should be avoided.

²² This idea lies at the heart of Strawson's interpretation of Kant's critique of rational psychology in the "Paralogisms" chapter of the First *Critique* (1966, 165-6).

2.3 THE PROBLEM OF BODILY SELF-KNOWLEDGE

In order to evaluate Descartes' thesis we must first determine which of the ways one has of knowing oneself provide one with self-conscious self-knowledge, or knowledge of oneself as oneself. We have already seen that not all knowledge one has of oneself is self-conscious knowledge of oneself. If we mean by a "self-conscious capacity" a capacity whose nature it is to yield knowledge of oneself as oneself, then our task is to find a criterion for establishing which of our epistemic capacities are self-conscious ones.

A good way to begin on this task is by focusing on a famous passage in *The Blue Book* in which Wittgenstein draws a distinction between two uses of "I" (or "my"), which he calls its use "as subject" and its use "as object". Some examples of its use "as object" are, he tells us, "my arm is broken", "I have grown six inches", and "I have a bump on my forehead", while its use "as subject" includes "I see so-and-so", "I think it will rain", "I try to lift my arm", and "I have a toothache". He distinguishes these uses as follows:

One can point to the difference between these two categories by saying: The cases of the first category [the use of "I as object"] involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: The possibility of an error has been provided for. It is possible that, say in an accident, I should feel a pain in my arm, see a broken arm at my side, and think it is mine, when really it is my neighbor's. And I could, looking into a mirror, mistake a bump on his forehead for one on mine. On the other hand, there is no question of recognizing a person when I say I have toothache. To ask "are you sure that it's *you* who have pains?" would be nonsensical And now this way of stating our idea suggests itself: that it is as impossible that in making the statement "I have a toothache" I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me. (66-7)

In explicating this passage Shoemaker labels the characteristic of the use of "I" "as subject" to which Wittgenstein is calling attention "immunity to error through misidentification"

(1968, 81). As James Pryor (1999) points out, though, there are actually two different ways in which a belief might be liable to error through misidentifying its subject, and these are not clearly distinguished by Shoemaker. One would make a *de re* misidentification if he believes that a is F because he knows, in a certain way, that some distinct particular b is F, and he mistakenly takes b to be a. However, it is also possible for a subject to believe that a is F on the basis of some way of knowing such that he is entitled to accept the existential generalization that something is F and he believes partly on this same basis that it is a that is F when it is really some distinct individual or individuals that is or are F. In this case the subject's belief would be in error through what Pryor refers to as "which-object" (or *wh-*) misidentification.²³ This means that there are also two corresponding forms of immunity to error through misidentification.

Of these two forms of immunity to error through misidentification, immunity to *wh-* misidentification is more fundamental. Wittgenstein appears to think that when "I" is used "as subject" it will be immune to both sorts of misidentification. He says that there is no question of recognizing a particular person as yourself when you think you have a toothache (and we must add: when your belief is based on the way of knowing you have a toothache that is clearly at issue, viz., by feeling it). He also writes that it would be nonsensical to wonder whether it is you who has the toothache, and it would amount to the same sort of nonsense if you were to claim to know on this basis that someone has a toothache but wondered whether it was you who has it. The entitlement that this way of knowing about yourself gives you for believing that the property of having a toothache is instantiated is sufficient to entitle you to believe that it is you who has it. But the belief in question is immune to the former sort of misidentification (*de re*

²³ As these formulations make clear, whether a certain singular judgment is liable to error through misidentification in either sense depends not only on its content but also on the way of knowing about an object on which it is based. Wittgenstein does not recognize this point in the quoted passage above, but it is explicitly made by Shoemaker (1968, 82).

misidentification) *because* it is immune to the latter sort (*wh*-misidentification). Since your way of knowing that something has a toothache suffices on its own to entitle you to believe that it is you who has it, your entitlement for believing that you have a toothache does not rest upon a belief about some object that it has a toothache and a belief that you are that object (cf. Pryor 1999, 285).

The concept of immunity to error through misidentification affords an interpretation for the idea of a self-conscious capacity as I introduced it: it is a capacity that is immune to *wh*-misidentification relative to the first person pronoun and thus immune to *de re* misidentification relative to it as well. Self-conscious capacities need not be infallible, but if it is the nature of a certain epistemic capacity to yield knowledge of oneself as oneself then, when the capacity does provide one with knowledge that some property F is instantiated, it provides one with knowledge that it is oneself that is F. Therefore, when one's belief about oneself is based on such a capacity it will not be possible for one to be mistaken due to the fact that she has misidentified which object it is that is F.

First person thought is thought about oneself as oneself. Thus for any expression to qualify as a version of the first person it must be tied to capacities on the part of the person to whom the expression belongs to gain knowledge of herself in ways that are immune to *wh*-misidentification, or that correspond to Wittgenstein's use of "I" "as subject". The English pronoun "I" is such an expression, but any expression not tied in part to such capacities fails to qualify as a version of the first person, no matter how similarly to "I" it might otherwise behave. This point is brought out by Anscombe in her essay "The First Person", when she contrasts our use of "I" with the use of "A" by the members of an imaginary community, which she describes as follows:

Imagine a society in which everyone is labeled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various: “B” to “Z” let us say. The other, “A”, is stamped on the inside of their wrists and is the same for everyone. In making reports on people’s actions everyone uses the names on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterance of the name on his own chest and back in the sort of way and circumstances in which we tend to respond to our own names. Reports on one’s actions, which one gives straight off from observation, are made using the name on the wrist. Such reports are not made on the basis of observation alone, but also on the use of inference or testimony or other information. (1975, 143-4)

In this community, “A” is a device for referring to the speaker of the utterance in which “A” occurs, just as “I” refers, in the basic case, to the speaker in English.²⁴ Yet unlike “I”, “A” does not express the first person; as Anscombe puts it, the “A”-users lack self-consciousness (ibid., 144). The reason is that the primary way they have of acquiring information about themselves is through observation. An “A”-user may directly observe the actions of an “A”-user, including her own, or she may infer that an “A”-user is doing such-and-such from some other observationally acquired evidence, and the “A”-users can share the information they acquire in these ways. But insofar as she acquires information that A is doing such-and-such on the basis of observation and inference, an “A” user’s judgments about A are liable to error through (*wh-*) misidentification. For instance, B might be tangled up with C, see a fist being clenched and believe on this basis that A is clenching her fist, though it is actually C who is engaged in fist clenching. In this case B’s observation gives her evidence that someone is clenching her fist, on the basis of which B incorrectly judges that A is clenching her fist.

²⁴ John McDowell (1998) argues that if we try to think through the “A”-language we realize that in the end it may be incoherent. The very idea of being a language-user may require the kind of self-consciousness that the “A”-users are supposed to lack. This incoherence would not affect Anscombe’s deployment of it in her thought-experiment, though, since her purpose in introducing it is merely as an aid in getting clear about the kind of self-consciousness presupposed by our concept of the first person.

The members of the “A”-community are not self-conscious because they can acquire knowledge of themselves only in the ways that they acquire knowledge of others, i.e., through observation and inference.²⁵ Unlike the “A”-users, we also have knowledge of ourselves that is not based on observation or inference. For example, in his list of the uses of “I” as subject Wittgenstein includes judgments about our own beliefs (“I think it will rain”) and sensations (“I have a toothache”). These sorts of judgments are not normally made on the basis of observation, inference, or any other ground, and when they are made groundlessly they are immune to *wh*-misidentification. If it should turn out, say, that my groundless belief that I believe that *p* is false due to self-deception, I would not still be entitled to believe in this way that someone believes that *p*. In order to know that someone other than myself believes that *p* I would need empirical evidence to that effect (cf. Wright 1998, 20). There is thus an intimate connection between self-consciousness and what philosophers call self-knowledge, or groundless knowledge of oneself. Self-conscious capacities provide knowledge of oneself that is immune to *wh*-misidentification, and beliefs that have this property are beliefs that are not justified by observation or inference.

We do not normally think, however, that the only ways we have of gaining knowledge of ourselves are via our self-conscious capacities. Though these capacities provide us with knowledge of ourselves that we can express using “I” in a way that corresponds to Wittgenstein’s use of it “as subject”, we also use this expression in contexts in which it is used “as object”. To borrow one of his examples, I might see a certain human being with a bump on his forehead and, believing that I am in front of a mirror, form the belief that I have a bump on my forehead. I would be justified in this belief if “I” has the same sense in this sort of case as when it figures in my thought about myself as myself, for then I could justify my belief that I

²⁵ Observation and inference are their ultimate sources of self-knowledge, though such knowledge can also be transferred by way of testimony once acquired.

have a bump on my forehead by citing the fact that I am justified in believing both that *that man* whom I see in the mirror has a bump on his forehead and that I am that man. However, I argued in the previous section that the judgment that I am a human being has a clear sense only if I can have self-conscious knowledge of myself as a material object. Otherwise, in its use “as object” “I” could only mean something like “the human body to which I am closely joined in some unique way.”

This latter, Cartesian conception of the self is widely rejected and I am not about to recommend it. But given that self-conscious capacities are capacities for self-knowledge, the following conditional is true: if all of our capacities for self-knowledge yield knowledge only of our purely psychological states and activities, then Descartes’ thesis that we are self-consciously aware of ourselves only as bearers of psychological properties is also correct. Therefore, let us call the problem this conditional makes for the view that one is a certain human being “the problem of bodily self-knowledge”. This problem should be one for many philosophers, for it is still commonly supposed that Descartes was at least right that we have non-empirical access only to our own minds.²⁶ If what I have said here is correct, then we must reject this restriction if we are to entitle ourselves to a more adequate conception of our own nature.

Note that it is not the phenomenon of immunity to *wh*-misidentification as such that generates the problem of bodily self-knowledge. When one refers to an object in her perceptual field demonstratively, her judgment is also immune to *wh*-misidentification relative to the demonstrative pronoun (cf. Shoemaker 1968, 83; Evans 1982, 179-91). And it would be absurd

²⁶ Colin McGinn is unusual in this respect. He accepts the Cartesian restriction and, as a result, detaches oneself from one’s body. He writes: “[A]s-object uses [of ‘I’] are always replaceable by ‘my body’, which brings out the fact that the object of reference about which the mistake is made is really something distinct from myself... to say of a certain body that it is mine is to imply that it is the body through which I speak and act and is the vehicle through which I undergo various sensations...” (1983, 49). Even if his conclusion is an unappealing one, I have argued he is right in thinking that such a conclusion follows from his starting point.

to think that there is a problem about how something identified as say, *this man* could be identical with a certain human being. Rather, the problem arises from the premise that there is an intimate relation between self-consciousness and self-knowledge, when this is combined with the further premise that we do not have self-knowledge of any of our physical states or activities, and the premise that one must possess a capacity for self-consciousness of such states or activities in order to be in a position to identify oneself with a material particular picked out through observation. The fact that some judgments about oneself are immune to *wh-*misidentification enters the discussion solely as part of the argument for the first of these premises.

2.4 ANSCOMBE ON KNOWLEDGE IN INTENTION AND SELF-REFERENCE

A solution to the problem of bodily self-knowledge would involve showing or making it plausible that we have one or more self-conscious capacities that provide us with knowledge of some of our physical states or activities, described as such. I believe that this challenge can be met, and that resources are to be found in the epistemology of action. It is the central claim of Anscombe's *Intention* that rational agents have a special epistemological relation to their intentional actions, such that they know what they are doing intentionally without observation or inference. She calls this way we have of knowing our intentional actions "knowledge in intention" (1963, §32).

That we have non-observational knowledge of what we are doing is plausible for ordinary cases of intentional action. Suppose I am looking through the pantry. You ask me what I am up to and I say that I am searching for some flour. Wanting to know more, you ask me why

I need flour and I reply, “I’m baking a cake”. There are two things to note about your query. First, your question here is the question “Why?” that asks for my reason for acting as I do. Anscombe argues that the fact that this question is applicable to my baking a cake means that I am doing it intentionally (ibid., §5). The second thing to notice, though, is that the question has application because the epistemic relation that I bear to my own intentional actions is different from the one that you bear to them. Without further evidence, you would be irrational to believe that I am baking a cake. Yet when I answer your question you accept it because you take it that I am in a position to know it, despite the dearth of empirical evidence.²⁷

Our immediate, non-observational beliefs about what we are doing, when they constitute knowledge, are also immune to error through *wh*-misidentification. Take, for instance, the man whom Anscombe alludes to who thinks to himself “Now I press Button A” (ibid., §32). If it should turn out that his belief is false, as it does in Anscombe’s example (he is actually pressing Button B), then he does not retain any entitlement on this basis for thinking that someone is pressing Button A. In order to be justified in thinking that someone other than himself is pressing Button A he would require some evidence to that effect.

It is in the nature of the will to provide the one who possesses it with knowledge of himself as himself, and this means that it is a self-conscious capacity. But the knowledge of oneself that is provided by the will, so conceived, includes knowledge of oneself as engaged in physical activities. Moreover, the sort of physical activities that one knows oneself to be engaged in in this way are the activities characteristic of the life of a certain kind of embodied

²⁷ The history of the philosophy of action in the latter half of the twentieth century was characterized by its focus on the first of these two points while virtually ignoring the second (the exceptions here being, besides *Intention*, Hampshire 1959, Harman 1976, Velleman 1989). It has recently become a topic of renewed interest, however (cf. Falvey 2000, Moran 2001, 125-8 and Setiya 2007; 2008).

being, a being that one could learn is an animal of the species *homo sapiens*.²⁸ Therefore, if Anscombe is right, knowledge in intention constitutes a solution to the problem of bodily self-knowledge.

Anscombe's contention that knowledge of action "in intention" is both immediate and non-observational is not uncontroversial, however. In his paper "Intention and Uncertainty" H.P. Grice argues that beliefs about action that were not grounded in observation or inference would be epistemically deficient, no better than cases of wishful thinking. Grice's objection must be taken seriously and it will be my goal in the next chapter to show why it is not ultimately compelling. Before turning to this, however, I want to address a worry about the use to which I propose to put Anscombe's account of knowledge in intention that comes from Anscombe herself. I have claimed that her account enables us to understand how "I" can refer to a human being, when used by one of us. But the fact that I am enrolling Anscombe in such a project is sure to raise eyebrows, for she is notorious precisely for denying that "I" is a referring expression.

In fact, Anscombe does think that knowledge in intention is important for understanding the first person, and in particular, for understanding the meaning of propositions of the form "I am N.N.":

The reason why I take only thoughts of actions, postures, movements, and intended actions is that only those thoughts both are unmediated, non-observational, and also are descriptions (e.g., "standing") which are directly verifiable or falsifiable about the person of E.A. Anyone, including myself, can look and see whether that person is standing. (1975, 157)²⁹

²⁸ Anscombe writes that the term "intentional" refers to a certain form of description of events and she says that what event-descriptions that manifest this form have in common is that they fall under the category of "vital descriptions" (ibid., §47).

²⁹ In saying that the thoughts of actions, postures, movements in question are "unmediated", Anscombe means that they are immune to error through (wh-) misidentification (see the discussion of Baldy on pp. 158-9). It is significant that in this passage she includes thoughts about postures as well as actions. We know the relative positions of the

Although her emphasis in this passage is on “verification”, what she denies is precisely Descartes’s thesis that we have self-conscious self-knowledge of ourselves as purely psychological beings. However, Anscombe also thinks that truths of the form “I am N.N.” are grammatically misleading, as she denies that they should be interpreted as claiming an identity between particulars picked out in different ways. And though she agrees that it may be true that I am F if and only if (in my case) J.S. is F, she denies the reason that such biconditionals hold is that “I” as said by me and “J.S.” are co-referential singular terms.

Anscombe’s view is surely implausible. She may be correct that we cannot tell by the mere syntax of propositions like “I am F” that “I” is a referring expression. There are also sentences, such as “It is raining”, that have a subject-predicate structure but in which the subject term does not play the role of a referring expression (ibid., 151). However, the claim that her thesis is implausible does not turn merely on this syntactical feature but also on the logical and semantic properties of propositions whose subject term is the first person pronoun. For instance, the truth or falsity of a given use of “I am F” is determined by whether a certain individual is F. Moreover, if the term “I” is understood in the same way in its occurrences in both “I am F” and “I am G” then it necessarily follows that if both of these are true then there is some one thing that is both F and G. In this respect an ordinary use of “I” behaves like other singular referring expressions and unlike the use of the pronoun “it” in “It is raining”.

Anscombe is aware of all this; she simply thinks that the price of taking “I” to be a referring expression is so high that we should accept her view with all of its attendant difficulties rather than be stuck with the alternative. For she claims that if “I” is a referring expression, it

various parts of our bodies through one of a family of epistemic powers that are commonly grouped under the heading of “bodily awareness”. I discuss whether this family of powers are self-conscious ones in the Appendix.

refers to (a stretch of) a Cartesian Ego (ibid., 152). But it seems to me that the right response to this is to take the conclusion “either ‘I’ refers to a Cartesian Ego or it does not refer at all” as itself a *reductio* of the conjunction of whatever premises entail it. We should therefore examine Anscombe’s argument for her conditional.

She arrives at it by assuming that if “I” were a referring expression its way of referring would have to be modeled on the way in which demonstratives refer. Since a singular term’s sense is the way that the object it designates figures in the thought of a rational being (Frege 1892, 152), Anscombe’s claim is that the sense of “I” must be modeled on the sense of demonstrative expressions. A demonstrative expression refers by picking out an object present to the perceptual consciousness of the thinker of the demonstrative thought (cf. Evans, ch. 6). Now, one could refer to oneself even in a state of complete sensory deprivation, but the only thing that could be said to be “present”, in any sense, to one’s consciousness in this state is consciousness itself. Anscombe infers from this that the only possible referent for “I” is a thing of pure consciousness, (a stretch of) a Cartesian Ego.

Anscombe’s argument hangs on her assumption that if “I” refers, then the way in which it refers must be modeled on that of another kind of referring expression. If one is forced to choose from such a list, then perhaps she is right that demonstrative thought would be the best or only sort of expression which could serve as a model. But why should we be forced to choose at all? Why could it not be that the sense of “I” is *sui generis*, as Frege believed (1918, 333)? Indeed, if we read her argument as a *reductio* as I have suggested, then what her argument actually shows is that we have good reason to adopt Frege’s position.

Anscombe does consider the Fregean view but she quickly dismisses it, on the ground that on its own it provides no positive account of how “I” refers, and so is empty (1975, 149). I

agree that anyone who takes this route must provide some specification of the sense of “I”, but I also think this criterion can be met. I can bring this out by pointing to a certain contrast between thought about oneself and demonstrative thought. The way that the object of a demonstrative expression figures in thought depends on its being perceptually present to the thinker. Our knowledge that an object picked out demonstratively has some property is the result of certain cognitive capacities, the senses, whose nature it is to provide one with knowledge of things that affect them. Since it is no part of the nature of the senses that what affects them should be oneself, if in a given case the subject of perception is the same as the object perceived then this identity is merely accidental (Rödl 2007, 8). On the other hand, we have seen that self-conscious cognitive capacities are capacities whose *nature* it is to provide knowledge of oneself as oneself. It follows that self-conscious self-reference is not a sort of demonstrative reference.³⁰ However, just as our perceptual capacities underwrite one sort of way that something may figure in the thought of a rational being, our self-conscious capacities underwrite a different way. For in being the one who possesses a self-conscious capacity one has available to oneself a way of gaining knowledge of the one who is exercising it, as such. In this case, it will be no accident that the subject of the capacity and the object known through the capacity are one and the same. There is therefore no barrier to the idea that the function of the first person is to refer to an object.

Anscombe’s idea that as rational agents it belongs to us to know what we are doing intentionally immediately and without observation is a promising way for overcoming the problem of bodily self-knowledge. But we must now confront the objection that beliefs about

³⁰ Anscombe is therefore wrong to assimilate the way in which one’s own consciousness is “present” to oneself to the way that perceived objects are. If I know that I am thinking “I won’t let this happen again!” in a sensory deprivation tank, this is not because I perceive myself thinking.

action not formed on the basis of observation or inference could not amount to cases of knowledge. Replying to this objection will require us to provide an account of the origins and nature of knowledge in intention.

3.0 PRACTICAL KNOWLEDGE AND PRACTICAL REASONING

“[T]he will is a particular way of thinking—
thinking translating itself into existence [*Dasein*],
thinking as a drive to give itself existence.”—
G.W.F. Hegel, *Philosophy of Right*, §4

The purpose of Chapter 2 was to set up a problem that arises when we reflect on ourselves and to indicate a solution. I dubbed this problem “the problem of bodily self-knowledge”, and I argued that in order to be entitled to a satisfactory self-conception we must have self-knowledge of some of our own physical states or activities, described as such. By “self-knowledge” I mean knowledge of ourselves that is not based on observation or inference. I suggested that if Anscombe is right that rational agents typically know what they are doing intentionally without observation or inference, then the epistemology of action contains the resources for responding to the problem of bodily self-knowledge. My goal now is therefore to investigate the origin and nature of knowledge in intention.

I organize my inquiry around responding to an objection to the very idea of knowledge of action that is not grounded in observation or inference. The worry is that beliefs about actions that are not so grounded could be no better epistemically than cases of wishful thinking, a paradigm of epistemic irrationality. In Section I, I explain how H.P. Grice uses such a worry to motivate an inferential account of knowledge in intention. According to this view, when one non-observationally knows what one is doing intentionally, one’s knowledge should be seen as

the result of an inference from one's knowledge of what one wills to do. Grice's view presupposes that groundless knowledge is not always epistemically deficient, however, since he assumes that we have self-knowledge of our will. This assumption also opens up a strategy for the defender of Anscombe to pursue. For if intentional actions are themselves acts of the will, our knowledge of our intentional actions will simply be a case of the self-knowledge we have of our will. I argue that we should understand Aristotle as holding just such a view of action. This is significant because unlike the Gricean account, the Anscombian one has the resources for resolving the problem of bodily self-knowledge.

Sections II and III are devoted to developing an interpretation of Aristotle's doctrine that acting can be the conclusion to, and not merely the effect of, practical reasoning. I do this through examining the accounts of practical reasoning proposed by Davidson and Anscombe. Both of these authors think the Aristotelian doctrine embodies an important truth about action, but I think that in the final analysis only the latter gives a satisfactory interpretation of it. While Davidson gives us a theory according to which the conclusion of practical reasoning is a certain kind of judgment, Anscombe provides us with a picture on which practical reasoning is formally distinctive precisely insofar as its conclusion is not a judgment but acting in a certain way. In Section IV I return to the task of explaining how accepting the Aristotelian doctrine enables us to respond to Grice's worry about groundless knowledge of action and I clarify the legitimate role empirical knowledge has to play in the context of intentional action. Finally, in Section V I compare my Aristotelian story of knowledge in intention with that of cognitivists such as J. David Velleman and Kieran Setiya.

3.1 GRICE'S OBJECTION, AND THE OUTLINE OF A REPLY

The idea we are investigating is that as rational agents we have the power to have knowledge of what we are doing intentionally that is not grounded in observation or inference. While this capacity is operative in every case of intentional action, this is liable to be obscured by the fact that we often also know what we are doing by way of the senses. In order to bring into view the special access we have to our own intentional actions it is therefore useful to consider a case in which an agent knows he is performing an intentional action but in which ordinary empirical knowledge of it is out of the question. I gave an example with this shape in the last chapter. Suppose you see me looking through the pantry and ask me what I'm doing. I reply "I'm baking a cake". Before I told you, you would have been irrational to believe I was in the initial stages of cake-baking. Yet when I reply to your query you accept it because you recognize that I am in a position to know. In this case I know what I am doing intentionally even though I seem to lack sufficient evidence on the basis of which I could rationally form the belief that I'm baking a cake. Certainly, everything my senses tell me about my immediate surroundings is not enough to indicate I am doing so, as opposed to making some other culinary item.³¹

Yet the thought that I have the ability to know I'm baking a cake in this situation immediately and without observation may seem epistemologically alarming. After all, what I claim to know about in this way is a certain process occurring in publicly accessible space—a process that as it unfolds will also be an object of ordinary empirical confirmation. Beliefs about such processes not formed on the basis of sufficient evidence are normally epistemically

³¹ This sort of case also creates difficulties for the so-called "two factor" thesis. According to this view one knows without observation only what one intends to do. The knowledge that one is actually executing one's intention must be observationally grounded (Donnellan 1963). In the example above, I know more than that I intend to bake a cake, I know I am actually baking one, though this knowledge is not based on observation (cf. Falvey 1999).

deficient in a way that prohibits them from counting as knowledge. Moreover, if I am intentionally baking a cake then I want to bake one. So if Anscombe is right about knowledge in intention then when rational agents act intentionally they typically form beliefs that are about empirical processes, whose occurrence they desire, which are not formed on the basis of sufficient evidence. This means that these beliefs share several features in common with those that are formed as a result of wishful thinking. It is for this reason that Grice claims that if we did form beliefs about our intentional actions that were not grounded on observation or inference they would be epistemically no better than cases of wishful thought (1971, 268).

Grice's response is not to reject the very idea of knowledge in intention, but to argue that Anscombe gets its epistemic structure wrong. While in my example above it may appear I lack sufficient evidence on the basis of which I form my belief that I'm baking a cake, according to Grice this appearance is misleading. It is true that the information about my immediate situation that my senses provide would not be a sufficient basis on which to rationally infer that I am baking a cake, but my evidence is broader than this. Grice's position is that knowledge of action in intention is the result of an inference based on the content of what one wills to do.³² So on his view my knowledge that I am baking a cake is grounded on an argument one premise of which is that I will to bake a cake, and the other that I tend to do what I will to do (ibid., 277-8).³³ The first premise expresses my knowledge of my will, while my knowledge of the second is based on past experience.

³² A more recent, sophisticated version of the inferential theory of knowledge in intention is defended by Paul 2009.

³³ If I know some particular feature of my circumstances that could prevent this tendency from realizing itself, then instead of supporting the conclusion that I am doing A, my past experience may only support the weaker conclusion that there is some probability I am doing it.

The inferentialist account of knowledge in intention is also problematic, however, in the light of my current project. That project is to try and make sense of the idea that human beings have self-conscious knowledge of themselves as bodily creatures, and I have argued that self-conscious knowledge is self-knowledge.³⁴ What made the turn to the epistemology of action a promising one was precisely Anscombe's claim that knowledge in intention is knowledge of our bodily activities that is a form of self-knowledge. But if Grice were correct, then one's self-knowledge would be restricted to one's knowledge of a volition that is not itself a bodily state or activity described as such. One's knowledge of what one is *doing* intentionally would not itself be a genuine form of self-consciousness. This motivates the project I will pursue in the rest of this chapter, which is to pursue an Anscombian alternative to the inferentialist account while attempting to reply to Grice's worries about its epistemic credentials.

It is important to note that Grice does not hold that *in general* beliefs not formed on the basis of observation or inference are epistemically defective. His own account presupposes we have this sort of knowledge of our own will. In this he is following the generally accepted philosophical opinion that we typically have self-knowledge of our own mental states, such as our beliefs and acts of will (e.g., intentions or volitions). The basis of this sort of knowledge might also seem puzzling, though, and in two respects. First, given that some creatures (such as non-human animals and very young children) with mental states do not have self-knowledge of them, why do mature human beings? Second, given that this higher-order awareness is not based on sufficient evidence, why does it normally constitute knowledge?

³⁴ See Chapter 2.3.

The answer to both of these questions is that the beliefs and acts of will formed by mature human beings are typically acts (in the sense of actualizations) of reason, and reason is by its nature a self-conscious power.

Consider our knowledge of our own beliefs. When a belief is an act of a rational creature's power of theoretical reason, it is that creature's settled response to the reasons for believing its content, as such. But her response to those reasons would not be what it is if she were not aware of it *as* her settled response to them. It is unintelligible to suppose that a person's belief that *p* could manifest her power to be responsive to reasons *qua* reasons and yet it turn out that she was blind to the fact that a belief with that content was rationally appropriate relative to the other considerations she was aware of.³⁵ Thus, when a rational creature with the concept *belief* forms a belief as an exercise of her power of theoretical reason, she will be able to express her awareness of her belief as her response to the reasons for believing it with the words "I believe that *p* because such-and-such". And since this awareness follows from the nature of the power in question, it is no accident that a rational creature's beliefs about its own beliefs are true, even if they are not formed on the basis of observation or inference from sufficient evidence.

A particularly illuminating instance of this occurs when one exercises one's power of theoretical reasoning in acquiring a new belief through deliberation. Evans imagines a situation in which he is asked whether he believes there will be a third world war and writes, "If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I

³⁵ These considerations can include second-order reasons such as the fact that one would not have a belief on a particular topic if one had not learned it from a reliable informant. One need not be able to remember the exact source of all of one's beliefs in order to think oneself justified in holding them.

believe that p by putting into operation whatever procedure I have for answering the question whether p” (1982, 225). Moran uses the term “transparency” to denote the relation that holds between the question whether I believe that p and the question whether p (2001, 60). Why should the first of these questions be transparent to the second in these circumstances? The reason Evans can answer a question about his own belief as to whether there will be a third world war by answering the question whether there will be one is that his belief is his answer to this latter question. And answering a question is not the *kind* of activity one can be engaged in without being conscious of the fact that one is doing so. Therefore, Evans can answer a question about what he believes by being aware of his answer to the question to which it is transparent.

This sort of case is illuminating but it is also special, insofar as it concerns how one knows a belief that one has arrived at by making up one’s mind what to think. However, we also normally know what we believe on matters about which our minds are already made up. But even if one does not exercise one’s power of theoretical reason in consciously deliberating in these cases, they remain acts of that power. That is, they remain one’s rational answers to the question what to think about the relevant topics, and this means that one holds them in the awareness that they are one’s responses to the reasons for believing as one does. This sensitivity is something that one can manifest by saying why one holds the beliefs one does, should the question arise.³⁶

³⁶ Moran calls expressions of belief knowable in this way “avowals”, and he stresses that not every expression of a belief about oneself is an avowal (2001, 86). A person might believe she believes that she is not respected by her colleagues only because she learned it from her psychiatrist. If she expressed this belief with the words “I believe my colleagues do not respect me”, her expression would not be an avowal. In psychoanalytic terminology, her belief that her colleagues do not respect her is *unconscious* or *under repression*. Repressed beliefs can be formed by defective exercises of reason or by psychological mechanisms that Mark Johnston calls “mental tropisms” (1988, 86). The important thing to note is that cases in which a belief cannot be avowed are pathological ones. For our beliefs to be the objects of self-consciousness, they must be formed on the basis of a non-defective exercise of reason.

So far I have focused on our knowledge of our own beliefs, but our power of practical reason is also a self-conscious power. Korsgaard arrives at this conclusion by a somewhat different route, by reflecting on a problem about psychological causality raised by Nagel. Nagel points out that it is not sufficient for rational choice that it is caused by a belief and desire. For instance, we can imagine a man who has been conditioned in such a way that when he wants a drink and believes that the object before him is a pencil sharpener, he puts a coin in the pencil sharpener (1970, 33-4). The problem here is that the belief and the desire are not causally connected in such a way as to engage the rationality of the person whose belief and desire they are.³⁷ What Korsgaard notices is that it is not sufficient for the rationality of the person to be engaged that the belief and desire that are responsible for the person's choice have conceptual contents that would make what he does on its basis rationally intelligible. After all, a person could be psychologically conditioned to perform an intelligible action such as putting a dime in a soda machine just as well as a pencil sharpener. The lesson is:

Neither the joint causal efficacy of the belief and the desire, nor the existence of an appropriate conceptual connection, nor the bare conjunction of these two facts, enables us to judge that a person acts rationally. For the person to act rationally, she must be motivated by her own *recognition* of the appropriate conceptual connection between her belief and desire (1998, 221; italics added)

What is missing in cases of non-rational action such as conditioning is the subject's awareness of the conceptual connection between his desire and belief and the choice that he makes on the basis of them. E.g., one must recognize that one's choice to put a dime in a soda machine is a rational response to the end of getting a soda given that it is possible to get a soda

³⁷ This is a version of the problem of causal deviancy raised by Davidson (1973, 78-80). For our purposes we need not go into Nagel's own solution to the problem.

by putting a dime in the machine. In other words, insofar as they are rational, acts of will no less than acts of belief are inherently acts of a self-conscious power.

The fact that the will is a self-conscious power means that Grice is not unjustified in his assumption that we have the power to know without observation or inference what we will to do. But he also assumes that our knowledge of our will extends only to a mental act of willing and not to our intentional actions themselves. On his view one's self-conscious power of the will only provides one with knowledge of something one can use as a premise from which to infer that one is acting in a certain way.

Grice's theory of the extent of our self-knowledge of our will appears to rest on a certain assumption about the metaphysics of action. The assumption is that a physical intentional action is a complex entity made up of an act of will and a bodily movement, where the relation that obtains between them is presumably causal.³⁸ We might label this model of action the "Cartesian model", since it was held by Descartes, though it would also be appropriate to ascribe it to Locke. The proponent of the Cartesian model may grant that rational agents have self-knowledge of their wills, but for him the scope of this self-knowledge is restricted to purely psychological acts. Since an agent's bodily goings-on are external to the exercise of the will, they do not fall within the domain of its self-consciousness.

We can understand what it means to say an agent's bodily goings-on are "external" to the exercise of the will by considering Hume's rather different claim that reason alone cannot influence the will. What he does *not* mean by this is that a passion (which includes both desires

³⁸ Here "bodily movement" might be taken in either a transitive sense or an intransitive one. In the first sense what intention or volition causes is the agent's moving her body, while in the second it causes the process of her body's movement. These are two different theories, but they both conceive of intentional action in terms of the causation of a physical event by a psychological one.

and volitions) cannot be caused solely by a belief. He explicitly asserts that a belief that a certain object would cause a feeling of pleasure or pain can by itself cause a desire or aversion for that object (*T* 1.3.10).³⁹ His claim is rather that passions are not generated by the process of reasoning, the process which occurs when the power of reason is exercised. On Hume's view it is the function of reason to compare ideas and impressions, which are essentially true or false, in order to infer relations between them by drawing demonstrative or causal inferences (1.3.4-7). Since he thinks that passions lack representational content, they are the wrong sort of thing to be the result of drawing either sort of inference. It may therefore be true on his view that one reasons, say, that touching a certain orange chunk of metal would cause one to be burnt and thereby becomes averse to touching it. However, whereas one's belief that one will be burnt can be generated from other beliefs by an inferential process, the process whereby that belief causes the desire not to be burnt cannot be an inferential one. Rachel Cohon summarizes Hume's view as follows:

[O]n [Hume's] account of production by reason alone, the *conclusion* of a piece of reasoning (a belief lively enough that it is believed) might perfectly well cause a passion to come into existence. The resulting passion need not, and indeed could not, be the product of reasoning alone, because its causation would not be a step in any piece of reasoning. A passion is not produced by the comparing of related ideas and is not found evident—it is not generated by a reasoning activity—even if it is caused by a belief that was so produced. (2008, 76-7)

There is thus a sense in which on Hume's view the formation of a passion is external to the exercise of reason. It is in this same sense that on the Cartesian model one's bodily movements are external to the will. For on that (non-Humean) view practical reasoning, which it is the function of the will to carry out, always concludes in a psychological act of will. The

³⁹ Thus Hume does not hold what has come to be known as "the Humean theory of motivation" (see M. Smith 1987).

causation of the appropriate bodily happenings by volition is not itself a step in a piece of practical reasoning.⁴⁰

It is because he holds the Cartesian model of action that Grice thinks the immediate and non-observational beliefs we have of our own will, whose status as knowledge he never doubts, must be restricted to purely psychological acts. This suggests that one could resist Grice's conclusion by denying the central claim of that model, that when one acts intentionally one's bodily goings-on are external to the exercise of the will's activity of practical reasoning. If we can make sense of that we would also be in a position to say that self-knowledge of the will extends all the way to those goings-on.

I believe a careful reading of *Intention* reveals that Anscombe holds just such a view. She claims in that work that the inability of "modern philosophy" to understand knowledge in intention results from the fact that it has "an incorrigibly contemplative conception of knowledge" (1963, §32). The object of contemplative knowledge is one that exists prior to and independently of one's knowledge of it. I get to know, say, that there is a tree in my yard because I have seen it there, and the tree is an element of reality that exists whether I know about it or not. Anscombe's central target is the view that the only knowledge that rational agents have of their intentional actions is contemplative knowledge. Thus, while she initially characterizes the sort of knowledge of action she is concerned with negatively, as knowledge that is not by

⁴⁰ Kieran Setiya has pointed out, in personal correspondence, that a theorist could think of an act of will (e.g., an intention) as causally distinct from the bodily happenings that occur when we act and yet not hold that the bodily happenings are external to reason, if he also holds that the relation between them is not "different in kind from the transitions among mental states that count as practical reasoning". For this reason Setiya denies that he and Velleman (and perhaps others) hold the Cartesian model of action. While in the absence of more detail I find Setiya's suggestion obscure, I can perfectly well grant this. For my purposes all I need is the following: (1) that the Cartesian model is one natural way of conceiving the relations that hold between the will, self-knowledge, and bodily action, (2) that Grice can be plausibly interpreted as presupposing such a model, and (3) that the Aristotelian conception of action that I aim to put in place below is a straightforward way of resisting the Cartesian model. I critique Velleman and Setiya's theories of knowledge in intention in Section V using arguments that do not presuppose they are instances of the Cartesian model.

observation (§8) or inference (§28), her rejection of the view that this knowledge is contemplative allows her to provide a positive characterization of it as well. It is *practical* knowledge, or knowledge that is the cause of that which it is the knowledge. As Anscombe makes clear, the sense of “causation” in play here is what the medievals called “formal” causation: it means that if a rational agent does not have self-conscious knowledge of what she is doing under a certain specification, then the action is not intentional under that specification (§48).⁴¹ To put the point in the language of powers, if an action is intentional then it is an act of a power whose *nature* it is to have self-conscious knowledge of its acts.

Two further comments on the notion of practical knowledge are in order. First, this sort of knowledge derives its name from the fact that it is the (formal) cause rather than effect of what it knows. We should therefore not be misled into thinking that the only application of the concept of practical knowledge is in the realm of practical reason, even though that is the context in which Anscombe introduces it. If I have been correct about how we know our own beliefs when these beliefs are acts of theoretical reason, our knowledge of them is equally “practical” knowledge, for it also belongs to the nature of theoretical reason to have self-conscious knowledge of its acts. In order to avoid the possibly misleading implication, we might therefore do better to borrow a bit of Kantian terminology and refer to what Anscombe calls practical knowledge as *spontaneous* knowledge.⁴²

⁴¹ Cf. Moran 2004, 54. Velleman misunderstands the sense in which practical knowledge is the cause of what it understands. He takes Anscombe’s view to be that intentions are beliefs that are reliable *efficient* causes of actions, and so embody knowledge of them (2007, xiv). This is no doubt a manifestation of the inability to properly comprehend practical knowledge that Anscombe ascribes to modern philosophy.

⁴² Velleman also refers to knowledge in intention as spontaneous knowledge of action (2008, 66-9). An advantage of this term is that—especially in its Kantian usage—it has overtones of “self-production” and is associated with reason. Kant also contrasts spontaneous knowledge with receptive knowledge, or knowledge due to affection by an object (*CPR* A19/B33). Kant’s notion of receptive knowledge is therefore very similar to Anscombe’s concept of contemplative knowledge.

Second, Anscombe's contention that modern philosophy has no notion of practical knowledge appears to be overly strong. A state or act that is practically (or spontaneously) knowable is a subjective state or act, for to say that something is subjective means that that thing does not exist independently of a subject's awareness of it. And some modern philosophers have at least recognized that one's psychological states and acts are subjective. Indeed, I argued above that our self-knowledge of our own acts of reason results from the fact that they are subjective. This may be a place where Anscombe is misled by her own terminology, which she borrows from Aquinas. For once it is clear that the kind of knowledge she is concerned with is not limited to action it also becomes clear that modern philosophy is not completely in the dark about it.

Still, it *is* characteristic of modern philosophy to confine the scope of the subjective to states and processes that are purely "inner", whereas our intentional actions include "outer" activities that involve bodily changes. The point at issue between Anscombe and modern philosophy would therefore be more aptly framed in terms of the extent of spontaneous knowledge; in particular, whether it is possible for the being of material processes that occur in publicly accessible space to be constituted by a subject's consciousness of them and thus for them to be spontaneously knowable.

The restriction on the scope of spontaneous self-knowledge characteristic of modern philosophy is clearly embodied in the Cartesian model of action, according to which one's actions or bodily movements are external to the actualization of practical reason. I believe it is for this reason that Anscombe says that knowledge in intention can only be understood once we have properly understood the nature of practical reasoning (§33). By her lights, Aristotle was the first to grasp what is distinctive about practical reason, namely, that it can *conclude in action*, but

his insight has been “obscured” by modern philosophy (ibid.). While understanding what it means to say that practical reasoning can conclude in action requires elucidation, we should already be able to see that if the Aristotelian doctrine is true then bodily action is not external to an agent’s reasoning.

I have now arrived, at least in outline, at a response to Grice’s worry that beliefs about action that are both non-observational and immediate would be epistemically defective. I agree that the will is a self-conscious power, so it is a legitimate *source* of self-knowledge. But I disagree about the *scope* of this epistemic power. If intentional actions are themselves conclusions of practical reasoning, self-knowledge of the will can include knowledge of action. Again, the motivation for holding this view is that it would make available a solution to the problem of bodily self-knowledge.

Our prospects for filling in this outline turn on re-acquiring a view of what Aristotle perceived clearly. The next two sections are therefore devoted to shedding some light on this matter. I will examine two accounts of practical reasoning influenced by Aristotle, those put forward by Davidson and Anscombe, and I will argue that only the latter does his insight justice.

3.2 DAVIDSON ON PRACTICAL REASONING

An action is intentional when Anscombe’s question “Why?” is applicable to it. What gives Anscombe’s question its special sense is that it asks for a reason for acting. By providing her reasons for acting an agent reveals the favorable light in which the action appeared to her, a

favorable light that makes it intelligible why she is motivated to act as she is.⁴³ In “Intending” Davidson argues that when an action is explainable by reference to the agent’s reasons, we can render intelligible the logical relations that hold between the action and the beliefs and pro attitudes that motivated it in terms of practical reasoning. Insofar as he aims to understand purposive action through seeing it as the upshot of practical reasoning Davidson is a follower of Aristotle. Furthermore, he approves of the Aristotelian idea that action can be the conclusion to a piece of practical reasoning, referring to it as “an important doctrine” (1971, 96).

Davidson is critical, however, of what he calls the “more or less standard approach” to practical reasoning, which he also associates with Aristotle. According to this approach practical reasoning is deductive. Suppose I see you eating a piece of candy. I ask you why you are doing so and you reply “Because it’s sweet”, thereby revealing the aspect of candy eating that appeared desirable to you. The proponent of the “more or less standard” approach theorizes that you arrived at the decision to eat a piece of candy from your beliefs and desires, the relevant contents of which can be given as follows (ibid., 86):

Any act of eating something sweet is desirable.

Eating this candy is an act of eating something sweet.

Therefore, eating this candy is desirable.

The two premises of this syllogism are supposed to express the contents of a desire you have to eat something sweet and a belief that eating this candy, which is something you are able to do, is

⁴³ These sentences review claims I discussed in the Introduction.

a way of eating something sweet. On the basis of these two premises you conclude that eating this candy is desirable, which is equated with a decision to eat it.⁴⁴

Davidson thinks this account of practical reasoning is inadequate. In particular, he is critical of the idea that a desire to do A corresponds to a judgment that an act of A-ing is desirable (along with the related idea, that an aversion to doing A corresponds to a judgment that an act of A-ing is not desirable). For suppose that you are now in a situation in which you want to eat something sweet, but you do not want to eat something poisonous, and you believe that a certain piece of candy is both sweet and poisonous. Given the above assumption about desire, you could conclude on the basis of this reasonable set of beliefs and desires that eating that piece of candy both is and is not desirable, and that is a contradiction (ibid., 98).

Davidson's alternative account of practical reasoning involves making a distinction between judgments of *prima facie* desirability and judgments of unconditional or "all-out" desirability. Desires to do A he takes to be or correspond to *prima facie* judgments: e.g., your desire to eat something sweet should be seen as being or corresponding to a judgment that an action is desirable *insofar as* it has that property, and similarly for your aversion to eating something poisonous. From the contents of these judgments one can arrive at the judgment that eating this candy is both desirable insofar as it is sweet and not desirable insofar as it is poisonous. And there is no contradiction in judging that an act of eating the candy has both

⁴⁴ In the first instance, the concept of deductive validity applies to an argument, which is a set of propositions, rather than to a piece of reasoning, which is a psychological process (Harman 1999). However, nothing stands in the way of speaking of a piece of reasoning as deductive in a secondary sense. We can say that a piece of reasoning is deductive just in case the reasoner takes it that the contents of the judgments that figure in her reasoning are assessable by the canon of deductive validity. The concept of a "conclusion" is also one that is applicable both to arguments and pieces of reasoning. The conclusion of an argument is the proposition that purports to be supported by the premises of the argument. But when a subject infers that p on the basis of contents of beliefs she takes to support the conclusion that p, we can also say that she *concludes* that p on the basis of the relevant beliefs or that she *draws the conclusion* that p. This last remark applies directly to theoretical reasoning. As will become apparent in the next section, I believe it must be modified when we turn to practical reasoning.

desirable and undesirable features. However, Davidson also thinks that *prima facie* judgments are not themselves directed at action. For that, one must arrive at a judgment that a certain type of action within one's power is all-out desirable, which on his view corresponds to, or is identical with, an intention (ibid., 99).

On the Davidsonian account practical reasoning consists in moving from one's desires and beliefs to an intention. This process involves weighing up various considerations that make courses of action under consideration desirable or undesirable in order to arrive at a verdict about which of them is all-out desirable. The contents of the *prima facie* judgments that correspond to one's desires rationally support, though they do not logically imply, the content of the all-out judgment made on their basis, which is why Davidson denies that practical reasoning is deductive.

This theory raises a number of concerns. Davidson identifies intentions with all-out desirability judgments on the basis of an alleged principle that it is not reasonable to perform an action unless one judges that it is unconditionally desirable. The mere fact that it has a desirable characteristic is not sufficient (ibid., 98). This principle looks dubious, though, if—as seems plausible—there are choice situations in which the courses of action under consideration are ones whose values are incomparable.⁴⁵ In these cases it can be reasonable to plump for one course of action over others even if it is not possible to make a judgment that the action one decides to do is all-out desirable. The identification of intentions and unconditional desirability judgments also makes it difficult to make sense of weakness of will. There seem to be cases in which an

⁴⁵ Two items are incomparable when no positive value relation, such as “is better than” or “is worse than”, holds between them. Incomparability is stronger than incommensurability. The latter relation holds when two items cannot be measured on a single scale with any precision, but a positive value relation can still hold between items that are incommensurable in this sense (Chang 1997, 1-4).

agent would deny he is executing an intention to perform an act he takes to be the most desirable one available to him.⁴⁶ But the feature of Davidson's account I want to focus on is its apparent incompatibility with the Aristotelian doctrine that action can be the conclusion of practical reasoning. Davidson speaks approvingly of this doctrine, but it does not come into view, either in the more or less standard account of practical reasoning or in his own.

I noted above that Davidson associates the more or less standard account of practical reasoning with Aristotle. This is because he holds that Aristotle believed practical reasoning is deductive (ibid., 97) and that its conclusion is a judgment that an action is desirable (1963, 9 n.7). Now, in the *De motu animalium* Aristotle does say that a practical syllogism, which displays a piece of practical reasoning, contains two sorts of premises, one of which relates to the good and the other to the possible (701a23-5). And some of Aristotle's examples of practical reasoning in the *De motu* could be made to approximate the more or less standard account. For instance, he imagines someone whose practical syllogism contains the two premises, "I should make something good" and "a house is something good" (701a17-18). However, the structure of another one of his examples, the "cloak-maker's" syllogism, is radically at odds with that theory:⁴⁷

I need a covering; a cloak is a covering. I need a cloak. What I need, I have to make; I need a cloak. I have to make a cloak. And the conclusion, the "I have to make a cloak", is an action. (701a18-21)

⁴⁶ Davidson's well-known attempt to deal with cases of weakness of will involves a distinction between all-out and all things considered judgments. He thinks that a weak-willed agent judges it is all-out desirable to perform a type of action she judges it is not desirable to perform all things considered (1969, 39-41). Like many others, I do not find this treatment of weakness of will satisfactory, but a critical analysis of this issue would take us significantly beyond the scope of this chapter.

⁴⁷ So too is the practical syllogism from *Metaphysics* Z 1032b7-10, among others.

Here we are to imagine a man who starts with the idea that he needs a covering, and reasons that a cloak will fulfill that need. The proposition he arrives at then serves as a lemma from which his reasoning continues. There is no hint that the man begins from a premise in which desirability is predicated of a certain type of action, and the reasoning is evidently not deductive. Most important, however, is the fact that Aristotle is explicit that the conclusion of the man's reasoning is not a judgment that making a cloak is desirable; rather, it is an act of *setting to and making a cloak*: "Now, that the action is the conclusion, is clear" (701a24).⁴⁸

Davidson writes that, at least in certain cases, nothing stands in the way of identifying the conclusion of practical reasoning with an action (1971, 99). There is a fundamental question about how seriously he can really mean this, though. For it looks like what he really *ought* to say is that on both the more or less standard account of practical reasoning and his own the conclusion of such reasoning is *a judgment that doing A is desirable*. Davidson's reason for thinking we can identify acting with drawing the conclusion of a practical syllogism is that "there is no distinguishing the conditions under which an agent is in position to infer that an action he is free to perform is desirable from the conditions under which he acts" (1969, 32). But this does not really help with the puzzlement we are likely to feel as to why we should follow Aristotle in referring to action using terms that are self-consciously derived from his logical treatises. The alleged fact that I would become disposed to look at my watch upon figuring out that it would satisfy my desire to know the time does not change the fact that it looks appropriate to say that the conclusion of my reasoning is the formation of the judgment *that* looking at my watch is desirable.

⁴⁸ Though of course we can also invent a form of words by which one could also express one's conclusion verbally. Aristotle appears to envisage the man expressing his conclusion with the words "I have to make a cloak".

I believe that when Davidson says that we can sometimes identify the conclusion of practical reasoning with an action he is merely paying lip service to the Aristotelian doctrine. Indeed, he admits as much himself, for despite his expression of approval of it he also writes, “The identification of ... action with the conclusion of a piece of practical reasoning is not essential to the view I am endorsing” (1978, 99). If we are to carry out the project of responding to Grice that I outlined at the end of Section I, however, we require a view on which such an identification is essential, at least in those cases in which the time for acting is at hand when one concludes one’s practical reasoning.

3.3 ANSCOMBE AND THE FORM OF PRACTICAL REASONING

In theorizing that practical reasoning is non-deductive reasoning whose conclusion is a judgment that a certain type of action is desirable, Davidson thinks he is being faithful to Anscombe (1963, 9n.7). However, this flies in the face of what Anscombe actually says in *Intention*. Her line in §33 is that the conclusion of practical reasoning is an action, at least in the cases she considers, and she means this quite literally. She calls practical reasoning one of Aristotle’s best discoveries, but laments that its true character has been “obscured”. In particular, Aristotle’s insight is obscured when practical reasoning is taken to be “ordinary reasoning” towards the truth of a proposition such as “I ought to do A” whose truth is purportedly proved by premises one believes (*ibid.*).

Davidson is right that Anscombe denies that practical reasoning is deductive. But her emphasis is not on the fact that practical reasoning is non-deductive, but on the fact that it is not reasoning towards the *truth* of a proposition about action. This is shown by her mocking of the

so-called “mince-pie syllogism”. The idea is that if what makes practical reasoning distinctive is simply that its subject matter is practical, there is no reason to even introduce the topic of “practical” reasoning as if it were its own species of reasoning in contrast with its “theoretical” counterpart, as Aristotle did. For then we could equally distinguish the mince-pie syllogism as embodying a distinct species of reasoning, its *differentia* being that it is about mince-pies, which is absurd. And this point holds regardless of whether the reasoning we envisage is deductive. The lesson Anscombe draws is that what distinguishes practical reasoning from theoretical reasoning is not a matter of its content but of its *form* (ibid.). We should therefore explore what it means to say that practical reasoning has its own formal character.

In her later paper “Practical Inference”, Anscombe calls the search for a geometrical construction “a pure example of practical reasoning” (117). In this respect, she follows Aristotle. One of his most sustained treatments of practical reasoning occurs in Book III of the *Nicomachean Ethics*, where he discusses it under the heading “deliberation” (*bouleusis*). He writes that deliberation is a form of “searching” (*zetēsis*) and he models deliberation on the activity of a geometer who analyzes a certain figure in order to construct it (1112b20-5).⁴⁹ This presupposes the audience has some familiarity with certain geometrical concepts, such as “construction” and “analysis”, and the full import of Aristotle’s analogy must be understood against that background. It will therefore be profitable to see how these concepts are deployed in the context of Greek geometry as codified in Euclid’s *Elements*.⁵⁰

⁴⁹ I here follow Ross and Cooper, the latter of whom writes, of the claim that what Aristotle is referring to in this passage is a problem of construction, that it “seems guaranteed by the reference to γένεσις (1112b24), a word that is not aptly applied directly to a proof except where construction is involved” (1984, 10 n.11).

⁵⁰ I need not suppose that Aristotle himself was familiar with this work. All I need is the weaker assumption that he was drawing on a stock of geometrical knowledge which was compiled and presented by Euclid, and this does appear to be justified by the text (see the previous footnote).

The *Elements* begins by setting forth a number of postulates, common notions, and definitions which are taken as starting points. After that there follows a series of propositions that are then shown to be derivable from these starting points. It will soon become clear, however, that the notion of “proposition” in play has the potential to be radically misleading to modern readers. The propositions themselves come in two different varieties: problems and theorems. An example of a theorem from Book I is the fifth proposition, the so-called “Pons Asinorum”, which runs as follows: *In an isosceles triangle the angles at the base equal one another, and, if equal straight lines are produced further, then the angles under the base equal one another.* This proposition is demonstrated in the body of the theorem, by showing how its truth follows from the starting points and from other propositions that have already been shown to follow from them.⁵¹

The reasoning employed in proving a theorem is deductive, which is a recognizable kind of ordinary theoretical reasoning. But besides theorems the *Elements* also contains problems. These are more relevant for us, for it is in connection with them that the idea of “construction” is employed. Let us therefore examine a problem and its solution more carefully. A typical example is the first proposition of Book I:

Book I, Proposition I

To construct an equilateral triangle on a given finite straight line

Let AB be the given finite straight line.

Thus it is required to construct an equilateral triangle on the straight line AB.

With center A and radius AB, describe the circle BCD (post. 3).

With center B and radius BA, describe the circle ACE (post. 3).

⁵¹ This is not to say that the text begins straightaway by setting out the starting points and theorems from which the proposition follows, for given only an isosceles triangle as it is described in the statement of the theorem it is not obvious how those starting points and theorems could be applied to demonstrate the required conclusion. The proof therefore begins by giving the reader instructions to draw certain other lines so as to produce a more complex figure, in order to get him in a position from which they are immediately applicable.

Find the point C at which the circles cut each other and draw a straight line from C to A (post. 1).

Draw a straight line from C to B (post. 1).

Then, ABC is the equilateral triangle required.

Since A is the center of BCD, AC is equal to AB (def. 15).

Since B is the center of BCE, BC is equal to BA (def. 15).

Therefore, AC is equal to AB and AB is equal to BC.

But things which are equal to the same thing are equal to each other (common notion 1).

Therefore, AC is equal to BC.

Therefore, the three lines AB, BC, and AC are equal to one another.

Therefore, the triangle ABC is equilateral, and it has been constructed on the given finite straight line AB.

The first thing that should strike us here is the statement of the proposition, which is not a statement of something for which truth or falsity could arise but is rather the specification of an action. In other words, it is a “proposition” in the more general sense of “something that needs dealing with”, and it is in this sense that both theorems and problems can be classified as propositions. What differentiates them is the manner in which each needs dealing with, and in the case of a problem dealing with it consists in performing an action of the required type.

The body of the problem is divided into two parts. The first consists of a series of specifications of further actions, arranged in the order that they are supposed to be performed. These are followed in the second part by a set of propositions, in the more familiar sense, whose function is to show that performing the actions in the order set out in the first part is sufficient to create the figure described in the statement of the problem.

This process of creation is what construction is. The geometer begins with only a compass, straight edge, and some means of drawing figures. The first three postulates state types of action he can perform using these devices: he can use the straight edge to connect two points, or further extend a given straight line a finite distance, and he can draw a circle with a given

center or radius using the compass. These are the sorts of actions that every student is supposed to know how to do upon commencing his geometrical studies. The steps of a problem then show one how to construct a new sort of figure by successive performances of these known types of action. The goal of the problem is not to teach one a bit of knowledge of something that is the case, but rather to teach one how to do something one did not know how to do before dealing with the proposition.

A construction contains the solution to a geometrical problem that has already been solved by earlier geometers. But Aristotle does not say that practical reasoning is like a student's following a construction that is already given. He says that is like a geometer's analyzing a figure in the course of *searching* for a construction. So let us imagine a geometer looking for a construction of an equilateral triangle along the lines of proposition I.1. He is sitting in his study with a straight edge and compass, and he has pen in hand. He has already drawn line segment AB and his goal is to construct an equilateral triangle with it as the base. He knows he needs to find a point C, such that $AC=AB$ and $BC=AB$. Employing his geometrical knowledge, he reasons that he could find a point C such that $AC=AB$ if he had a circle such that A was at its center and B and C both lay on its circumference. Likewise, he would have a point C such that $AB=BC$ if he had a circle such that B was at its center and A and C both lay on its circumference. Finally, he knows that if he placed C at the point where these circles intersect, this would result in lines AC and BC meeting at the same point. This would form an equilateral triangle since if $AC=AB$ and $BC=AB$ then $AC=BC$.⁵² Since our geometer knows how to draw the relevant circles and lines, he first draws circle BCD, then circle ACE, then line segment AC.

⁵² This process of breaking down a complex figure into the simpler figures that compose it is what an analysis of that figure consists in.

Finally, he draws line segment CB. Once he has drawn this final segment, he has accomplished his goal of drawing equilateral triangle ABC on a finite straight line.

What is it about this geometer's reasoning that makes it *practical*? The first thing to note is that it is situated in a distinctively practical context, insofar as he is looking to *do* something. What he is seeking to do is set by the problem he is attempting to solve, and the solution is not already at hand. We speak of the geometer doing some reasoning to accomplish his task because he uses his knowledge of facts about circles, lines, etc. to arrive at a way of solving it. But the conclusion of his reasoning is not the formation of a certain judgment. It is his *drawing* the relevant circles and lines in the appropriate order.

If we return now to *Intention*, we can find a similar understanding of practical reasoning at work in Anscombe's discussion of it. She provides an example which is supposed to be modeled on some of Aristotle's own illustrations. She gives us the contents of some of a certain man's beliefs, which he can use as premises in some reasoning, as follows (1963, §33):

Vitamin X is good for all men over 60.

Pigs' tripes are full of vitamin X.

I'm a man over 60.

Here's some pigs' tripes.

If the man were to reason with these premises, what conclusion could be rationalized by his knowledge of, or belief in, them? The answer is that it depends on the *purpose* to which they are put. For example, the man may be standing in front of a plate and wondering whether what is on it is good for him. In that case he could use his knowledge of the premises to arrive at a belief that what's on it is good for him, and the conclusion of his reasoning would be a belief

rationalized by his beliefs in the premises. But this would be a case of “ordinary” theoretical reasoning, not of practical reasoning.

The man engages in practical reasoning when, like the geometer we imagined above, he has a certain aim and he uses his knowledge of the premises to arrive at the performance of an action that would certainly or likely accomplish it. So suppose the man is at a buffet trying to decide what to eat, and that he has the objective of eating something good for him. He can then use his knowledge of the premises to arrive at an act of eating some of the food on a certain plate, which he identifies as “here”. In this case, his conclusion is not the formation of a certain belief, but rather his *act* of eating the pigs’ tripes, and this action is rationalized by his beliefs in the premises insofar as they show that eating some of the pigs’ tripes is a sufficient means of accomplishing his goal.⁵³

The fact that the man has the goal of eating something good for him means that he has a desire or pro attitude towards an action of this type, and this in turn implies that he thinks that an act of eating something good for him is desirable insofar as it has that property. But the way its desirability figures in his practical reasoning is quite different from the way Davidson theorized. According to Davidson, the first premise of a piece of practical reasoning includes a reference to a type of action and asserts *that* such an action is *prima facie* desirable. On the other hand, on Anscombe’s interpretation the first premise does not predicate desirability of a type of action, rather it mentions something the agent wants as an end. That the agent finds this thing desirable

⁵³ Practical reasoning as Anscombe understands it thus consists in putting means to a chosen end, where this is understood broadly enough so as to include both causal and constitutive means. This does not imply that it is impossible for agents to reason about which ends are desirable to pursue. It simply means that such practically oriented thought does not fall under the concept “practical reasoning”.

is not stated in the premise. It is *shown* by the fact that she is motivated to pursue it.⁵⁴ In this respect practical reasoning proceeds under the aspect of the good in the same sense that theoretical reasoning proceeds under the aspect of the true. When we specify the premises of someone who believes that p, because q and if q then p, we do not specify the content of the premises from which she reasons as “‘q’ is true” and “‘if q then p’ is true”. Rather, she reasons from “q” and “if q then p” and the fact that she concludes that p from these premises shows that she takes their contents to be true.

I remarked above that Anscombe claims that what distinguishes theoretical and practical reasoning is a matter of form. We are now in position to clarify how they differ in their formal characters. These two sorts of reasoning belong to a common genus insofar as they are both species of an activity that can be abstractly described as *the application of one’s rational faculty to the answering of a certain sort of question*. Where they differ is in the sort of question they are supposed to answer, and the nature of the answer they provide.

Consider first the activity of reasoning theoretically. This is the sort of reasoning that concludes with a judgment that something is the case. As Wittgenstein has pointed out, the very idea of grasping the sense of a proposition is tied up with the ability to understand a question of the form “Is it so?” The answer to this question must be either “yes” or “no” (*TLP* 4.023). To judge a proposition true is to answer the question whether it is so in the affirmative, while to judge it false is to answer it negatively, or to return an affirmative answer to its negation. A

⁵⁴ This also sheds light on Aristotle’s remark in the *De motu* that practical reasoning involves two different sorts of premises, one sort of which has to do with “the good” and the other with “the possible” (see Section II). He does not mean that it is essential to the practical syllogism that one of its premises states that something is good and that another states that doing a certain type of action is a possible way of bringing it about. What he means is that it needs to include a premise that asserts that a certain action or state of affairs has a characteristic that the agent finds intrinsically desirable, even if it does not state that she finds it desirable. The role of the other kind of premise is to show that the action the agent performs is a possible way of performing the act or bringing about the state of affairs she finds intrinsically desirable.

subject engages in the activity of reasoning theoretically when she has grasped the sense of a certain proposition but has yet to decide whether things are as it represents them as being. In other words, she aims at returning an answer of “yes” or “no” to the question “Is it so?” Her reasoning then consists in working out the answer based on other things that she knows or believes, and her judgment constitutes her answer to this question.

The act of judgment may thus be said to be an affirmation or denial in response to a certain theoretical question. It is therefore noteworthy that Aristotle understands acting as analogous to an act of judgment in this respect: “What affirmation and denial are in the case of [theoretical] thought, pursuit and avoidance are with desire” (*NE* 1139a21-2). For rational animals acting is like making a judgment insofar as it is the employment of reason in answering a certain sort of question. But the question with which practical reasoning is concerned is not the question what to think but *what to do*. This is not a yes/no question, in the primary sense, but it can be seen as one in a secondary sense. Practical reasoning occurs when the agent has an end and she needs to figure out how to accomplish it. In any deliberative context, there will be a range of possible actions that are intelligible candidates for performance. This range may expand as an agent imaginatively considers different possibilities and contract as she excludes certain actions that on reflection she realizes she cannot perform. Acting is saying “yes” to some of these possibilities and “no” to others in response to the question what to do in order to pursue her end.

It may have been noticed that I have so far spoken of the will only in the context of acting intentionally. I have not yet said anything about intention for the future. In this respect, I have followed Aristotle. Davidson accuses him of neglecting intention for the future and as a result saying things that “sound fatuous” (1978, 96). But I think it is more probable that Aristotle’s

omission is simply due to the fact that his discussions of practical reasoning are very compressed, and he thought it obvious how his account could be extended to cover cases where the time for acting is at some temporal distance from the deliberating agent. To see how this might be so, take Davidson's own example of writing the word "action". If one intends to write this word then while he is writing "a" he must intend to write "c" a moment thereafter, followed by "t", "i", "o", and "n". This example brings to the fore the fact that most of the actions we perform have some duration, so that in order to perform them intentionally an agent must possess the capacity to do things in a certain sequence and the capacity for memory. Thus, to affirm writing the word "action" as the conclusion of practical reasoning is to affirm writing the letter "a" and to become disposed to write the rest of the word thereafter. But these are the same capacities that are required in order to be able to form an intention for the future. For an agent who possesses them can conclude a present deliberation with not doing anything now but becoming disposed to do something at the appropriate future time, so that when the time comes and if memory does not fail she will do it intentionally then. Intention for the future poses no new conceptual difficulties for understanding practical reasoning, or at least none that are relevant for my purposes in this chapter.

3.4 DE-MOTIVATING THE INFERENTIAL THEORY

The reason for examining practical reasoning was to make intelligible what it would mean to say that intentional action is not "external" to the process of practical reasoning. Recall that to say that a certain entity is not external to a rational process means that we can understand its

production as itself a step in that process. If acting intentionally is the conclusion to a piece of practical reasoning, it will not be external to that reasoning.

We are now in position to finish the reply we began in Section I to Grice's objection that beliefs about what we are doing not formed on the basis of observation or inference would necessarily be epistemically deficient, no better than wishful thinking. This objection overlooks the fact that there is a significant difference between our beliefs about our intentional actions and beliefs formed on the basis of wishful thinking. In paradigm cases of the latter, beliefs are formed in such a way that any connection they have to the truth is purely accidental. Indeed, this is precisely what makes it the case that such beliefs are epistemically deficient. But to assimilate immediate, non-observational knowledge of what we are intentionally doing to wishful thinking is to overlook the fact that such knowledge is not receptive but *spontaneous*. When one is aware of what one is doing through having chosen to do it, one's awareness of one's action is an implication of the proper exercise of the self-conscious capacity of practical reason, whose actualization it is. This ensures that there is a non-accidental connection between one's belief about what one is doing and the fact that one is doing it. In this respect, our special epistemic access to our own intentional actions is analogous to our self-knowledge of our own beliefs.

Our beliefs about what we are doing intentionally are thus different from beliefs that are held on the basis of wishful thinking in an epistemically significant way. If there is a further and more general reason why even non-accidentally true beliefs can never amount to knowledge when they are not formed on the basis of sufficient evidence Grice has not told us what it is. Furthermore, as we saw in Section I, he is not in a position to make such a claim. This is

because the inferential model that he endorses presupposes that an agent is entitled to the non-observational and non-inferential belief that she wills to perform a certain action.⁵⁵

Although I think we should reject the inferential account of knowledge in intention, I also think it is important to defuse the motivation that lies behind it. We can put the motivating thought in the form of an objection: Our intentional actions include things like opening a window and painting a wall yellow. So if I know I am engaged in these activities then I must know, e.g., that a window is going up and that a wall is becoming yellow. And surely I could not know these things are happening without observation or inference. This is what I take to be the source of the urge to push spontaneous knowledge of the will back to a purely interior and psychological act.

Anscombe deals with an objection with this same shape in *Intention*, and her response is instructive. Her central move is to concede that we can normally know in intention what we are doing intentionally only in a context in which we also have some observational or inferential knowledge. But she denies that it follows that knowledge in intention is *itself* observational or inferential. In a crucial passage she writes:⁵⁶

When knowledge or opinion are present concerning what is the case, and what can happen—say Z—if one does certain things, say ABC, then it is possible to

⁵⁵ Velleman also says that the fact that our beliefs about our intentional actions are non-accidentally true helps to explain why they have positive epistemic status (2007, 62-4). However, for him this claim follows from a more general thesis that one is justified in forming a certain belief just in case it would be supported by sufficient evidence once formed (*ibid.*), whereas I have not endorsed any such general thesis. My argument has been more piecemeal, consisting in three points. First, these beliefs are instances of a kind of self-conscious belief that is already acknowledged by all parties as being in epistemically good shape. Second, the fact that self-conscious beliefs of this kind are spontaneous removes the most obvious reason for thinking that they are not in epistemically good shape. Finally, because of the first two points, the burden of proof falls on someone who thinks that immediate and non-observational beliefs about what we are doing could not constitute knowledge to say why there is a *further* problem in their case. Skepticism about them cannot rest merely on the fact that they are not formed on the basis of observation or inference.

⁵⁶ The important role that this passage plays in responding to objections about the very idea of knowledge of action not grounded in observation or inference is also noted by Moran (2004, 49).

have the intention of doing Z in doing ABC; and if the case is one of knowledge or if the opinion is correct, then doing or causing Z is an intentional action, and it is not by observation that one knows one is doing Z; or in so far as one is observing, inferring etc. that Z is taking place, one's knowledge is not the knowledge that a man has of his intentional actions. (§28)

Observational and inferential knowledge is often required for knowledge in intention because one must know, or at least believe, certain facts in order to so much as *try* to do some act Z. For instance, one cannot try to open a window by pushing it up unless one knows or believes that it is possible to open it by applying upward pressure. And knowledge that this is possible can be had only on broadly empirical grounds. Once this knowledge is in place, though, one can choose to open the window by pushing it up, and when one intentionally does open it in this way one will have spontaneous knowledge in intention of one's action under the description "opening the window" through having reasoned to it.

Empirical knowledge is required not only to try to do something but also to *sustain* doing an intentional activity that takes any substantial length of time. For example, if one is to intentionally paint a wall yellow for longer than a few moments, one must have some observational knowledge that one can keep painting it yellow and that the paint has not worn off one's brush. Again, though, if one has knowledge that one can continue painting the wall yellow by moving the brush across it, one can know one is actually *painting* it yellow without observation or inference.

The case of acquiring observational knowledge that one is still painting the wall yellow also highlights that the fact that one can normally know what one is doing intentionally does not imply that one cannot also have observational knowledge of this same action. Anscombe herself makes this point in the final clause of the last sentence in the passage quoted above. Knowledge by observation and knowledge in intention are two different *ways* of knowing, and nothing

precludes a rational agent's knowing a single happening in each of these two ways. However, unlike observational knowledge, spontaneous knowledge in intention is *self-conscious* knowledge of a happening as one's doing. I might see a certain human being painting a wall, for example, and I might know on this basis that I am painting it. But insofar as it was observationally acquired, this knowledge would not be self-conscious self-knowledge. The occurrence of the first person in my belief about myself would be an instance of its use "as object". On the other hand, if I was intentionally painting the wall, then I would know my action under that description, as *my* painting the wall, in intention. And insofar as this knowledge was an instance of knowledge of the actualization of my will, it would be a case of knowledge of myself "as subject".⁵⁷ This last point is crucial. It means that knowledge in intention is a form of self-consciousness of oneself as a bodily creature, and thus a way of overcoming the problem of bodily self-knowledge.

3.5 COMPARISON WITH COGNITIVISM

The goal of the preceding sections has been to provide an account of knowledge in intention according to which it is non-observational and non-inferential. The theory I have arrived at is not the only one available that has this feature. It is shared by a rather different theory of knowledge in intention, which goes under the label "cognitivism about practical reason". Versions of cognitivism have been recently defended by Velleman and Setiya. My goal in this

⁵⁷ For the distinction between the use of the first person "as object" and "as subject" see Chapter 2.3.

final section is to compare cognitivism with the “Aristotelian” account I have offered. I will suggest that the latter emerges looking more attractive.

Cognitivists believe that when a rational agent acts intentionally her action is caused by her intention, and that intentions are beliefs about one’s actions, or at least psychological states with a belief-like direction of fit. Therefore, forming an intention is coming to be in a psychological state that both represents oneself acting and causes one to perform the represented type of action.

Velleman’s theory of intention occurs within the context of a larger and ambitious account of the nature of agency and the standards of practical rationality. His central contention is that practical thought is a region of theoretical thought. This means both that the norms that govern practical deliberation are the norms of theoretical reasoning, and that intentions are beliefs. Specifically, “Intentions to act...are expectations that issue from reflective theoretical reasoning. They are self-fulfilling expectations of acting that are adopted by the agent from among potentially self-fulfilling alternatives because he prefers that they be fulfilled, and they represent themselves as such” (2007, 98). Velleman hypothesizes that intentions are rendered self-fulfilling by special desires to know what one is doing and why that are constitutive of rational agency (ibid., 27-32). These desires normally restrain an agent from acting until she has formed a belief about what she is going to do, at which point they move her to perform the relevant action.⁵⁸ Because her beliefs about what she is going to do are self-fulfilling, an agent can still see them as representations of what is the case even if she does not form them on the

⁵⁸ In his more recent work, Velleman has preferred to refer to his hypothesized aims for self-knowledge as “drives” or “tendencies” rather than desires. The reason is that the term “desire” carries the connotation of an attitude that embodies a personal-level, conscious goal whereas he thinks the tendency for self-knowledge might operate sub-personally and sub-consciously (cf. 2000, 19; 2007, xx).

basis of sufficient evidence (ibid., 128). Furthermore, the rational force that certain considerations bear in practical deliberation is a function of how well they satisfy the desires for self-knowledge and self-understanding (ibid., 264).

Velleman's account faces several different sorts of problem, of which I will mention two.⁵⁹ The first concerns his theory of what it is to be a practical reason. He contends that a consideration counts as a reason for action just in case it would fulfill the aim of rational agency, which is self-understanding. Velleman says that the norms for constructing an understanding of oneself are those of theory construction more generally, e.g., generality, coherency, and simplicity (ibid., 273). But despite his suggestion to the contrary, if this were true it seems *prima facie* extremely implausible that what we would have most reason to do would be anything like what we normally suppose. After all, a pure egoist seems to have a self-conception that is extremely general, simple, coherent, and that provides a basis for precise predictions about her future actions. She would therefore seem to meet the relevant criteria just as well as, if not better than, someone who is committed to justice and benevolence in addition to prudence about her own future well-being.

Second, Velleman's theory of intention faces what Michael Bratman calls "the problem of promiscuity" (1991, 257-61; cf. Setiya 2007, 109-10). If Velleman is right then whenever a person comes to believe that he will do A he should acquire at least some motivation to do it. However, this is not the case: if I carelessly step off the sidewalk and realize that I'm about to be hit by oncoming traffic, I do not *ipso facto* acquire even the slightest motivation to be hit by it.

Setiya's theory of intention is meant to avoid both these problems while retaining a commitment to cognitivism. According to him an intention is a *sui generis* mental state that is

⁵⁹ For a third difficulty see Clark 2001.

both belief-like insofar as it represents oneself as acting and desire-like insofar as it tends to motivate one to make its content true (2007; 2008). Since Setiya builds motivation into intentions themselves he does not need to posit special desires to know what one is doing and why that will ensure one acts on one's intentions. The fact that he is not committed to the claim that one will in general be motivated to perform actions one believes one will perform means he can avoid the problem of promiscuity. Nor need he endorse the idea that practical reasons are considerations that would promote the fulfillment of such desires.

On Setiya's account, to act intentionally is to be motivated to act by a desire-like belief about the psychological causes of one's action. So, for instance, when an agent does A intentionally for the reason that p her action is caused by the desire-like belief that she is acting partly because she believes that p. I say "partly" because Setiya thinks intentions must also represent themselves as a partial cause of the actions they motivate (2007, 45). This is due to the fact that an agent might act because she believes that p without acting for a reason or on the basis of an intention. So for an agent to perform an action for the reason that p that action must be caused by the desire-like belief that she is acting both because she believes that p and because of that desire-like belief itself. Assuming that there is some story to be told about why these desire-like beliefs are warranted when they are formed, the fact that intentions are such beliefs explains why agents have self-conscious self-knowledge of their intentional actions.

How should we evaluate Setiya's theory? One mark against it is simply that, when compared to the Aristotelian account it is bound to appear overly complicated and even Ptolemaic. Setiya gives us an explanation of why knowledge in intention should belong to rational agency, but only with the help of unnecessary and dubious additions. His theory requires, as part of what it is to act intentionally, that agents form belief-like representations of

the psychological processes which they undergo when they so act. It may be true that rational agents typically know what they are doing under the descriptions under which their actions are intentional, and that they normally know the reasons for which they are acting. But it is much more questionable to contend that what *explains* how they know this is that intentional action is essentially action motivated by beliefs about its psychological genesis. The capacities involved in the ability to be conscious of the psychological explanations of one's actions go well beyond the capacity to be conscious of the descriptions of one's actions under which they are intentional. These additional capacities are also ones that require the mastery of extremely sophisticated concepts, namely, the concept of intention as a desire-like belief and the concept of the kind of non-deviant causality involved in rational explanations.⁶⁰ The Aristotelian account, by contrast, does not set overly demanding requirements on the capacity to act intentionally. On that account, the only concepts that an agent needs to do A intentionally are the ones specifying what it is to do A, which one acquires in learning how to do it.

Second, as I stressed in Section I, we often have self-conscious self-knowledge not only of our own actions but of our beliefs, and we can add to this attitudes such as anger, regret, and certain emotions. What these have in common is that they are motivated by the reasons for having them, and this in turn suggests that knowledge in intention is simply one form of the more general phenomenon of self-conscious rational activity. While the Aristotelian makes it perspicuous how knowledge in intention is a special instance of self-conscious rational activity that is also applicable to theoretical reasoning, Setiya's cognitivism does not. It would be wrong

⁶⁰ Wedgwood notes we routinely credit young children with the ability to act intentionally even before they have reached a stage where it is plausible to credit them with having such sophisticated concepts (2008, 408). The point goes even further, though, since adults do not appear to invoke their mastery of intention and psychological causality when they say, decide what to say next in the course of a conversation.

to suggest that theoretical believers are self-conscious believers because when someone forms a belief that *p* she is caused to form it by a desire-like belief that specifies the beliefs she regards as reasons for *p*. This would entail that we form our beliefs intentionally, which is false.⁶¹ While Setiya's version of cognitivism is an improvement over Velleman's, the Aristotelian account is still to be preferred. It is both simpler and better illuminates the way in which self-consciousness is bound up with rationality.

Of course, even if I have been right so far about the superiority of the Aristotelian metaphysics of action, our acceptance of it still depends on its being applicable to rational agents like us. But the worry might arise that the fact of our finitude means that for us there must be exactly the sort of gap between thought and action that this theory is meant to bridge. It may be conceded that the world would always conform to the desires of an omnipotent agent. We are not omnipotent, though, and as a result it is possible for us to fail to do what we set out to do. We sometimes try and succeed in moving our bodies, and other times we try and fail. Does this not show that in both cases our reasoning concludes in an act of trying, which may or may not cause a bodily movement? In addition, our finitude ensures there will be cases in which we act intentionally without being able to know whether we are succeeding. Since the will is an essentially self-conscious power, do cases of this sort not prove that the will's acts must be purely psychological ones whose occurrence we infallibly know?⁶² The purpose of the next chapter is to formulate these objections more carefully and to offer replies.

⁶¹ In his recent paper "Knowledge of Intention" Setiya attempts to show that his theory is consistent with there being a general connection between rationality and self-consciousness. The contention I am making is not simply that one's theory of knowledge in intention must respect this general connection, but that it is plausible that practical and theoretical self-consciousness are two forms the same phenomenon takes in different domains.

⁶² In addition, someone may think that the fact that we sometimes form an intention for the future which we never act on shows that there is also a "gap" between intending and doing. The remarks on intention for the future in

4.0 TRYING, BODILY AGENCY, AND THE LIMITS OF PRACTICAL KNOWLEDGE

*Between the idea /And the reality
Between the motion /And the act
Falls the Shadow. –T.S. Eliot,
“The Hollow Men”*

In the previous chapter I defended Anscombe’s thesis that as rational agents it belongs to human beings to have a special mode of epistemic access to their intentional actions. We can describe knowledge in intention negatively by saying that it is knowledge of action not acquired on the basis of observation or inference. But we can also describe it in positive terms as *practical* or *spontaneous* knowledge, knowledge that is the cause of that of which it is the knowledge. The sort of causality at issue here is formal causality, which means that an action is not intentional unless it is an act of the will or practical reason, a power whose nature it is to have self-conscious knowledge of its acts. Intentional action is not external to the will’s activity insofar as it can be the conclusion to practical reasoning. I argued that this Aristotelian conception of self-conscious agency is superior to what I called the Cartesian model, according to which an agent’s immediate knowledge of what she is doing intentionally is limited to purely “inner” acts such as intentions or tryings, because only the former helps to provide an answer to the problem of bodily self-

Section III already contain a response to this sort of worry. There is no more a gap between intention for the future and intentional action than there is between a disposition and its realization.

knowledge. The Aristotelian conception also provides a simpler account of knowledge in intention than cognitivist accounts of intention, and it renders more perspicuous the connection between practical self-consciousness and its theoretical counterpart.

Yet as T.S. Eliot vividly reminds us, our actions are not always successful. Anscombe may be right that failure to execute intentions is necessarily the exception rather than the rule (1963, §48), but it still occurs more frequently than we would like. The fact that we are not omnipotent means that failure is an ever present possibility. In this chapter I discuss the implications of our finitude for the metaphysics and epistemology of intentional action. In the *De motu animalium* Aristotle notes on at least two occasions that reason (or *phantasia*, which plays an analogous role in non-rational animals) will be efficacious in moving the parts of a creature's body only if that creature is not prevented from performing the action hit on as the conclusion of its practical reasoning (701a16; 702a17). But he does not appear to think that this fact poses any threat to the account of self-movement he there puts in place.

There are two lines of thought that purport to show that Aristotle moves too quickly, however. Both begin from the fact that it is possible for us to fail to accomplish what we set out to do and conclude on this basis that the Aristotelian account of rational agency cannot be true. The first is an argument for volitionalism, or the thesis that events that are describable as an agent's performing a bodily action are acts of trying that occur prior to her bodily movements. In Section I, I set out the argument for volitionalism, while in Section II I show that this argument is invalid. The alternative to volitionalism I put in place is a version of what Brian O'Shaughnessy calls the "dual aspect" theory of the will, and Section III I discuss this feature of it. In Section IV I turn to the second of the two arguments against the Aristotelian account of self-conscious agency. This argument proceeds from the claim that because we are finite we can

act intentionally without *knowing* that we are succeeding, but knowing only that we intend to act and are trying to do so. It concludes that acts of the will must be purely psychological acts such as intentions and tryings. Again, I think that this argument is not valid and I show why in Section V.

I will argue that despite their differences both arguments suffer from the same defect. Both fail to understand the significance of the concept of fallibility. If I am right, the mere fact that we are finite does not impose special constraints on how we must understand the nature of practical reason in us that would not hold if we were considering an omnipotent rational agent. The right lesson to draw, rather, is that our lack of omnipotence means that in us this power is not infallible in certain respects.

4.1 THE ARGUMENT FOR VOLITIONALISM

According to Anscombe, modern philosophy's inability to understand knowledge in intention is bound up with its inability to properly grasp what is distinctive about practical reasoning. Specifically, it has lost sight of Aristotle's insight that this sort of reasoning can *conclude in action* (*Intention*, §33). By contrast, in "Practical Reasoning" John Broome denies that Aristotle's claim about practical reasoning is an insight at all. He writes:

Aristotle took practical reasoning to be reasoning that concludes in an action. But an action – at least a physical one – requires more than reasoning ability; it requires physical ability too. Intending to act is as close to acting as reasoning alone can get us, so we should take practical reasoning to be reasoning that concludes in an intention. (2002, 83)

Insofar as he takes it that the furthest out, so to speak, that practical reasoning can take us is to an intention, Broome is a paradigmatic modern philosopher.

All should agree that in order for one's practical reasoning to be effective when it is directed towards bodily action, one must possess the appropriate physical abilities. But why should we think that in that case reasoning ability cannot itself *include* physical ability? After all, if Aristotle is correct, then a human being's conclusion to a piece of practical reasoning can consist in her exercising her knowledge how to engage in some physical activity. If Broome's assertion is to be more than simply question-begging, a reason is needed for thinking that physical ability is extrinsic to reasoning ability.

Broome does not provide such a reason, but in his commentary on the quoted passage Nishi Shah does. Shah endorses Broome's conclusion and adds, "If I intend to raise my arm now to ask a question but my arm suddenly becomes paralyzed or I become paralyzed by fear, my deliberation has not failed, although I have" (2008, 4). Clearly, I would not have succeeded in raising my arm in this scenario, but the crucial question is how we should describe the way in which I would have failed. According to Shah, I formed an intention as the conclusion of my practical reasoning, but having been struck by paralysis I did not do anything else. Why should this tell us anything significant about how to describe what happens in cases in which my action is successful? If reflection on the unsuccessful case is to have any bearing on how we describe the successful one then this can only be because it tells us something about intentional agency in general, whether successful or not. Shah's thought must therefore be something like the following: the unsuccessful case shows that practical reasoning always concludes in an intention. The difference between that case and the successful one is that in the successful case the intention I form causes an event that makes true the proposition that I move my body.

This argument is objectionable, though. Shah assumes that when I am unknowingly struck by paralysis I only form the intention to act and do nothing else, and this assumption is false. Even if I am suddenly paralyzed at the moment I am supposed to raise my arm, I still *try* to raise it. But trying to do such-and-such is something that an agent actively engages in. In the proposition “S is trying to do A” we are able to modify the verb by certain adverbs of action. For instance, one might try very hard to raise one’s arm, or one might try to raise it attentively or inattentively.

The considerations that Shah adduces do not support the conclusion that practical reasoning can get us no closer to action than intention. However, the style of argument that he goes in for still constitutes a challenge for the Aristotelian conception of practical reasoning, even if the nature of that challenge must be reconceived. For we need only substitute “trying” for “intending” and if we find this kind of argument compelling we will be led to the conclusion that practical reasoning always concludes in an act of trying. The difference between the successful and unsuccessful cases would then be that in the former but not the latter my trying has an effect, a bodily movement, that makes it true to say that I not only try to move the relevant portion of my body but that I actually move it. In other words, we will be pushed towards *volitionalism*, the theory that intentional actions are causally prior to bodily movements.⁶³ The truth of volitionalism is incompatible with the truth of the Aristotelian conception of rational agency, for according to the Aristotelian conception when an intentional action is describable as an agent’s moving her body, her bodily movement is itself included in the action rather than being an effect of it.

⁶³ “Volitionalism” is perhaps not the most felicitous label, but it is the one that McCann and Smith use to describe their position. Lowe (2008) also puts forward a theory he calls “volitionalism”, but it differs from the theory under consideration here. Lowe believes that acts of will, or volitions, cause happenings such as bodily movements, but he does not identify the cause of these happenings with an action (148).

As I will understand it volitionalism is the conjunction of the following two claims: (1) all physical actions can be described as tryings and (2), these tryings are related to bodily movements as cause to effect. It follows from (1) and (2) that actions are the cause of bodily movements. Furthermore, since trying is a psychological concept, it also follows that actions are psychological causes of bodily movements. Volitionalism in contemporary philosophy is inspired by H.A. Prichard (1945), and has been defended by A.D. Smith (1988), Hugh McCann (1975), and by Jennifer Hornsby in her 1980 book, *Actions*.⁶⁴

Two points of clarification are in order. First, volitionalism is a thesis about physical action, or action that involves bodily movement. We are therefore setting aside cases of mental action such as calculating a sum in one's head. It is not that such cases hold no importance for the philosophy of action. They do not bear on the present issue, though, which concerns the relationship between the psychological and physical aspects of physical action. Second, it is important to be clear about the sense of the term "action" that appears in (1). As Hornsby points out, in referring to someone's action one might be referring either to something that an agent does or to the event of her doing it. The first of these refers to a universal that can be repeatedly instantiated by different agents on different occasions and the second refers to a non-repeatable spatiotemporal particular (1980, 3). Since volitionalism is a thesis about the nature of particular events or processes, it is the second of these senses of "action" that is intended.

⁶⁴ Hornsby has since renounced volitionalism. According to McCann, whenever an agent tries to do A, her trying must consist in some other action, whether it be A-ing itself or something that she does as a way of doing A (1975, 102-3). Smith and Hornsby place no such condition on trying. As we shall see, there seem to be cases where an agent tries to move but without making any physical movements at all. Smith and Hornsby can say that in such cases the only way of describing the agent's action is as a trying to move, while McCann must say that her trying consists in something else, which he proposes is the mental action of willing or volition (ibid., 108). In what follows I will assume the view of trying held by Smith and Hornsby, but all of the arguments could be recast to fit McCann's conception without loss of substance.

If volitionalism is true, then whenever anyone does A intentionally their action can be described as their trying to do A. This claim is not obviously true, and in fact it may seem to make trying much too ubiquitous. Wittgenstein writes, “When I raise my arm I do not usually *try* to raise it” (1958, I §622). Wittgenstein’s thought accords well with our normal use of the verb “to try”. We would not normally say of someone that he tried to do something unless either he was unsuccessful or he needed, or expected that he would need, to make some special effort to do it. But while such usage tells us something about the conditions under which it is normally *appropriate to say* that someone is trying to do A, this does not by itself settle when it is *true* that an agent is trying to do A. Whether it is appropriate to utter a proposition in a given context is determined by more than just its truth-condition, since some true propositions have non-semantic implications that are highly misleading in that context (Grice 1961). Therefore, its being inappropriate to say that S is trying to do A if S is not failing, having difficulty succeeding, or expecting to have such difficulty may be due to some non-semantic implication of its utterance.

The volitionalist has an argument that is designed to show that whenever one does A intentionally one tries to do it.⁶⁵ It proceeds by pointing to particular cases that show the truth of “S is trying to do A” does not depend on conditions the holding of which renders it appropriate to say that S is trying to do A. The recipe for generating such cases is to construct a situation in which two conditions hold: first, one person S believes that she will do A without difficulty but another person E has an excellent ground, g, that S does not have and that if true would defeat S’s confidence. Second, though E is justified in believing g it turns out that g is false and S does A without difficulty. In such a situation E would think that S would try to do A and fail, but E would still speak truly if when S A’s she said that though she was wrong to think S would fail,

⁶⁵ This argument can be traced to O’Shaughnessy 1974.

she was at least right in thinking that S would try to do A. As Hornsby emphasizes, the point of the argument is not that for any agent S and any action A, we could imagine someone with the requisite false ground and justified belief, but rather that the class of cases for which we can do this shows something about trying *überhaupt* (1980, 37-8).

So much for (1), what about (2)? The volitionalist's argument for (2) presupposes that someone can try to move her body even though her body completely fails to move. This assumption is a plausible one. McCann notes, for instance, that a recovering paralytic who is uncertain whether he has regained use of his arm could still comply with his physician's order to try and raise it even if, as things turned out, the limb was still completely immobile. Indeed, the paralytic would learn that it was still immobile precisely *because* he tried to raise it and was unsuccessful (1975, 99).

We will allow, then, that one can try to move one's body even when one's body fails to move. This raises the question what the difference is between cases in which one tries to perform some bodily action and fails and cases in which one's trying succeeds. McCann seems to think that the volitionalist conclusion follows directly from the fact that the difference between these cases is one of whether things are allowed to take their normal course (*ibid.*, 108-9). But what is required is an argument to the effect that things taking their normal course should be understood in terms of causation of bodily movement by an act of trying.

It is a virtue of Smith's article "Agency and the Essence of Action" that in it he provides just such an argument. Smith connects his argument with what he refers to as Anscombe's and Davidson's approach to action individuation, contending that volitionalism is its "logical conclusion" (409).⁶⁶ I say "what he refers to as" this approach because though Anscombe and

⁶⁶ Hornsby claims, for similar reasons, that volitionalism is the "logical continuation of Davidson" (1980, 20).

Davidson are often seen as speaking with one voice on the topic of the individuation of actions there are actually important differences between them. What they share in common is that they both think we can often refer to a single particular action under many specifications, and that some of these specifications describe it by reference to one of its effects. But there are also at least two ways in which they diverge. First, in his illustration of the approach Davidson typically uses as examples of actions processes that have reached their completion, which are referred to using the perfective aspect (1971).⁶⁷ By contrast, when she speaks of actions Anscombe almost always has in mind processes that are still under way, which are reportable by the progressive aspect. The two therefore often mean different things by the use of the word “action”. If volitionalism is true it should apply to both of these kinds of processes, so from here on out we will take the class of actions, in the sense of the particulars that are multiply specifiable, to include both completed doings and doings that are still in progress. For instance, S may *have moved her finger* in order to flick a switch, and she may *have flicked a switch* in order to turn on the light. But though S *did* at least three things here, there was only one action that was her doing and it can be described variously as her moving her finger, her flicking the switch, and her turning on of the light. It may also be the case that S *is moving* her finger in order to flick a switch and thereby turning on the light, and here what she *is doing* is describable as moving her finger, flicking a switch, and turning on the light. In both cases, S’s turning on of the light is so describable because it is also describable as a flicking of a switch, and the switch’s being flicked causes the light’s going on, while it is a switch flicking because it is also a finger moving, and the finger’s movement causes the switch to be flicked.

⁶⁷ A fact noted by Thompson (2008, 136, n.17).

The second divergence between Anscombe and Davidson is that the latter claims that descriptions of an action that do not specify it as an agent's making bodily movements always describe it by reference to an effect, whereas the former does not commit herself to any such proposition. This is noteworthy because Smith includes the additional Davidsonian claim in his description of the approach to action individuation he wishes to push to its "logical conclusion". What is plausible is that although some of the act-specifications that are true of a certain action specify it by reference to an effect, not all do. But Davidson takes it that there is a certain *kind* of act-specification that describes actions as they are in themselves, namely, the kind that describes them in terms of agents moving their bodies.⁶⁸ One reason for this turns on the conceptual truth that an action is essentially an agent's activity together with certain temporal considerations. Consider again someone's turning on the light. If her action would have continued to occur after she moved her finger, this would entail by the conceptual truth that her active contribution also would have continued after she moved her finger. But after she moved her finger it was unnecessary for her to do anything else; she'd made her contribution. Whether or not the switch was flicked and the light went on was not a matter of her activity but of the obtaining of certain conditions in the mechanism of the switch and the electrical circuit. Davidson puts the point by saying that everything that happens after we move our body is "up to nature" (1971, 59).⁶⁹

Smith argues that Davidson is wrong to think that in describing a bodily action as an agent's moving her body we are not describing it by reference to an effect, and he thinks that our

⁶⁸ In fact, he also makes the stronger claim that every physical action is identical with a bodily movement (1987, 102). I think that the volitionalist is right to dispute this stronger claim. It will become apparent why in the next section.

⁶⁹ Davidson in fact misstates his own point. What he actually says is, "We never do more than move our bodies: the rest is up to nature". But here Davidson has confused "action" in the sense of something agents do with "action" in the sense of the event of an agent's doing something. The point he ought to make is that though agents do many things, e.g., flick switches and turn on lights, all of these descriptions of things that we do describe events of our moving our bodies by reference to their effects.

earlier considerations about trying support this conclusion. According to Smith, what distinguishes those specifications that characterize action by an effect is that their truth depends on the obtaining of a condition that is independent of the agent's being active. A condition is "independent" in the relevant sense just in case its failure to obtain would not imply that the agent was not active (1983, 405). However, Smith also thinks that to describe an action as an agent's moving her body is still to describe it in a way that is dependent on the obtaining of an independent condition. For instance, when I move my finger, it moves in part because certain causal connections hold between the finger and the muscles whose contractions cause it to move, and between those muscles and my nervous system. These causal connections are independent conditions because even if, unbeknownst to me, they had failed to hold I would have still been active insofar as I would have still tried to move my finger. So just as an act of flicking a switch might not have been an act of turning on the light if certain conditions in the circuit had failed to hold, so too my trying to move my finger might have failed to be an act of moving my finger had certain conditions in my muscular or nervous system failed to hold.

Smith concludes that the very same considerations that lead Davidson to the conclusion that certain non-bodily act-specifications characterize actions by reference to their effects should also lead us to think that to describe an action as an agent's moving her body is to describe it by reference to an effect of her trying. To parrot Davidson: all we have to do is try, the rest is up to nature. And since an agent tries to act every time she acts intentionally, there is no reason why the scope of these considerations should be limited to anything less than physical action as such. This yields (2), the second part of volitionalism.

4.2 FAILURE AND FALLIBILITY

Volitionalism does not have the consequence that practical reasoning cannot conclude in action. But it is incompatible with the Aristotelian doctrine, according to which physical intentional actions include bodily goings-on. For my purposes I need only show that plausible considerations about the nature of trying do not force us into volitionalism. But I think it will be useful to first point out two ways in which volitionalism is problematic in its own right. The fact that my alternative also avoids these difficulties should lend it additional support.

First, volitionalism has unsatisfactory epistemological implications. If actions are causally prior to bodily movements, then they are, at least typically, wholly invisible to the naked eye. But intentional actions are not normally wholly invisible. We can often know at least some of the things that others are doing intentionally by observation (Lowe 1981). Hornsby's reply is that we might see physical actions "in virtue" of seeing other things, such as their effects or the place where they occur (1980, 103). John Hyman rightly counters that we no more see events in someone's nervous system by seeing their effects than we see the pistons of a car moving in virtue of seeing the car move. And we no more see events in the nervous system by seeing the place they occur than we see someone walking in the park by seeing the park where he is walking (2006, 159).

Volitionalism also fails to do justice to the phenomenology of embodied agency. According to the volitionalist, our powers as agents are distinct from the capacities of our bodies to move, and the actualizations of the latter capacities are not our doings but their effects. Consider, by way of analogy, what happens when someone grasps certain items by operating a metal claw. The agent has the power to grasp things with the claw, and she has this power in virtue of the fact that she does something else, such as manipulating a controller, which causes

the claw to exercise its capacity to grasp. If volitionalism is correct, then the relation between an agent and her body is similar to that between the agent and the claw. This is problematic insofar as the latter is a paradigm case of a relation between an agent and an instrument or tool, and it would be wrong to think of our bodies as related to us in a similar manner.⁷⁰

The objectionable element of volitionalism lies in its claim (2) that the relationship between action and bodily movement is a causal one. This thesis is independent of both the claim that physical actions can be described as tryings and the claim that an agent can try to move her body without there occurring any bodily movement. These other claims are not pernicious by themselves and indeed, I think that the arguments we have seen on their behalf are compelling. My focus in what follows will therefore be on Smith's argument for (2).

His argument hangs on the alleged principle that act-specifications characterize an action by reference to an effect when their truth depends on the obtaining of an "independent condition", or a condition such that its failure to obtain would not imply the agent was not active. Smith thinks this principle shows that Davidson is right in saying that all act-specifications that go beyond describing actions as the making of bodily movements describe them as the effect of these movements. What Davidson failed to realize, on his view, is that the same principle also entails that to specify an action as the making of a bodily movement is to specify it by reference to an effect. However, even if we bracket Smith's distinctive application of this principle for a moment, there appear to be counterexamples to using it even to go as far as Davidson's original

⁷⁰ It is this kind of thought that lies behind McDowell's remark that if we adopt a model of agency on which action withdraws inwards, prior to bodily movement, then the result is a picture on which the acting subject is alienated from her body (1996, 90). Some philosophers hold that even when we do things like raise our arms (in the normal way) we do so *by* doing something else, namely willing (cf. Lowe 2008, 147). However, I believe that the purpose of saying this is simply to mark out the fact that one's arm-raising is an exercise of one's will, and not, e.g., a spontaneous reaction. When one intentionally raises one's arm in this way, one does not do so by exercising one's knowledge how to do something else, which is why this sort of case contrasts with that of manipulating the claw.

conclusion. For example, consider my carrying a backpack to class. The description of my action as an act of carrying a backpack depends on the obtaining of an independent condition, namely, the pack's being slung over my shoulders. The condition is "independent" in Smith's sense because had the pack failed to be slung over my shoulders, I would still have been active in walking to class. But it is not the case that the backpack's getting carried is an effect of my walking. Rather, when I walk with a backpack over my shoulders, my walking *constitutes* my carrying a backpack. It thus also cannot be the case that the description of my action as one of carrying a backpack describes it by reference to an event of a backpack getting carried that it causes.⁷¹

It may be that act-specifications whose truth depends on independent conditions often characterize actions by reference to their effects. But we have seen that this is not true in general. It is therefore possible to accept (a) that when an agent moves her body, the movement of the relevant bodily parts depends on the holding of certain conditions having to do with, e.g., her muscles and her nerves, and (b) that these conditions are such that had they failed to hold, the agent would still have been active, and yet reject Smith's conclusion. One way to do so would be to say that the connection between an agent's bodily activity and the events that occur in her muscles and nerves when she engages in it is a constitutive one.⁷²

In fact, there is an alternative to volitionalism that is both appealing in its own right and avoids the problems we have raised for that theory. On this account, as an embodied agent I have powers to move my body in certain ways. As powers to move my body they are fully exercised only when the relevant bodily parts move, but they may also be only partially

⁷¹ Similar cases are discussed by Hornsby (n.d.).

⁷² This is true even if the relationship between the action and these events is otherwise very different than the relation that holds between walking and carrying a backpack.

exercised if something interrupts, interferes with, or otherwise prevents me from fully exercising them. This explains why if my muscles fail to contract or my nerves fail to conduct the appropriate electrical signals due to paralysis, I am still active. But on this alternative it does not follow that, when such preventing factors are absent, my activity is not constituted in part by my bodily movements. The difference between the two cases is not one of whether my activity has certain effects but whether it is completed or prevented from reaching its completion.

Another way to put the point is that our powers for bodily action are *fallible*. In general, a power is defined by its proper acts, but if the power is fallible in a certain respect then it is not necessarily the case that whenever it operates the one who possesses it does what it is defined as a power to do in that respect. When one exercises the power, one will act in the appropriate way only if its operation is not impeded or inhibited.

An action that is an agent's trying to move her body is a process that is extended in time. As such, it is composed of various other events and processes that, having been bound together into a temporally unified whole, constitute the parts or phases of that whole. Earlier phases of this process involve the occurrence of certain events in the motor cortex and its later phases include events in the peripheral nervous system, contractions of muscles, and finally, the movement of certain parts of the body. When all goes well the agent's trying to move is successful and the action which at first could only be characterized as her trying to move can also be described as her moving her body. But if for some reason it is not successful, e.g., because the body is prevented from moving, then it can be described only as her trying to move.

The upshot is that one must be careful to distinguish the relation that obtains between an agent's action and her bodily movements from the relation that obtains between the sub-personal events that occur in her nervous and muscular systems when she acts and her bodily movements.

The latter relation is a causal one, but the former is not and it is the mistake of the volitionalist to think that it is. An action, which we are agreeing may be described as an agent's trying to act, does not cause the movement of her body. Rather, the movement is the final part or phase of the action whose earlier phases include events in her motor cortex, peripheral nerves, and muscles.

What unifies the different phases of action into a single process? The answer is that the unity is a *teleological* one, for the phases are related as elements of a single teleological order. To understand the nature of this order, it will be useful to contrast it with a teleological order of a different sort, namely, the sort that can make a number of things that an agent is doing intentionally into a single complex whole that she is also doing intentionally and whose phases they constitute. This is the sort of teleologically constituted unity that Anscombe draws attention to in *Intention* and that Michael Thompson has recently discussed under the heading of "naïve action explanation" (2008, 90-1). As Thompson explains, a person may be doing something such as building a house which he does by mixing mortar, laying bricks, and nailing together pieces of wood. Here the simpler tasks that make up the parts of the whole are parts because they are all directed at or for the sake of the whole. Furthermore, what makes it the case that they are so ordered is the fact that the agent who is doing them conceives of them as so ordered. It is *because* the builder thinks that mixing mortar, etc. are means to his end of erecting a house that when he mixes the mortar his action is also describable as a phase of the process of house-building. That he thinks of his actions as so ordered is in turn something that he can manifest by his ability to say that he is mixing mortar and the like in order to build a house.

When a human being intentionally moves her body, the process of her moving it is also composed of a teleologically ordered series of phases, but the ground of their unity is different in kind from the one highlighted by Anscombe and Thompson. For example, if I try to raise my

arm and certain muscles contract, then their contraction is no accident; rather, they contract in order to cause my arm to rise. More generally, the events that take place in a person's nervous and muscular systems when she tries to move her body are directed at or for the sake of the movement whose occurrence would make it the case that her trying is successful. However, that they are so directed is not due to any fact about how that person conceives them: the principle of their ordering is not thought but nature.⁷³ The sort of unity that obtains between the sub-personal events that constitute an action is thus comparable to the sort that holds between the phases of vital processes such as digestion. Digestion is the process whereby food is broken down into a state in which it can be absorbed by the body, and has as phases the ingestion of the food, the action of certain enzymes upon it, and its chemical decomposition. These phases form parts of a single process because they are all for the sake of getting some ingested food into a state suitable for absorption by the body.

The capacities that human beings have to willfully move their bodies are like their capacity for digestion. The possession of both of these sorts of capacities is a part of the nature of the being whose capacities they are. Since they belong to that nature, when they are actualized and the processes that constitute their actualization reach their completion, their reaching this completion is no accident. But in addition to having powers to intentionally move their bodies, human beings also possess powers to cause other sorts of happening in virtue of being located in an environment with a regular causal order. Because we inhabit such an environment, we also have powers to cause other sorts of things to happen, powers we acquire through being in physical contact with portions of it. Anscombe's "small selection" of causal verbs in her essay "Causality and Determination" provides a sample of some of the further things

⁷³ The contrast between thought and nature is not meant to imply a denial that in another sense thought itself is a natural process in the lives of human beings.

we can do: “*scrape, push, wet, carry, eat, burn, knock over, keep off, squash, make* (e.g. noises, paper boats), *hurt*” (137).

4.3 A DUAL ASPECT THEORY

The theory of bodily action I endorse is a version of the dual aspect theory that O’Shaughnessy puts forth in his book *The Will*. A theory of action is dual aspect if it holds that physical intentional actions, which are describable as acts of bodily powers, can also be characterized as psychological acts of trying. In the course of his argument for volitionalism, Smith objects to O’Shaughnessy’s theory and one might think that I would also be a target of Smith’s objection.

O’Shaughnessy contends that we must factor physical action, which itself admits of both physical and psychological descriptions, into a part that is purely psychological and a part that is purely physical and non-psychological. In particular, we must say that an action is composed of a bodily movement that is non-psychological and a “non-autonomous psychological part”, where to say that this part is non-autonomous is to say that it is a phase of the action and not a separate, causally-related event. This is because he thinks that if a certain event is psychological then it is necessarily so. But it is not the case that a particular bodily movement that occurs when someone willfully moves her body is necessarily psychological (1980, 2:207-8). Consider, for instance, a case in which a person tries to raise her arm, but dies after the nerve signals have been sent out from her brain and before her arm goes up.⁷⁴ In such a case the rising of the arm—or

⁷⁴ This assumes, of course, that we can make sense of her dying as a process whose duration is brief enough to allow the thought-experiment to make sense, and though one may question whether this is in fact possible, I will let it stand for the sake of argument.

better, the rising of the matter that made up her arm when she was alive—would not be a psychological event.

Smith thinks that O'Shaughnessy is committed to the implausible view that an action, which is essentially an agent's activity, is itself a causally related series of intrinsically non-active events. He considers an unsuccessful action T that fails to include a bodily movement because a certain circumstance C fails to obtain. Smith argues that if C had obtained, then according to O'Shaughnessy what we picked out as T would have caused a bodily movement and thus would have been a non-autonomous psychological part of a different action T' of which the bodily movement was a non-psychological part. But since non-autonomous parts of actions are not themselves actions, whether some process is an action or merely a non-autonomous psychological part of an action is dependent on the circumstances (1988, 418).

Smith's argument depends on the assumption that O'Shaughnessy is committed to the claim that had C failed to obtain, T would have been a cause of a bodily movement in T', and as such would not itself have been an action. But this is incorrect; all O'Shaughnessy is committed to saying is that, had C obtained, what we picked out as T would have been a *part* of an action T' that also had a bodily movement as a non-psychological part. It would therefore not be a non-active event even in that case.

I therefore think that Smith's objection does not touch O'Shaughnessy's version of the dual aspect theory. Nonetheless, I think that it is more in the spirit of what I've argued so far to say that in better circumstances, some process we pick out as an unsuccessful trying in the actual world would have been *itself* a successful trying—a physical action—rather than merely a non-autonomous psychological part of one. O'Shaughnessy is barred from holding this because of his assumption that being psychological is an essential property of psychological processes. This

claim is not argued for, however, and I do not see why there is any reason to accept it. Psychological processes are a species of psychical processes, in Aristotle's sense of processes engaged in by living beings, and the counterpart of O'Shaughnessy's assumption does not hold for other psychical processes. Consider again the process of digestion. Digestion is a nutritive process in living things and is composed of various mechanical and chemical processes wherein food is ingested and decomposed. It is also possible that a creature should die after ingesting some food but before its digestion has taken its course, and that some of the chemical processes that constitute digestion in creatures of its kind take place after its death. Even so, it would be wrong to say that these chemical processes constitute digestion in the creature's corpse. Only living things have the capacity to nourish themselves by digesting food. Therefore, in absence of an argument otherwise, I see no reason for not taking a similar line when it comes to intentional bodily action. When someone successfully tries to raise his arm and his arm rises, he fully exercises a psychological power possessed by human beings. Here it is not that a part of his action is psychological and another physical; rather, the whole process of his raising his arm is *psychophysical through and through*. But if he had died before his arm went up, then at the time of death he would have lost the power to intentionally raise his arm. So even if some nerve signals had gone out before he died and happened to jolt the matter that used to compose his arm into rising, this rising would then be a purely physical event and not a psychophysical one.⁷⁵

Both O'Shaughnessy's version of the dual aspect view and my own avoid the problems that arose for volitionalism. First, since physical actions are processes of which bodily movements are parts, they do not occur inside our skins and are not wholly invisible to

⁷⁵ Another difference between O'Shaughnessy's theory and mine is that whereas he conceives of intention as one of the "act-progenitors" that causally precede willed action, I agree with Anscombe in thinking that we should not conceive of intention as a causal intermediary between reason and action.

observation. Second, we stand in a very different relation to our own body than we do to a tool or instrument. When we use an instrument, our action gets to be describable as an operation of it in virtue of the fact that we do something else we know how to do, such as move our bodies, which causes it to operate. But we do not move our bodies by performing an act that causes our bodies to move. When we try, e.g., to raise our arm, the rising of our arm is not an effect of our action but is the final stage of it. Our bodily powers are not powers we have in virtue of being in contact with our environment. They are among the powers we have in virtue of our nature.

4.4 FAILURE TO KNOW WHAT ONE IS DOING

The volitionalist begins with the undeniable premise that we sometimes fail to do what we set out to do and infers from it that when our tryings are successful our actions are causally related to the movements that we make. I have argued that we need not accept this inference, and that there are good reasons for resisting it. In this section I will examine a second argument against the Aristotelian conception of practical reason. Though the argument has a very different shape from the argument for volitionalism, it too proceeds from the fact that we are not omnipotent.

In Chapter 3 I said that the will is essentially a self-conscious power. A self-conscious power is one that provides one with knowledge of oneself as oneself. I also argued that the acts of the will include intentional actions, so that the self-conscious knowledge that the will provides can include knowledge of oneself as acting. Finally, I said that this is a good way of stating one of the central theses of Anscombe's *Intention*, and I contrasted this conception of the will with what I called the Cartesian model. The proponent of that model agrees that the will is a self-

conscious power, but she insists that the will's acts are restricted to purely psychological acts such as intentions and tryings.

The argument I want to examine is one for the Cartesian model of action. It begins from the shared premise that the will is by nature a self-conscious power. Its second premise is that it is possible for an agent to be doing such-and-such intentionally without knowing that she is doing so, but merely knowing that she intends to do such-and-such and is trying to do so. The conclusion inferred from these premises is that acts of the will must be purely psychological acts of intending and trying.⁷⁶

Evidence for the second premise includes Davidson's well-known case of a man signing a certain document with the intention of making ten carbon copies of his signature (1978, 50). Ten papers being a thick stack, the man could not be said to know that his signature is showing up on all ten sheets as he signs the document. But it is possible he is successfully impressing his name on all the papers, and if he is then he is making ten copies of his signature intentionally, even though he does not know that he is. He knows only that he intends to write his name and that he is trying to do so.

If we are sympathetic to the idea that we normally know what we are doing intentionally without observation, this argument may be thought to rule in favor of the inferential view of knowledge in intention. We encountered this view, which is associated with Grice, in Chapter 3. Recall that according to the inferential view if one has knowledge in intention that one is doing A this knowledge is the conclusion of an argument, one premise of which is that one wills to do A and the other that one tends to do what one wills to do under certain conditions. The

⁷⁶ The argument could also proceed by asserting in its second premise that it is possible for an agent to know what she is doing intentionally only by relying on observation or inference. But it is most vivid if we focus on cases in which it looks plausible that an agent does not have any knowledge of what she is doing, as I do here.

inferential view can explain both why a rational agent often knows what he is doing intentionally and why Davidson's carbon copier lacks such knowledge. When one knows what one is doing intentionally, it is because one has a warranted belief that there are no features of one's circumstances that would prevent one's tendency to do what one wills to do from realizing itself. One can therefore rationally infer that one is doing what one wills to do. On the other hand, the man's past experience with carbon paper does not warrant the belief that his signature is showing up on all ten papers. Therefore, he cannot rationally infer from the fact that he wills to impress his signature on every paper in the stack that it is actually appearing on all of them. He is only warranted in inferring that there is a certain probability that this outcome is occurring.

Though both the argument for the Cartesian model of the will and the argument for volitionalism turn on the possibility of failure, they differ in two ways. First, the sort of failure which was in question in the argument for volitionalism was the failure to do certain things that one attempts to do, whereas the sort of failure that the proponent of the present argument points to is a failure to *know* what one is doing. Second, the argument for volitionalism focused solely on cases where the difference between success and failure was constituted by whether certain bodily movements occurred, while in the present argument the difference between them is not restricted in that way. It seems possible to do things like write one's signature on a piece of paper without knowing one is doing so.

Similar arguments to the one I have outlined in this section have been made in the literature on *Intention*, but my way of stating Anscombe's central thesis differs from the one usually attributed to her. I flag this now because it will be important in what follows. Whereas I have argued that her thesis can be put by saying the will is a self-conscious power of action, she has been interpreted by several commentators as claiming that necessarily, if S does A

intentionally, then S knows she is doing A (Setiya 2007, 25-6; Paul 2009). Counterexamples such as Davidson's carbon copier are supposed to show that this cannot be correct. In response, Setiya suggested that we weaken the claim to say that if S does A intentionally, then there must be something she is doing intentionally in the belief that she is doing it (ibid). But he now recognizes that even this claim can be defeated by counterexamples, at least if we limit ourselves to descriptions of S's action that characterize it as the successful performance of some bodily activity (2008, 390). For instance, consider a scenario in which a patient who is recovering from paralysis in her hand is asked to try and clench her anaesthetized fist behind her back. If prior to clenching the patient has only medium confidence that she has recovered, then even if her clenching is successful she will not know that she is clenching her fist though she is doing so intentionally. Furthermore, the patient is not successfully doing anything else intentionally in the knowledge that she is doing it.

Thompson attempts to defend the claim that one necessarily knows what one is doing intentionally from these alleged counterexamples. His response to Davidson's carbon copier case rests on the fact that claims of the form "I am doing A" are cast in what linguists call the imperfect aspect, which is expressed in English by the progressive present tense. The progressive has a certain property, often referred to as its "broadness", such that someone may be doing something in the broad sense at a time when she could not truly be said to be doing it in a narrower sense. This allows him to object that Davidson's copier does not in fact constitute a genuine example of someone who is doing something intentionally without knowing it (n.d., 12). Suppose the signature of the carbon copier is not appearing on all ten sheets of paper on his first attempt. Thompson points out that this fact does not show that the copier does not know he is making ten carbon copies. When he is finished with the first attempt he would simply check and

if necessary take out the sheets of carbon paper with his signature on them and try again. Even if he must repeat this procedure several times, it is still true that the entire process constitutes his making ten carbon copies and he knows throughout the process that he is doing so. It is in the broader sense that Davidson's carbon copier is making ten copies even if his signature is not appearing on all ten sheets of paper the first time he signs his name.

There are other counterexamples to the claim that we always know what we are doing intentionally, however, which cannot be dissolved by the broadness of the progressive. First, suppose that it is the carbon copier's intention to make ten copies of his signature *on this very attempt*.⁷⁷ Then, if he makes ten copies of it on that attempt he is doing so intentionally, without knowing he is. Second, notice that in the case of the recovering paralytic even if we take into account the broadness of the progressive it would be wrong to say that the patient knows she is clenching her fist if it turns out that it will take several more days or even weeks for her to recover. The progressive may tolerate some degree of broadness, but a temporal interval of this length between attempt and execution is too long for it sensibly to apply, at least for this sort of activity.

Thompson's reply to such examples is to deny that they are instances of the phenomenon that interests Anscombe. His thought is that the concept normally expressed in English by the phrase "doing such-and-such intentionally" includes more than what Anscombe talks about under that heading (*ibid.*, 12). However, I see no evidence that Anscombe believes she is discussing anything other than the will, the power that is at issue when ordinary English speakers speak of intention and intentional action. Thompson's suggestion therefore seems both desperate and uncharitable.

⁷⁷ This is arguably the sort of case that Davidson himself had in mind.

What emerges from the examples we have looked at in this section is that the second premise of the argument for the Cartesian model is a strong one. If we wish to avoid the conclusion of that argument we will have to dispute its validity rather than the truth of its premises.

4.5 THE WILL AS A FALLIBLE EPISTEMIC POWER

We should grant that it is possible for an agent to be doing A intentionally without knowing she is successfully doing A, or any other bodily action. Indeed, we can see that if an Anscombian conception of knowledge in intention is correct, then such cases are bound to arise. In Chapter 3 I argued, following Anscombe, that knowledge in intention of what one is doing is only available in a context in which one also has empirical knowledge, or at least true beliefs, about the possibility of accomplishing one's ends in one's circumstances.⁷⁸ For instance, in order to have knowledge in intention that one is opening the window one must know, or at least truly believe, that the window will move up if one applies upward pressure to its base. This in turn requires knowledge that the window is unlocked and not rusted shut. It will sometimes be the case, however, that one will only be in a position to know or believe that acting or trying to act in a certain way is more or less likely to produce a certain result. In such circumstances, one will know at best only, e.g., that if one does A one will *probably* do B, or if one tries to do C one *stands a good chance* of doing it (or at least *not too bad a chance*). In these circumstances one won't be able to know that one is actually succeeding at doing B or C without being able to

⁷⁸ Chapter 3.4.

observe the happenings or states the obtaining of which would make it true that one is B-ing or C-ing. Therefore, if (1) one is not in a position to acquire empirical information about such facts, if (2) conditions are such that Thompson's point about the broadness of the progressive is inapplicable, and if (3) one's action is successful, *then* one will perform an intentional action without knowing one is doing so. We have seen it is possible to construct examples in which all these conditions are met. The important task is to determine the significance of such cases for our understanding of knowledge in intention.

Recall that the argument for the Cartesian model of the will has two premises. The first is that the will is an essentially self-conscious power, while the second is that it is possible for an agent to be doing A without knowing that she is doing so, but merely knowing that she intends to do so and is trying to do so. The proponent of the argument then infers from these premises that acts of the will must be purely psychological acts of intending and trying. But even if we grant both the premises there is still room to take issue with the conclusion. This is because the inference from the premises to the conclusion will be a good one only if the following principle is true: if one's power to do A is a self-conscious power then necessarily, when one does A intentionally one knows one is doing A. Given our earlier discussion of fallibility, this principle ought immediately to strike us as dubious. I pointed out that although powers are defined in terms of their proper acts, if a power is fallible in a certain respect then it is not necessarily the case that whenever it operates the one possessing it does what it is defined as a power to do in that respect. This occurs when the relevant aspect of its operation is impeded or inhibited. The fact that we sometimes fail to know what we are doing intentionally therefore cannot by itself prove that we lack a self-conscious power to act intentionally. At the most it shows only that this power is not infallible, and that it is fallible insofar as it is a self-conscious power.

Compare our power to acquire knowledge of the environment by way of vision. I can know that things are thus-and-so by seeing them to be so. In order for my judgment that they are thus-and-so to constitute a piece of knowledge, I must also know that I see them to be so. I cannot know that they are so if my belief that I see them to be so is a mere guess, even if it turns out to be true. Moreover, my capacity to visually perceive that things are thus-and-so is a self-conscious capacity. While it is not part of the nature of visual perception to yield knowledge of oneself as oneself insofar as one is an *object* of perception, one's capacity to perceive visually is self-conscious in the sense that it provides one with knowledge of oneself as oneself as the *subject* who is perceiving. I know on the basis of seeing that things are thus-and-so that I see them to be so, and when I form the belief that I see them to be so in this way my belief is both groundless and immune to error through (*wh-*) misidentification.⁷⁹

Yet it can sometimes happen that I think I see that things are thus-and-so when they are not. To take a classic example, I may mistakenly think I see that there is a bent stick in front of me because I am looking at a straight stick that is partially submerged in a tank of water. In this situation, both my belief that there is a bent stick in front of me and my belief that I see a bent stick would not amount to instances of knowledge. Nonetheless, we need not take this scenario to prove the skeptical conclusion that I do not have a self-conscious power to see that things are thus-and-so. Rather, it shows only that the power in question is fallible, and in two respects. First, my power is fallible insofar as it is a power to see that things are thus-and-so. Second, it is fallible insofar as it is a self-conscious power, or a power to *know* I see that they are so. In particular, sub-optimal viewing conditions which inhibit my power to see that things are thus-

⁷⁹ The argument that a mark of a belief's being an instance of self-conscious self-awareness is its immunity to error through (*wh-*) identification relative to the first person pronoun is found in Chapter 2.3.

and-so also inhibit my power to know that I see they are so. But this does not contradict the claim that when viewing conditions are more favorable, I both see that things are thus-and-so and know that I do.⁸⁰

Similarly, we need not conclude, from the fact that there exist conditions under which we would do A intentionally without knowing that we are doing A, that the will is not a self-conscious power of action. It is open to us to conclude instead that this power is fallible insofar as it is a self-conscious power. In fact, that our wills are fallible is just what we should expect, given that we are finite rational agents. Like our power to see that things are thus-and-so, our self-conscious power to act intentionally is fallible in two respects. First it is not the case that whenever one exercises her power to do such-and-such intentionally, one is guaranteed to succeed in doing such-and-such. That is the conclusion I drew from our discussion of volitionalism. Second, it is not the case that when an agent does such-and-such intentionally she is guaranteed to know that she is doing so. Conditions which potentially inhibit the success of one's doing such-and-such intentionally may also inhibit one's power to know that one is doing so when the action is successful. When such conditions are absent, however, one knows what she is doing immediately and without observation.

In *Intention* §6 Anscombe says that an intentional action is one that is susceptible to the question "Why?", in the sense that asks for a reason for acting, and that this question is rejected by the response "I was not aware I was doing that." In §8 this condition is strengthened so that the response "I knew I was doing that, but only because I observed it" is also counted as a rejection of her question. In these passages she seems to be saying both that it is a necessary truth that if one does something intentionally one knows one is doing it, and that it is a necessary

⁸⁰ For more on the role of fallibility in the epistemology of perception, see McDowell 1982, Rödl 2007, 146-58.

truth that one has non-observational knowledge of what one is doing intentionally. It is therefore significant that she also says things that counteract this appearance. As an alternative way of making her point in §8 she says that intentional actions are a sub-class of the class of things that one knows without observation. In the course of giving other instances of this class she writes “E.g., a man *usually* knows the position of his limbs without observation” (italics mine). And when she returns to the topic of non-observational knowledge in §28 she says that when one knows the position of one’s limbs, one’s knowledge *ordinarily* does not rest on a sensation. This makes room for the thought that there can be a connection between a certain kind of state of affairs’ obtaining and a person’s non-observational knowledge that it obtains without that connection being absolutely necessary. While both of these passages concern a different species of non-observational knowledge than knowledge in intention, there is no hint that what she says about them is not supposed to be applicable to the whole genus.

If this is true, does it destroy the intimate connection between action and knowledge that Anscombe clearly means to assert? No. Aristotle often discusses what happens *hôs epi to polu*, a locution that is usually translated as “typically” or “for the most part”. Thus certain fish “for the most part” conceive at a certain age (*HA* 560b19-20) and teeth “typically” come to be by a certain process (*Physics* 198b34). The point of Aristotelian science is not to study mere statistical regularities, though. It is to discover natures or essences. But creatures are not always situated so that their natural powers and tendencies are manifested, for the actualization of these capacities may be blocked or interfered with. In other words, the phrase *hôs epi to polu* marks the fact that these natural powers are fallible ones.⁸¹ However, the fact that the powers are fallible does nothing to impugn the fact that certain acts characterize the natures of those powers.

⁸¹ For my understanding of Aristotle’s use of *hôs epi to polu* I am indebted to Moravcsik 1994.

I suggest that we hear Anscombe's uses of "usually" and "ordinarily" as versions of this Aristotelian locution. If this is right then when she remarks that an action's being intentional rules out its agent's saying "I was not aware I was doing that" or "I knew I was doing that but only because I observed it", we should understand her to have in mind the central cases of the phenomenon she is investigating. These cases are the ones in which conditions are suitable for the nature of the power in question to be manifest. And if the Aristotelian account I have defended is correct, it belongs to the nature of intentional actions, insofar as they are acts of practical reason, to be self-consciously knowable by their agents.⁸²

The fact that it is possible for one to do something intentionally without knowing one is doing it does not undermine the idea that the will is essentially a self-conscious power whose acts include actions. Rather, it highlights the fact that this power is fallible insofar as it is a self-conscious power. This possibility of failure does not show that knowledge in intention is unfit for resolving the problem of bodily self-knowledge.

⁸² Thus the fact that the will is fallible in this respect does not contradict the claim that there is a *formal* connection between intentional action and knowledge. The formal connection exists because an action is not intentional unless it is an act of the will, a power whose nature it is to have self-conscious knowledge of its acts. This is true of every intentional action, even those which, because of some impeding circumstance, are not known to their agent, or are only knowable by observation. Even so, I can acknowledge that in one respect the will is infallible *qua* self-conscious power, for whenever we do such-and-such intentionally that act is an act of trying and we know it as such. The respect in which the will is fallible *qua* epistemic power is that it is sometimes the case that one does not know one's trying can be described as a *successful* trying.

5.0 REDUCTIONISM AND THE METAPHYSICS OF AGENCY

“I should not say of the movement of my arm, for example: it comes when it comes, etc. And this is the region in which we say significantly that a thing doesn’t simply happen to us, but that we *do* it.”—
Ludwig Wittgenstein, *Philosophical Investigations*, I §612

I have now completed the task set in Chapter 2, of making it plausible that we have self-knowledge of ourselves as bodily beings, by defending the idea that knowledge in intention is a form of self-knowledge. The challenge was an epistemological one, but the central argument of Chapter 3 was that in this region of philosophical inquiry epistemological questions are inseparable from metaphysical ones. In particular, in responding to Grice’s objection to the very idea of knowledge of action that is not based on either observation or inference I claimed that knowledge in intention should be seen as the sort of immediate knowledge that rational beings characteristically have of the exercises of their rational powers. Furthermore, in replying to an argument for volitionalism in Chapter 4, I argued that when one does A intentionally one is truly describable as trying to do A. Since the concepts “acts of reason” and “trying” are psychological ones, one way of expressing the conclusion of these chapters is that intentional action is psychological in nature. This should not be taken to imply that such actions are purely “inner”, however. On the contrary, it is often the case that acting intentionally includes changes of

configurations of one's body. Intentional action is not merely a psychological activity, but is an irreducibly *psychophysical* one.

Yet there is still a risk that this way of stating the upshot of these chapters will seem puzzling to many contemporary philosophers. For according to a common and entrenched view in the philosophy of mind what it is to be an intentional action is to be a bodily movement that is produced in the right way by certain psychological states or events. If this is right, then there is nothing specially psychological or psychophysical about intentional action. When I act intentionally, the bodily movements that occur are no more psychological occurrences than the one that occurs when someone takes hold of one of my limbs and moves it. The difference between these cases is not an intrinsic one but a matter of their respective etiologies.

The entrenched view can take different shapes depending on what psychological states and events are specified as causally necessary. But however it is spelled out it remains *reductionistic*, in the sense that it tries to provide a satisfactory theory of the nature of intentional action in terms of causal relations between bodily movements and various psychological states and happenings.⁸³

I believe there are grounds for rejecting reductionism that are independent of providing an elucidation of knowledge in intention, however. In his paper "What Happens When Someone Acts?" Velleman raises a problem for a certain version of an entrenched view of action, namely, that it fails to capture the role that agents play in the production of their behavior, and he traces the source of this problem to its reductionistic aspect. In Section I I clarify and expand both on

⁸³ I use the more generic "happening" in place of the more common term "event." The latter most naturally connotes happenings that have already been completed, and that are reportable using perfective aspect (Thompson notes the word "event" is derived from the Latin verb *evenire*, where "*X eventum est*" literally means "X has happened" [2008, 163 n.17]). Davidson's examples of events also tend to have this feature (cf. 1967). But along with events we should also countenance happenings that are still under way in our ontology. Such happenings are ones which are reportable using the progressive aspect.

the nature of the worry and the sort of reductionism that is at issue. But Velleman also thinks that we both *can* and *should* respond to the problem he has raised within a reductionistic framework. In Section II I cast doubt on both of these claims. Velleman believes that he can solve his problem by utilizing the concept of “participation” from Harry Frankfurt’s work on free will, but it is a confusion to think that the sort of conceptual material that Frankfurt makes available has any bearing on the relevant issue. Once we are clear about this, we will see that the problem Velleman raises afflicts any version of reductionism about action, including his own. This should make us wonder how compelling the motivation for accepting a reductionistic account of intentional agency really is. Velleman claims that reductionism is motivated by a certain view of the natural world that should claim our allegiance, but in Section III I will claim that he is mistaken in this. Finally, in Section IV I return to the task of specifying what it is to be an intentional action, this time within the non-reductionistic framework that has been in the background throughout this dissertation.

5.1 WHERE’S THE AGENT?

Consider the following scenario: I am walking down the street wondering whether I will be ready to pitch at an upcoming baseball game. I therefore move my arm in order to get a large rock, and I throw it and shatter the window of an abandoned house. According to this story, I am apparently the cause of certain happenings. For it is a necessary truth that from propositions of the form “ x V_T ’s y ” we can infer that x causes y ’s V_I -ing, where “ x ” and “ y ” denote continuants, “ V ” stands for a verb with both transitive and intransitive forms, and “ V_T ” stands for the verb’s transitive form while “ V_I ” stands for its intransitive form. When I move my arm I

thus cause my arm's movement, and when I shatter the window I cause its shattering.⁸⁴ In both cases, as an agent I cause certain happenings to occur and my action is my *causing* them (a point that has been noted by G. H. von Wright [1963] and recently reiterated by Maria Alvarez and John Hyman [1998]).⁸⁵ The question that I will be examining is the following: is it possible to give a satisfactory analysis of what happens in cases of intentional action that makes no mention of the agent of the action as a cause of various happenings? Reductionistic accounts of intentional action provide an affirmative answer to this question.

Reductionism is often assumed in contemporary philosophy of action. In fact, a certain reductionistic theory of intentional action is so dominant that Velleman can refer to it simply as “the standard story”. He explains this story as follows:

[An agent's] desire for the end, and his belief in the action as a means, justify taking the action, and they jointly cause an intention to take it, which in turn causes the corresponding movements of the body. Provided that these causal processes take their normal course, the agent's movements consummate an action, and his motivating desire and belief constitute his reasons for acting. (1992b, 123)⁸⁶

⁸⁴ In her book *Actions* Hornsby attempts to move solely from conceptual truths of this form to the conclusion that when an agent moves her body, an event of her moving her body occurs which causes the event of her body's movement (1980, 13). This move is fallacious. The conceptual truth as applied to the case of bodily movement entails only that when an agent moves her body *she* causes her body to move. Hornsby's conclusion follows only if one assumes the additional principle, which Hornsby does nothing to justify, that from the fact that an agent causes a certain bodily movement one may infer that an event of an agent's moving her body caused that bodily movement.

⁸⁵ When the verb that describes my action is capable of taking both transitive and intransitive forms, the happening that I cause when I act is the happening denoted by the intransitive form of the verb. In these cases we have a purely syntactic criterion for identifying the caused event. However, this criterion is insufficient to identify the relevant events in all cases of action for the reason that not all descriptions of action involve verbs that exhibit the transitive/intransitive alternation. Berent Enç provides a more generic, non-syntactic criterion for identifying the relevant events, which he deems the *results* of action, according to which event R is the *result* of agent S's action A if and only if (1) “S did A” entails but is not entailed by “R occurred” and (2) “R occurred” carries all the information that is contained in “S did A” except for the fact that some action of S's was involved in the production of R (2003, 9).

⁸⁶ In the Introduction to *The Possibility of Practical Reason* Velleman also include intention as one of the causal antecedents to bodily movement in his characterization of the standard story. Nothing in what follows hinges on this discrepancy.

There are several points that should be noted. First, the standard story is supposed to be an account of *intentional* action. Second, it is an account of *what it is to be* such an action. It does not just state that all intentional actions are in fact bodily movements with a certain mental etiology, but that it is the nature of intentional action to be such a thing. Another way of putting this is that the standard story is an attempt to provide an answer to Wittgenstein's famous question, "[W]hat is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?" (*PI*, I §621).⁸⁷ If Wittgenstein's question is a good one, then we can understand, say, an intentional act of one's raising one's arm as being an event of one's arm's rising plus something else. There is reason to think that Wittgenstein himself did not think that this question was a good one, but the proponent of the standard story thinks that it is and cites causation by suitable psychological states and events in answering it.

Third, we should clarify what the subject of reduction is supposed to be, since as Hornsby has pointed out, the term "action" in ordinary language can refer to things of two different metaphysical categories (1980, 3-4). First, it can refer to something *that is done* by an agent, such as moving, kicking, or throwing. When used in this sense, an action is a universal that can be multiply instantiated by various agents and on different occasions. On the other hand, the term "action" can also refer to an agent's *doing* of something, and here it stands for a particular happening. When the standard story purports to offer a reductionistic account of action it is the second of these two senses of action that is relevant, for it seeks to define what it is to be a *doing* of an agent in terms of states and happenings which are not themselves the agent's doings but mere happenings. Gideon Yaffe provides a particularly clear statement of this point when he writes: "If we think of mere happenings—events that do not reveal the presence of any agency

⁸⁷ It is obvious from the context that Wittgenstein is here talking about intentional arm raisings.

whatsoever—as the basic events that there are, . . . then the problem of agency becomes a problem of reduction: How do we take a group of mere happenings and add them together to create an action?” (2000, 75; cf. Velleman 1992b, 130). Here it is clear that it is a particular happening, not a universal, that is the target of analysis.

Fourth, it is important to be clear about the sense in which the standard story is reductionistic. One thing that is often meant by calling an account reductionistic is that it seeks to explain the meaning of a certain concept that is seen as problematic in terms of the meanings of other concepts that are supposed to be less problematic. But this is not the sort of reductionism that is at issue here. Though Velleman disagrees with the standard story, he wants his own account to stay within its reductive aspirations, in the relevant sense. As we shall see, he thinks that a more adequate account of the nature of intentional action will have to have as the proximate cause of bodily movements a desire to act in accordance with reasons. Such an account would be a non-starter if what was at issue were a conceptual reduction of action, because the notion of *acting* would appear in the content of the desire in the *analysans*. The sort of reduction at issue is not semantic but ontological. The goal is not to provide a reductive analysis of the concept of intentional action, but rather to provide an account of intentional agency that does not invoke agents as causes of their movements.⁸⁸

Finally, reductionism is distinct from Davidson’s (1963) thesis that reasons explanation is a species of causal explanation. This is significant because Davidson is widely, though wrongly,

⁸⁸ Velleman explicitly states this aspiration when he writes that the reductionistic project shared by himself and proponents of the standard story aims “to show how the causal role assigned to the agent by common sense reduces to, or supervenes on, causal relations among events and states of affairs” (1992b, 130).

credited as the leading contemporary defender of Velleman's standard story.⁸⁹ Confusion persists on this point in part because some theorists refer to reductionistic accounts of intentional action as "causal theories of agency" (e.g., Yaffe 2000, 121; Enç 2003, ch. 3). But the causal theory that Davidson defends is one about the *explanation* of intentional action and not its ontology. He argues that rationalizing explanations—the sort of explanation that makes sense of action by displaying it as aimed at an end its agent sees in a favorable light—is a species of causal explanation, on the ground that it is the best way to understand the "because" that we invoke when we say of a person who had or has multiple reasons for doing some action that she did it or is doing it because of a certain subset of those reasons (1963, 11). What Davidson here asserts is causally explained by certain psychological states is a person's doing such-and-such, and one can agree that he is right about this without thinking it is possible to give a reductive definition of what it is for a happening to be a doing of such-and-such in terms of the causation of bodily movements by those psychological states. Reductionism is independent of Davidson's argument and can draw no support from it.

Now that I have clarified the content of the standard story, I want to turn to an important objection that several philosophers, including Velleman, have recently posed for it. Velleman's preferred way of making this objection is to say that the standard story fails to cast the agent in her proper role, since on it she is merely an "arena" for certain psychological and physical happenings (1992b, 123). However, I think that another one of his formulations is more telling, namely, that it fails to include an agent at all. The two ways of stating the objection are closely related. The worry is that insofar as the standard story conceives of action as the result of the causal interaction of states or happenings that occur within the agent, it fails to provide the agent

⁸⁹ For example, M. Smith 2004, 165. Velleman himself is careful to avoid taking a stand on the issue of Davidson exegesis.

with any role in the production of her behavior other than being the subject or location of the states and happenings in question. In the words of Thomas Nagel, whom Velleman cites approvingly in this context, “it is not clear that [in the standard story] there is anything for the agent to contribute. . . —anything that he could contribute as source, rather than merely as scene of the outcome” (1986, 113-4). Similarly, John Bishop writes, “Even if it is my own mental states that cause me to behave, surely such behavior must fall short of a genuine action of mine in which *I* contribute to determining what happens” (1989, 44).

We do not think that we as agents sit passively by while our desires combine with our beliefs to produce intentions and actions; rather we think that we reflect on what we take to be our various reasons for acting, choose which shall express our will, and carry it out in action, and the standard story seems to leave these phenomena out of its account. But this way of putting the objection naturally leads to Velleman’s other, non-preferred way of stating it. For what it is to be an agent precisely *consists in* being the author or “source” of one’s behavior, rather than merely its “scene”. Therefore, what is referred to by proponents of the standard story as “the agent” could only be something that we had hoped to be able to reveal as an agent but is in fact something this line of critique shows we are not entitled to so reveal.

The upshot is that the way the standard story attempts to carry out a reduction of actions in terms of materials that do not include an agent as a cause is unsuccessful. This is clear from Velleman’s summary of his objection. He writes, “[On the standard story] reasons cause an intention, and an intention causes bodily movements, but nobody—that is, no person—*does* anything” (1992b, 123). But saying that nobody does anything on the standard story is equivalent to saying that it is not a story about *action* at all.

5.2 VELLEMAN AND FRANKFURT

It appears that Velleman's objection to the standard story should be applicable to reductionistic models of intentional action in general. The problem as he presents it is that it is difficult to see how any network of causally related mere states and happenings could add up to an agent's doing something, and this worry seems to apply regardless of what states are in question. Interestingly, though, Velleman does not think that his objection is quite as devastating as his initial way of putting it might lead one to believe. Rather, he thinks that the standard story is all right for certain "defective" forms of agency, e.g., the actions of non-human animals, halfhearted actions, akratic actions, and actions which are explained by appeal to psychoanalysis, and that it is problematic only insofar as it cannot account for actions that are not defective in any of these ways, a class he calls human actions *par excellence* (ibid., 124).

Velleman characterizes the problem as one of "participation". He takes it that saying that a person is the author of her action is equivalent to saying that she participates in her action, and he also takes it that actions in which agents participate are human actions *par excellence*. This in turn leads him to ask what accounts for the difference between defective action and human action *par excellence*. His answer is that in the latter kind of case there is an extra psychological state involved in the production of bodily movement. This state "plays the functional role of the agent" (ibid., 137). When this state is among the causes of a bodily movement that movement counts as a case of human action *par excellence*, and when it is absent the standard story applies and the resulting movement is a case of defective agency. What could this state be? Velleman begins with the plausible thought that the mark of practical deliberation, which is thought characteristic of rational agents as such, is that it is thought aimed at acting in accordance with

reasons. From this he infers that the additional state that is necessary for action *par excellence* is the presence of a desire to act in accordance with reasons:

We say that the agent turns his thoughts to the various motives that give him reason to act; but in fact, the agent's thoughts are turned in this direction by the desire to act in accordance with reasons. We say that the agent calculates the relative strengths of the reasons before him; but in fact, these calculations are driven by his desire to act in accordance with reasons. We say that the agent throws his weight behind the motives that provide the strongest reasons; but what is thrown behind those motives, in fact, is the additional motivating force of the desire to act in accordance with reasons. (ibid., 141)

By having a functional role that is realized by the desire to act for reasons, the agent can literally be said to get in on the production of her behavior in human actions *par excellence*.

Velleman therefore thinks that he can account for the sense in which an agent can be said to participate in her deeds that meet the ideal of human action *par excellence*. Yet his account raises a serious question about the status of actions that fall short of this ideal. Recall that he thinks the standard story is fine for describing what happens when someone acts “halfheartedly, unwittingly, or in some equally defective way” (ibid., 124). This is a bizarre consequence, given his criticism that the standard story makes a human being into an arena of causally related states and happenings. For we do not think of the movements of a halfhearted or akratic person as being the external result of psychological happenings without her being involved (Hornsby 2004, 182-3). While some such actions may not reach Velleman's ideal, they are nevertheless still *actions*, and agents who choose a less-than-ideal action as the result of their practical deliberations do not, solely on that count, become less than agents.

Nor do I think that it would help to weaken the content of the desire that is supposed to play the functional role of the agent, making it a desire to act for some reason or other. This is due to the fact that it seems possible to act intentionally without acting for any reason at all. I

can intentionally draw a doodle while I talk on the phone but my drawing does not serve any further purpose. This does not mean that such actions are completely unrelated to the capacity to act for reasons, since, as Anscombe points out, even when I do something intentionally but for no reason, the question “Why?” in the special sense in which it serves as a request for a reason for action is an appropriate one in a way in which it is not in a case of non-intentional actions (1963, §17). But such actions both cannot be caused by a desire to act for reasons and yet are still genuine actions.⁹⁰

Perhaps a reductionist could modify Velleman’s theory in light of the present line of objection. What leads Velleman to such an undesirable consequence is the fact that he thinks that the special desire to act for reasons must not only be causally efficacious in producing a bit of behavior in order for that bit of behavior to count as an action but must actually have sufficient force to determine the course of behavior. A more sophisticated version of his theory could avoid the objection by stipulating that the desire to act for reasons need only be exerting causal influence in one’s psychological economy in order for the movement produced by that economy to count as one’s intentional action. When the desire succeeds in determining the movement, that movement counts as a human action *par excellence*, and when it is overpowered by other desires, the resulting movement is a less than ideal action.

⁹⁰ Is it not the case that even these actions must be motivated, and thus caused by a desire to do them, and would not this desire itself be a reason for these actions—so that there is actually no such thing as actions done for no reason? I will not dispute the premise that an act of A-ing must be caused by a desire to A, but the conclusion required to save Velleman’s theory does not follow. This is because the phrase “reason for action” is ambiguous, and can denote either what has been called a “motivating reason” or a “normative reason” (M. Smith, 1994, 95-6). A desire is a reason for action in the first sense, which means that it plays a role in an explanation of why the action was performed. A reason for action in the second sense, on the other hand, is a fact in light of which the action would be a sensible or desirable one. For an action to be done “for a reason” in this sense would be for it to be done in light of such a fact. Furthermore, it is also the sense of “reason for action” in play in the content of the alleged desire to act for reasons, since that desire would be a desire to act in the light of rational considerations. Although it is true that intentional actions that are performed “for no particular reason” are caused by a desire to perform them, this does not show that they are after all done in the light of some particular rational consideration.

I do not think that such a revised theory can rescue the reductionistic program. While it avoids the problem for Velleman's original theory it falls into another one that theory was designed to avoid. Recall that the difficulty for the standard story was that it appeared to leave no room for the agent in the production of her movements. One could object here that the standard story does not have this consequence if we understand it as saying that for an agent's desires⁹¹ (and on one version, intentions) to determine her movements is what it is for an agent to produce those movements.⁹² But Velleman replies that this is inadequate, for we do not think that an agent causes her body to move in just any way, but that her role comes in at particular points in the causal chain, and the standard story still omits this in its account. The world comes to conform with reasons through one's activity. It does so through one's desires, but these desires influence the course of events only indirectly, by influencing *the agent* to produce various movements and states of affairs (1992b, 125). Velleman's theory attempts to accommodate this fact by having it that the desire to act for reasons must actually determine one's behavior in order to play the functional role of the agent. If we drop this feature, as the more sophisticated version we are now considering does, then we also forfeit our ability to explain the essential role of the agent as mediator between desire and movement.

I therefore think that Velleman is stuck with the unappealing consequence that I earlier argued was entailed by his view. Yet perhaps this result is not as bizarre as it initially sounds? Apparently Velleman does not think it implausible that human beings who act in some less than ideal way should be seen as mere arenas for psychological and physical happenings, without

⁹¹ I mean the only sort of desires that the standard story countenances.

⁹² This is the line taken by Goldman (1970, 80).

participating in a way that would render them agents. I will next explain the likely reason for this thought and I will argue that it rests on confusion.

At one point Velleman claims that his project is simply a modification of Frankfurt's (ibid., 133). Frankfurt is also concerned with situations in which we find it natural to say that an agent doesn't "participate" in her action. In a number of essays he has explored how this metaphor is systematically connected with several others. On Frankfurt's account persons are beings whose will exhibits a reflexive structure by which they are capable of stepping back and evaluating their own motivational attitudes (1971, 12). The capacity for reflection gives persons a capacity lacked by creatures not endowed with it, namely, the capacity to reflectively *endorse* some of their desires (1971, 12; 1987, 163).⁹³ Frankfurt has used this notion to explicate a family of interrelated metaphors for describing the psychology of a person when she endorses or fails to endorse a certain desire. Perhaps most central in his work is the idea that when a person endorses one of her desires she thereby becomes "identified" with it. In endorsing a desire, an agent makes it, in some sense, a part of herself. Of course, in another sense all of a person's desires belong to her, but in endorsing the desire the person authorizes it to speak for what she judges upon reflection to be worth desiring or caring about. It is because of this that a person can be said to participate in actions that are motivated by endorsed desires. Frankfurt calls actions motivated by these desires *wholehearted* ones (1987, 165).

Yet it is also possible for a person to act on a desire that she does not endorse, either because she is not sure whether to endorse it or because from the reflective standpoint she positively rejects it. The possibility of acting on such motives is what underwrites such familiar

⁹³ Frankfurt initially analyzed this capacity as the capacity to form higher-order desires, and this analysis has been much criticized (e.g., Watson 1975). Frankfurt has accepted some of these criticisms and subsequently sought to modify his approach in light of them. For our present purposes we can abstract from any particular analysis of this capacity.

phenomena as ambivalence, halfheartedness, and weakness of will. In such situations the agent does not identify with her motives for action. They do not carry the authority of her endorsement and because of this they are in a sense “external” to her. Frankfurt vividly illustrates this sense of alienation with the example of a narcotics addict who endorses quitting but who inevitably succumbs to his desire to have another hit. He claims that in this situation, “the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force other than his own, and that it is not of his own free will but rather against his will that this force moves him to take it” (1971, 18).

Velleman claims that he is engaged in the same project as Frankfurt, the difference between them being that whereas Frankfurt attempts to capture notions such as identification and participation in terms of the ability to form higher-order desires, Velleman thinks that this end is better served by a desire with a particular content. I think that this explains why Velleman does not think it is objectionable if his theory represents human beings who fall short of engaging in action *par excellence* as being passive arenas for certain happenings. For Velleman seems to want his idea of human action *par excellence* to function as an analogue to Frankfurt’s notion of wholehearted action, and Frankfurt does a lot to make plausible a sense in which less than wholehearted actions are actions which lack a certain kind of participation by the agent.

The question is whether the sense of non-participation or passivity that Frankfurt picks out is the same as the one that figures in the standard story as Velleman criticizes it. I think that upon a moment’s reflection we can see that it is not. Although the addict may think of his desire for drugs as an “external force” operating on his will, it would be a gross distortion to think that in such a case the desire literally pushes his limbs around in the way that, e.g., the impact of a heavy object would. As Gary Watson explains, when one is conquered by a desire one’s will is

not altogether bypassed. It is more appropriate in these cases to say that the person overcome by desire is more like a collaborationist than an unsuccessful freedom fighter: such a desire works by diverting one's attention and making one uncomfortable until one acts upon it (1999, 65). This brings out a sense in which the behavior of a drug-seeking addict is still attributable to *her*. Preserving this fact is important for our understanding of the moral dimensions of addiction: we feel sympathetic because we recognize the obstacles to autonomy that addiction creates and recognize that addiction can diminish or eliminate one's responsibility for one's actions in some situations, but we also hold that the addict is at least sometimes accountable for her behavior (Wallace 1999, 188). These points hold *a fortiori* for the less extreme cases of halfheartedness and weakness of will.

The upshot is that we should distinguish between at least two different versions of "participation" and related notions like "identification", "activity", and "internality". These metaphors have to do with the sense in which states and happenings in one's mental life can be said to be related to oneself. But there are different ways of conceptualizing what is internal or external to oneself. In the age-old picture of the battle of reason and the passions it is oneself *qua* rational agent that must be constantly on guard against besiegement by the forces of the lower parts of the soul. But when we turn to debates where the concept of agency is central the agent participates in all of her actions whether or not they are motivated by desires she can reflectively endorse, simply because they are hers. The contrast here is with outside forces that would render the agent's action involuntary.⁹⁴

Indeed, Frankfurt himself recognizes such a distinction insofar as he is careful to distinguish "free action" from what he calls "free will" (1971, 20). In Frankfurt's special use of

⁹⁴ R.J. Wallace makes a similar point when he says that we should be careful to distinguish a sense of identification that goes with responsibility from a sense that goes with autonomy (2000, 191).

the terms, an agent's action is free when she does what she wants, whereas her will is free when she endorses the desire she acts on. This means that even the acts of unwilling addicts are genuinely free actions on Frankfurt's account, and so there is a sense in which they are agents, even if they do not participate in their actions in the further sense required for them to enjoy free will.

Frankfurt does not have much to say about freedom of action, other than noting that it is subject to its own unclarity. This is perfectly justified given that his focus is on what he refers to as freedom of the will, but it is important to stress that he *takes it for granted* that some story about free agency can be told. His goal is not the more general one that Velleman is engaged in as "a fundamental problem in the philosophy of action—namely, that of finding a place for agents in the explanatory order of the world" (1992b, 128), or for that matter, even of finding a place for free agents. Rather, it is the different one of explaining another, further way a person can be said to be passive with respect to her actions assuming that we can already describe them as actions and, and more determinately, as free actions.

We are now in a position to see the source of the difficulties that I brought out for Velleman's view. Velleman misunderstands the nature of Frankfurt's project. He therefore runs together the sense in which an agent participates in any of her actions and the more specialized sense in which agents may be said to participate in actions that are free and meet a further condition of excellence. He begins his article by agreeing with Nagel and Bishop, who argue that a picture of action like the one told by the standard story cannot make sense of how agents participate in their actions in the first and most basic sense, the sense that must be intelligible if the very idea of agency is to have application. But the project he goes on to pursue is an attempt to modify Frankfurt's theory as to what accounts for an agent's participation in the second sense.

A consequence of this confusion is that on Velleman's theory the conditions of agency in general become coextensive with the conditions of agency *par excellence*; hence the strange result that human beings who do not meet this ideal can be nothing more than arenas in which psychological and physical happenings play out.⁹⁵

5.3 REDUCTIONISM AND NATURALISM

I have now diagnosed what I take to be a central confusion in Velleman's handling of the problem he raises for the standard story. Nor do I think that it should be surprising that his strategy does not work. The reason is that he treats his problem as a local one insofar as he assumes that it arises only in the context of rational human agency. He therefore feels free to draw on the characteristic activity of human beings, that they typically act in accordance with their understanding of reasons, in his attempt to solve it. But in fact the problem with reductionism extends to forms of agency beyond this scope. Snakes slither, sunflowers turn toward the light, and acids dissolve metal, and while not all of these are instances of *intentional* agency, they are still actions in the most general sense of the term. Thus, it is necessary to see if Velleman's problem occurs in some of these instances of agency as well. Perhaps one could plausibly think that no such difficulty arises in the third of these cases. I do not mean that one should not be a non-reductionist about the causal role of acids, but if a philosopher were to give a

⁹⁵ I think that it is this confusion that lies behind a curious footnote (n. 5) in Velleman's paper. Immediately after stating his objection to the standard story Velleman confidently states "Of course, every action must be someone's doing and must therefore be such that an agent participates in it, in the sense that he does it. But this conception of agential participation doesn't require anything that is obviously missing from the standard story." Here Velleman fails to notice that it is precisely the point of philosophers like Nagel and Bishop, whom he sees himself as following, that on a picture like the standard story there is *no* sense in which an agent can be said to do anything (cf. Hornsby 2004, 181). What accounts for Velleman's failure here is his confusion of the line of thought stemming from Nagel and Bishop with the one stemming from Frankfurt.

reductive analysis of their role in terms of the causal interactions of various states and happenings that occur during chemical reactions, his account would not be obviously objectionable on the ground that it somehow left the acids out. I do want to suggest, though, that this objection could be made in the case of living beings. It is surely just as plausible to think that other living things are more than arenas for the occurrence of mere happenings as it is to think that human beings are. But if the difficulty can be generalized to domains of agency beyond the human, then Velleman's attempt to resolve it with materials available only in the specifically human case is too parochial to succeed.⁹⁶

This raises the question of why we should accept reductionism at all. In this connection, Velleman writes:

To insist that the story mention only the agent himself as the object of rational influence, or as the author and executor of intentions, would be to assume *a priori* that there is no psychological reduction of what happens in rational action. One is surely entitled to hypothesize, on the contrary, that there are mental states and events within an agent whose causal interactions constitute his being influenced by a reason, or his forming and conforming to an intention. (1992b, 124)

Velleman here claims that to accept that an agent plays some role in the production of intentional behavior would be to rule out reductionism *a priori*. But I have not argued that we should rule it out *a priori*. Rather, I think that once we get clear on what the real issue is reductionism should seem *prima facie* implausible, and this is sufficient to undermine Velleman's confidence that he is surely entitled to assume it in a philosophical account of human agency. After all, once we agree that there is something problematic with conceptualizing an

⁹⁶ Yaffe also misses this point in his mostly very perceptive comments on Locke's theory of agency. Like Velleman, Yaffe thinks the "Where's the Agent?" problem arises only when we turn to providing an account of free human agency. But he overlooks the fact that the same problem would arise for Locke's theory of active powers in a more extensive range of cases of agency, since according to Yaffe Locke defines action in general reductively, as the modification of a substance that is caused at least in part by another one of its modifications (2000, 81).

action in terms of causation by mere happenings and states, it is hard to see how merely adding more of the same *kind* of things could help, and I have argued that Velleman's treatment does nothing to make this problem seem any less intractable.

There is, however, a deeper source of Velleman's opposition to non-reductionism. For he claims it is opposed to a view that he believes naturalistically-oriented philosophers are obligated to accept, namely, a "naturalistic conception of explanation" (ibid., 130), or alternatively, "the scientific view of the world". According to this view, "all events and states of affairs are caused, and hence explained, by [and only by] other states or events, or by nothing at all" (ibid., 129).⁹⁷ But I think we need to cast a critical eye on the claim that anyone who wants to accord due respect to the natural sciences is really committed to adopting this sort of ontology. Indeed, it is Velleman's own reductionistic "scientific view of the world" that appears to be at odds with our actual scientific practices, since it has trouble accommodating not only the existence of human agents but other living ones as well, while it is the task of many of the natural sciences to study the causal powers of such beings.

Velleman simply dismisses non-reductionism as being incompatible with naturalism without argument, and I think that at least part of the reason he feels justified in doing so is because he equates it with Roderick Chisholm's conception of agent-causation. Like Velleman, Chisholm is concerned specifically with the sophisticated form of action engaged in by human

⁹⁷ The addition in brackets is mine, but makes for a more accurate statement of what Velleman must have in mind. For it is clear that he thinks that the "scientific view of the world" is at odds with non-reductionism, and his statement needs the addition in order to be the statement of something incompatible with it. Without it his formulation does not preclude positing agents as non-reducible elements in the causal order so long as the effects that they produce are simultaneously caused by certain happenings or states. Velleman is not the only philosopher who supposes that reductionism is entailed by naturalism. Bratman (2001, 92), Enç (2003, 2), Yaffe (2000, 121), and Bishop (1989, 43) also make a similar assumption in their writings on agency.

beings.⁹⁸ Indeed, in “Human Freedom and the Self” he endorses a reductionistic account of the sort of agency had by non-human life-forms and inanimate objects, holding that everything that happens in these domains can be accounted for by appealing to causal relations that obtain among happenings or states of affairs (28). It is only when we come to acts of human beings that Chisholm thinks we need to introduce the notion of agent-causation, which is a sort of causal relation that holds between substances and happenings they bring about, and he argues that we must introduce it at this point in order to allow for the possibility of human freedom. Furthermore, Chisholm claims that when they act intentionally the events that human agents cause are neural ones, and that these agent-caused events in turn cause further events in the nervous system until a bodily event takes place, the occurrence of which is appropriately related to the content of the agent’s intention (*ibid.*, 32).

I fully agree that Chisholm’s theory should seem suspicious, insofar as he conceives of human agents as somehow interfering with processes that occur in their brains in a way that lower animals do not. It flies in the face of the canons of comparative physiology to posit such radical kinds of discontinuity between the biological workings of humans and other species of animal. But I think that this is a problem for Chisholm’s particular version of non-reductionism, not non-reductionism *per se*. It should already be clear that insofar as he introduces the notion of agent-causation to account for human freedom, Chisholm’s theory will be inadequate for getting around the problem for reductionism, as that was supposed to apply to a much wider class of agents than ones whose actions are characteristically free actions. In fact, Chisholm appears to conflate these issues in the same way that Velleman does. For there are at least two different

⁹⁸ I base these comments on Chisholm 1964. This is probably the most well-known version of Chisholm’s account, though he made many modifications to his view over the years. For our purposes it will not be necessary to consider the details of this development.

contexts in which the idea of agent-causation might figure, and in each of these contexts it denotes a different relation.

The first is the issue we have been investigating, whether an agent's being the cause of a certain happening can be cashed out in terms that mention only happenings and states as causes. This is the problem that Chisholm appears to be concerned with under the heading of agent-causation in the passage which Velleman cites as a statement of non-reductionism:

[T]he issues about "agent-causation"...have been misplaced. The philosophical question is not—or at least shouldn't be—the question whether or not there is "agent-causation". Thus, for example, if we have good reason for believing that Jones...kill[ed] his uncle, the philosophical question about Jones as cause would be: Can we express the statement "Jones killed his uncle" without loss of meaning into a set of statements in which only events are said to be causes and in which Jones himself is not said to be the source of any activity. (1978, 622-3; qtd. in Velleman 1992b, 130)

Notice that while Chisholm's example involves a human person doing something intentionally, the issue is really a more general one. That is, the question is whether we can understand propositions of the form "S did (or is/was doing) A" without invoking whatever takes the place of "S" as a cause, where what can take the place of "S" are a wide variety of continuants, both human and non-human. Notice too that for all Chisholm says here it need not be the case that what is caused by S is the doing of A. As we have seen the grammar of propositions about action supports the view that when an agent does A it does not cause its doing A but rather other happenings, the occurrence of which would make it true to say it did A, of which its action is the *causing*.⁹⁹

But there is a second debate in which a very different concept of agent-causation has figured, namely, the debate over what conditions must obtain in order for an action to be free.

⁹⁹ See Section I.

Traditionally compatibilists have argued that S's doing A is free just in case it is caused by suitable prior psychological states, while incompatibilists have argued that it is also a necessary condition for S's doing A to be free that it is not necessitated by these psychological states.¹⁰⁰ Proponents of the theory of agent-causation in this debate take a different line. They argue that for S's doing A to be free either it must also be partially caused by the agent herself, or the intention to do it must be so caused. There are three ways in which this debate is distinct from the one over reductionism. First, its subject matter is more restricted, since here only subjects capable of having psychological states can take the place of S. Second, according to the agent-causal conception of freedom the causal relation is supposed to hold between an agent and her action or an agent and her intention, not an agent and the happenings of which her action is the causing. Finally, in this context propositions of the form "S is doing/was doing/did A" are assumed to be understood and whether or how they can be further analyzed is not part of the dispute.

The announced topic of "Human Freedom and the Self" is this latter debate on freedom. However, Chisholm appears to think that the fact that only humans are free means they are the only things in nature that are agent-causes in the sense of being the only sort of substances that stand in causal relations with happenings or states of affairs (1964, 20). But of course this conclusion does not follow from any considerations about freedom, even that it consists in agent-causation in the relevant sense, for that would just be to confuse the different things that can be meant by that term. Furthermore, all the objectionable aspects of Chisholm's position derive from his conception of what human freedom requires. For he thinks first, that for a person's

¹⁰⁰ Note that incompatibilists need only deny that a free action can be necessitated by an agent's prior psychological states; they can agree that they can causally influence the action. That the very idea of causation does not involve necessitation is forcefully argued by Anscombe (1971).

doing A to be free she must be the cause of her doing A, and second, that in order to be the cause of her doing A a person must be the cause of a neural event. Both of these claims are dubious, but the present point is that they are simply irrelevant to the issue of this chapter, which is non-reductionism and not freedom.¹⁰¹

Non-reductionists therefore need not commit themselves to the excesses of Chisholmian metaphysics. I think this gives us good grounds for coming up with a different way of understanding non-reductionism, one that has a wider scope and does not suffer from the defects of Chisholm's theory. Fortunately there is a non-reductionist account of action that both meets these criteria and is immensely plausible. The core idea, which has been given its fullest and most enduring exposition by Aristotle (cf. *Physics* II-III, *Metaphysics Z-H*), is that the category of substance is a basic element of our conceptual scheme. Substances are things with various sorts of powers or capacities. Among these are active powers, or powers to bring about or cause certain effects, and passive capacities, which are capacities to be affected in certain ways. For instance, spiders have the power to spin webs, and when a spider exercises this power, it causes a web to be spun. They also have a passive capacity to see the prey that they capture on these webs, and this capacity is actualized when their eyes are affected by light in appropriate circumstances. This provides an obvious answer to Velleman's provocative question, "What else could an agent be?" if it is not a state with a certain functional role (1992b, 143). A substance is an agent when it exercises one of its active powers and it is a patient when it undergoes a change and thus actualizes one of its passive capacities. It can therefore be at some points an agent, at others a patient, and sometimes be both the agent and patient of a single act, as when a cat grooms itself.

¹⁰¹ Similarly, Davidson's dilemma for agent-causation obtains only for those who hold we can illuminate our notion of agency by arguing that agents cause their actions (1971, 52).

This account is non-reductive since it takes seriously the idea that substances cause effects in the world and it makes no attempt to cash this idea out in terms of causal relations among happenings that do not themselves imply the presence of agency. However, this does not mean there is nothing more that can be said by way of explaining how it is that certain kinds of substances are able to bring about certain sorts of characteristic effects. On the contrary, a good description of the task of many of the natural sciences is that these sciences investigate the causal powers of various sorts of things, and then seek to explain how it is that they possess them by appealing in part to the powers of the things that compose them.¹⁰² Non-reductionism is not incompatible with our actual scientific practices but only with a metaphysical interpretation that has been foisted upon them by philosophers.

5.4 A NON-REDUCTIONISTIC ACCOUNT OF INTENTIONAL ACTION

The reason I gave for examining reductionism was that the Aristotelian account of intentional bodily acting as a psychophysical phenomenon is bound to seem puzzling to those who hold reductionistic models of action. We have now seen that reductionism is both problematic and insufficiently motivated. This provides warrant for taking non-reductionism seriously. Therefore, in this final section I want to return to the question of what it is to be an intentional action, but this time from the non-reductionistic viewpoint that has been implicit throughout this dissertation. I will argue that intentional action never consists in the bodily movements that

¹⁰² “In part” because a substance’s environment and history can also sometimes be relevant to explaining how it possesses certain powers and capacities. For an idea of how a non-reductionistic framework might be applied to the natural sciences, see Harré and Madden 1975, which is full of examples taken from the history of physics, chemistry, and biology.

occur when human agents move their bodies (*pace* reductionism), and that it only sometimes consists in agents moving their bodies.

The basic non-reductionistic idea is that action is to be understood as an exertion of a causal power of a substance, where substances are taken as ontologically basic. A natural thought is thus that we can understand what makes an act intentional by specifying the nature of the causal power that is exerted when one acts intentionally. According to the account in Chapter 3, an act is intentional when it is the exercise of an agent's will, or power of practical reason. A rational agent exercises her will when she answers the practical question what to do by reasoning from an end she takes to be desirable to something she immediately knows how to do. And when the time for acting is here and now her action constitutes her answer to this practical question. Intentional action is therefore susceptible to being explained by showing how what one is doing, under a certain description, is a phase of a more encompassing action, or a step in a series of actions, that is sufficient or at least likely to accomplish a certain end. Furthermore, since the will is a self-conscious power, the agent herself is normally able to provide the explanation of what she is doing.

It is sometimes the case that an intentional action will be or consist in an agent's moving his body a certain way, e.g., if one raises one's arm intentionally. Note, though, that even here the agent's action is not his arm's rising but his *causing* his arm to rise. I argued in Chapter 4 that in a situation such as this one, one's arm's rising is not identical with the process of one's arm raising, but is the final phase of it. Or consider Anscombe's case of the assassin at the pump. In this scenario a certain man is trying to bring about a revolution. He is doing so by poisoning the water supply of the residents of a certain house, and he is doing this by pumping toxins into their water supply from a nearby pump. While the man may truly be said to fall

under the descriptions, “poisoning the water supply”, “operating the pump” and “moving his arm up and down”, these descriptions all describe one and the same happening (1963, §26). In these circumstances, the man’s moving his arm up and down on the pump *is* his operating it, and his operating it *is* his poisoning the water supply of the house.

It does not follow, however, that all of our intentional actions are identical to our moving our bodies, in the way that the pumper’s poisoning the well is his moving his arm up and down on the pump. It is possible for S to do A intentionally without there being any happening that is his doing it.¹⁰³ For example, if S agreed to bring the food and beverages to a party, then S may intentionally ruin the party simply by deciding not to show up. Of course, S might do plenty of other things in the meantime, such as watch television or go to bed. It is not plausible, however, that S’s act of ruining the party is his engaging in these activities. The direct object of our criticism would not be that S engaged in these other activities but rather that he failed to come to the party after agreeing to provide the food and beverages. What this example shows is that there can be cases of omission where what one does intentionally (e.g., ruin a party) consists in one’s intentionally *refraining* from doing anything about a certain manner. And refraining as such does not consist in moving the body.

Davidson thought that if we set cases of omission to the side, along with mental acts like calculating a sum in one’s head, then all the rest of our intentional actions consist in our moving our bodies (1971, 49). He claimed that as our acts of moving our bodies go on to have various effects in the world they also come to be describable by reference to their effects (*ibid.*, 59). In doing so he sought to generalize Anscombe’s example of the man at the pump to all cases of

¹⁰³ Here it is crucial to keep in mind the point that the term “action” can refer either to something that an agent does (a universal) or to an agent’s doing (a particular happening) (see Section I).

bodily intentional action. For in that case it is plausible to say that, e.g., it is because the man's operation of the pump has the effect of pushing toxins into the water supply that he is also truly describable as causing the water supply to become poisoned. But even within Davidson's limited domain, what is true of Anscombe's example cannot be generalized to all cases of intentional action.¹⁰⁴

To see that this is so, consider the pedestrian activity of making a cake. In order to make a cake one must perform a number of other actions in a certain temporal sequence. For example, one must crack a number of eggs, measure out some oil and water, pour these into a bowl along with cake mix, stir it all together, grease a pan, pour in the mix, put the pan in the oven, and wait for an hour. At any time one is engaged in one of these other less-encompassing activities, one is also engaging in the more overarching activity of baking a cake. But it is not the case that at any time one is engaged in one of the lesser activities one is moving one's body. There are two ways in which this might be so, which correspond to the two sorts of cases Davidson believed could be set aside. First, any activity that takes a significant interval of time to accomplish will tend to involve pauses in bodily movement. So after I put the cake mix in the oven, I will have to wait an hour for it to bake. During this phase of the process of cake-baking I need to do nothing, save being ready to take it out of the oven. And while I may do other things during this period of time, it is not the case that my cake-making consists in my engaging in these activities. Rather, during that time it consists in *waiting and standing ready* to take it out of the oven at the appropriate time. Second, complex bodily activity may also require that one perform certain mental acts, as when I calculate to myself how much oil and water I will need for a cake of such-

¹⁰⁴ In Chapter 4.2 I point out another sort of counterexample to Davidson's thesis.

and-such a size. So the shape my cake-baking might take at a certain time would therefore be my performance of an act of mental calculation.

The lesson to be drawn from these reflections is that it is a mistake to think that we can always neatly divide up actions into cases of refraining and non-refraining, or cases of mental and bodily action, in the way that Davidson imagined. While this might work in some simpler cases, more complex bodily activities will often have phases in which doing them will not involve making any bodily movements. What we can say is that, for creatures like us, act descriptions like “making a cake” can be classified as descriptions of bodily activities insofar as they require *some* bodily activity to be instantiated. The reason for the qualification “for creatures like us” is that it is at least possible to imagine beings that do not need to make bodily movements in order to effect changes in the spatiotemporal world. While it is an empirical fact that human beings do not have powers to cause such changes through thought alone, the very idea of such a power does not appear to be an incoherent one. It is in any case incorrect to define intentional action in terms of bodily movement. At the most abstract level, what it is to be an intentional action is to be an instance of the central act of practical reason.¹⁰⁵ We get a conceptual grip on this power by understanding the nature of practical reasoning. And insofar as the will is a rational capacity, it is an essentially self-conscious one.

¹⁰⁵ It is the central act because other acts of the will, such as the formation of intentions for the future, are to be understood by reference to it (see Chapter 3.3).

APPENDIX A

BODILY AWARENESS AS A FORM OF SELF-CONSCIOUSNESS

In Chapter 2 I argued that if one accepts Descartes' thesis that when one is self-consciously aware of oneself one is not aware of oneself as having any physical properties, one will be barred from giving a satisfactory answer to the question "What am I?" I identified self-conscious epistemic powers with powers for self-knowledge, or powers to acquire knowledge of oneself that is immediate and non-observational. And I concluded that in order to have an adequate conception of our own nature we must show that we have self-knowledge of some of our bodily states or activities, described as such. The main thread of my dissertation has been devoted to developing in detail one solution to this problem of bodily self-knowledge, which focuses on knowledge in intention.

The account of the epistemology of action I defend is sufficient to solve the problem of bodily self-knowledge, but I do not claim that it is the only way to do so. Besides having the power to know what we are doing intentionally, we also have the power to know the relative positions and motions of the parts of our bodies (proprioception) and powers to know certain somatic states such as temperature, hunger, fatigue, and pain that do not rely on the five traditional senses (cf. Bermúdez 1998, 132-3). Following recent usage, I will group this family

of epistemic powers under the heading “bodily awareness”. The question whether bodily awareness constitutes a genuine form of self-consciousness has recently become the topic of philosophical debate. Proponents of an affirmative answer include Gareth Evans (1982, ch. 7), Quassim Cassam (1997), and José Luis Bermúdez (1998), while M.G.F. Martin (1995) and Lucy O’Brien (2007) are among those who answer it negatively. Besides being interesting in its own right, this debate is obviously relevant to my overall project. Since it is also somewhat of a tangent from my main line of argument, however, I will only be able to explore it briefly in this appendix. My remarks will necessarily be limited; a comprehensive treatment would require a dissertation in its own right.

My central goal will be to connect the question whether bodily awareness is a form of self-consciousness with a different one, namely, whether such awareness should be classified as perceptual in nature. According to one popular approach to the epistemology of bodily awareness, these powers make up a group of “internal” senses to be placed alongside the five traditional “exteroceptive” ones. Understanding whether bodily awareness is perceptual is important for answering the question whether the modes of bodily awareness are modes of self-consciousness. We saw in Chapter 2 that observational knowledge of oneself acquired on the basis of the “exteroceptive” senses does not constitute self-conscious self-knowledge. I will argue that this is not an accidental feature of these senses; rather, it follows from the very idea of perceptual knowledge that knowledge of oneself had on this basis is not self-conscious self-knowledge.

In Section I I argue that insofar as an epistemic power is perceptual, it cannot be self-conscious. The reason is that the mark of self-conscious knowledge of oneself is that it is immune to error through misidentification, while it is incompatible with the nature of sense

perception to provide knowledge that is immune to such error. In Section II, I defend Evans' thesis that the knowledge of oneself provided by bodily awareness is immune to error through misidentification. Finally, in Section III, I argue that even though proponents of the perceptual model of bodily awareness are right to point out that perception and bodily awareness share some features in common, these commonalities are not sufficient to show that the latter is a species of perception.

A.1 SELF-CONSCIOUSNESS, PERCEPTION, AND RECEPTIVITY

Should we conceptualize the various modes of bodily awareness as a group of internal senses? One way to go about answering this question would be to follow the method that Shoemaker employs in the course of trying to determine whether our knowledge of our mental states, which he refers to as "introspection", is a form of perceptual knowledge (1986). Shoemaker lists several features of paradigmatic modes of sense perception such as vision and asks whether introspection shares these features. Since it does not, he infers that it is not a perceptual capacity. Bermúdez (1998, 136) utilizes Shoemaker's criteria for the purpose of determining whether bodily awareness is a form of perception and I think this is a promising place to start.¹⁰⁶

Shoemaker writes that ordinary modes of perception admit of our perceiving, either simultaneously or in succession, a multiplicity of different objects which are on a par *qua* objects

¹⁰⁶ Bermúdez thinks that bodily awareness is both immune to error through misidentification relative to the first person and that it is a perceptual power. I argue in this section that these two properties are incompatible. He thinks that bodily awareness is a perceptual power because one's awareness of oneself through bodily awareness is integrated with one's knowledge of the environment (1998, 141-2). Even though this is correct, we can still distinguish the knowledge we have on the basis of bodily awareness from that which we have on the basis of the senses. My claim will be that the knowledge we have of ourselves on the basis of bodily awareness is not at once both self-conscious and perceptual.

of perception. It is possible for us to single out one object from this perceptual manifold and distinguish it from others by its properties and its spatiotemporal location. We can identify and re-identify perceived objects, either by applying explicit criteria of identity or by perceptually tracking them over time (1986, 126). It follows that it is also possible for one to misidentify an object one perceives, when one identifies the object otherwise than demonstratively. This occurs when one's perceptually based belief that a is F is knowledgeable insofar as it is a belief that something is F, but one makes a mistake about which object is F. For instance, I might see a certain figure approaching and judge that Jones will be here in five minutes, when in fact it is Smith who will be here in five minutes.

Crucially, it is also possible for me to misidentify a perceived object as myself. For instance, I might see a certain human being with a bump on his forehead and, believing that I am in front of a mirror, form the belief that I have a bump on my forehead. But it may be that I am in fact standing in front of a window rather than a mirror, and that I actually see my *doppelgänger* on the other side who has a bump on his forehead. In the terminology I employed in Chapter 2, beliefs about myself formed on the basis of observation are liable to *wh-*misidentification.

One way of putting the lesson of this example is that it is not in the *nature* of the power of visual perception that its object is oneself. In Shoemaker's language, insofar as I am an object of vision I am one of a multiplicity of objects that are on a par *qua* objects of vision. As such, if I form the belief that a certain object that I see is myself, this belief rests on a further judgment that I am that object. The fact that this judgment may be false means that it is possible for me to misidentify a visually perceived object as myself. And a similar point holds for the objects of the other four senses. The conclusion I drew from this in Chapter 2 is that my knowledge of myself

had on the basis of observation is not a form of self-consciousness. For an epistemic power is a self-conscious power if it is in its nature to provide one with knowledge of oneself as oneself, and beliefs about oneself formed on its basis are not liable to error through *wh-* misidentification relative to the first person pronoun.¹⁰⁷

The feature of ordinary modes of perception that implies they are not self-conscious capacities (at least as regards their objects) is the fact that they admit of our perceiving a multiplicity of objects that are on a par as objects of perception. It might be objected, however, that while this is a feature of the five traditional senses, it does not show that there cannot be other perceptual capacities that lack it. My reply is that this feature is not merely an accidental property of the “exteroceptive” senses, but follows from the concept of a perceptual capacity.

In the First *Critique* Kant tells us, “The capacity (receptivity) to acquire representations through the way in which we are affected by objects is called sensibility” (*CPR* A19/B33). What this definition brings out is that is that sense perception is essentially a capacity of *receptive* knowledge. But, as Kant makes clear, the object that a receptive epistemic capacity provides knowledge of is not, or at least not wholly, determined by the capacity itself but by whatever it is that *affects* it. This is not to say that it is impossible for the nature of a receptive epistemic capacity to be such as to partially determine which objects can affect it in a manner suitable for providing knowledge of them. For instance, vision can represent things only insofar as they are colored. But while the nature of the capacity may put some constraints on the kind of object it can represent, the *particulars* it represents are not determined by the nature of the capacity itself. This means that even if things are so arranged that it is nomologically impossible for anything other than object a to affect receptive epistemic capacity C, it is nonetheless still metaphysically

¹⁰⁷ See Chapter 2.3.

contingent, or contingent so far as we are considering the nature of the capacity, that C should represent a.

The upshot is that we can conclude from the fact that perceptual capacities are receptive epistemic capacities that they are not self-conscious capacities, insofar as they are receptive.¹⁰⁸ The reason that this is important is that it provides us with a principle for deciding whether bodily awareness is a kind of perception. If the knowledge of oneself that a certain mode of bodily awareness yields is not liable to *wh*- misidentification, it follows that it is not a perceptual capacity. And since it is the mark of self-conscious capacities that they yield knowledge that is immune to error through *wh*- misidentification, if bodily awareness is not liable to such error we will also have good reason to classify it as a form of self-consciousness.

A.2 IS BODILY AWARENESS RECEPTIVE KNOWLEDGE?

In order to determine whether the modes of bodily awareness are receptive epistemic capacities, we must investigate whether they provide one with knowledge of oneself that is immune to error through misidentification. *Prima facie*, it appears that they do. In *The Varieties of Reference* Evans writes:

None of the following utterances appears to make sense when the first component expresses knowledge gained in the appropriate way: ‘Someone’s legs are crossed, but is it my legs that are crossed?’; ‘Someone is hot and sticky, but is it I who am

¹⁰⁸ This final qualification is important. While our perceptual powers are not self-conscious powers as far as their objects are concerned, they provide one with self-conscious knowledge of oneself as the subject who is perceiving. For instance, when I touch something I know that I am the one doing the touching, and my knowledge of myself as the one touching is immune to *wh*- misidentification. Bermúdez infers from the fact that touch provides one with knowledge of oneself that is immune to such error as well as knowledge of a multiplicity of touched objects that it is both a self-conscious power and a perceptual one (1998, 137-9). This is correct in one sense, but not in a way that undermines the conclusion of this section, because it is not both self-conscious and perceptual in the same respect.

hot and sticky?'; 'Someone is being pushed, but is it I who am being pushed?' There just does not appear to be a gap between a subject's having information (or appearing to have information), in the appropriate way, that the property of being *F* is instantiated, and his having information (or appearing to have information) that *he* is *F*; for him to have, or to appear to have, the information that the property is instantiated just is for it to appear to him that *he* is *F*. (1982, 220-1)

There is also room to challenge Evans' claim, however. Recall that immunity to *wh*-misidentification is defined in modal terms: if a subject's judgment about herself is immune to *wh*-misidentification relative to a certain way of knowing then it is *not possible* for her to be entitled to judge that something is *F* by that way without also being entitled to hold that it is she that is *F*. I also said above that the sort of modality at issue here is metaphysical rather than nomological. But now one might object that while it is normally the case that a subject is entitled to hold that she is the one who, e.g., has her legs crossed, that is a metaphysically contingent fact. For instance, D.M. Armstrong writes, "We can conceive of being directly hooked-up, say by transmission of waves in some medium, to the body of another. In such a case we might become aware e.g. of the movements of another's limbs, in much the same sort of way that we become aware of the motion of our own limbs" (1984, 113). If it is possible to become aware of the states and motions of another's body through bodily awareness in this way, then it appears that we could also conceive of a scenario in which one misidentifies oneself on its basis.

We can imagine situations in which person *X* has been wired up to another *Y* in the way that Armstrong envisages, and we can also imagine that as a result of this procedure *X* becomes inclined to believe that someone's legs are crossed when either his legs or *Y*'s legs are crossed. The crucial question, though, is whether the conjunction of these facts is sufficient to show that

judgments made on the basis of bodily awareness are liable to *wh*- misidentification. Evans answers that it does not (1982, 221), and I will argue that he is right to do so.

There are two sorts of cases to consider. In the first sort, X does not know that he has undergone the procedure. We may imagine either that X has no inkling he is “hooked-up” to Y or that he knows only that he may be. In either case, if X thought his legs were crossed but then looked and saw they were not, he would have to revise his belief. But it is not the case that he would retain his entitlement for believing that someone’s legs are crossed. For all he knows this could be the case but it could simply be the case that his own nervous system is malfunctioning. All he would know is that it seemed to him as though his legs were crossed when they were not, that is, that he was subject to a kind of illusion.

In the second sort of case we are to imagine that X knows that he is hooked-up to Y and that his own nervous system is working properly. In this case, X may be entitled to believe that someone’s legs are crossed even though he is not entitled to judge that his own legs are crossed. For instance, X could see that his legs are not crossed and on this basis form the belief that Y’s must be. He would therefore also be entitled to assert the existential generalization of this belief. Yet this still does not show that beliefs made on the basis of bodily awareness are liable to *wh*- misidentification. In order for it to show that, the case would have to be one in which X would thereby come to know that someone’s (but not his own) legs are crossed *in the same way* that he knows his legs are crossed when he knows this in the usual way, and this is not the case. As in the first scenario, X’s capacity for bodily awareness provides an illusory appearance that his legs are crossed. The difference is that in this case X also knows that this illusion may be correlated with how things stand with Y’s legs. He is thus able to infer, partially on the basis of his illusory appearance, that someone’s legs are crossed. But forming this belief on the basis of making an

inference is not forming it on the same sort of basis on which one normally forms beliefs about one's own bodily position when one forms them on the basis of bodily awareness.¹⁰⁹

Lucy O'Brien puts forward a somewhat different argument for the claim that bodily awareness is not a self-conscious power. She claims that although in normal circumstances the content of bodily awareness is first personal, this is not an essential feature of it. Like Armstrong, O'Brien relies on counterfactual cases involving the cross-wiring of two people's nervous systems. She imagines a situation in which X has a device that connects her with another person Y in such a way that when the device is switched on it seems to X that her leg is bent when Y's is bent. Suppose that X does not know whether the device is on and has no other way of knowing how things stand with herself or with Y. O'Brien points out that in these circumstances even if X were to be aware of a bent leg through her capacity of bodily awareness, she would not know that it was her leg that was bent. This is because she would be in no position to rule out its being Y's leg rather than her own which was bent. O'Brien thinks that this scenario shows that it is implausible that in the envisioned situation X's capacity for bodily awareness provides her with knowledge that her own leg is bent (2007, 208).

O'Brien's conclusion appears to rest on the principle that if bodily awareness is essentially a self-conscious capacity, then the awareness that it provides must be sufficient by itself to rule out the possibility that Y's leg is bent in the circumstances she imagines. This principle is implausible. The fact that one can be mistaken as to whether her leg is bent does show that the relevant mode of bodily awareness is a fallible epistemic power. As I explain in Chapter 4, if a power is fallible in some respect it means that it is not necessarily the case that

¹⁰⁹ This final claim parallels Evans' response to Shoemaker's claim that memory is not immune to error through misidentification (244-5). Pryor (1999) misses this aspect of Evans' reply to Shoemaker in his evaluation of their debate.

whenever it is actualized one does what it is defined as a power to do. But if it is the nature of a fallible power to provide knowledge of oneself as oneself, one central way in which things could go awry would be if the power were interfered with in such a way that it purports to deliver knowledge that one is F when one is not. It is consistent with O'Brien's scenario to say that X does not know whether her leg is bent precisely because, given her background beliefs, she does not know whether she is suffering just such an illusion. But by hypothesis we are assuming that bodily awareness is a self-conscious capacity. Therefore, if X seems to be aware of a leg's being bent she seems to be aware that it is *her* leg that is bent, though she does not know that it is due to the fact that she cannot rule out the possibility that it merely seems to her that it is because Y's leg is bent.

A.3 A FAMILY OF NON-RECEPTIVE EPISTEMIC POWERS

Although I deny that bodily awareness is a perceptual capacity, I agree with proponents of the perceptual model that the two sorts of capacity have several important features in common. M.G.F. Martin motivates the thesis that bodily awareness is a form of perception by showing how it avoids the defects of a theory he refers to as "the traditional view" (1995, 268). According to the traditional view, bodily awareness provides us only with sensations that lack intentional content. We normally speak of bodily sensations as being located in different regions of the body, but proponents of the traditional view interpret our speech here as locating the causes of these sensations. Sensations themselves are conceived of as unlocated psychological entities. For instance, according to the traditional view when I say that I have a pain in my toe,

this means that I am experiencing a painful sensation which is being caused by a disturbance in my toe.¹¹⁰

Martin claims, rightly in my view, that the traditional view is not phenomenologically plausible, even for those modes of bodily awareness to which it most readily applies, such as the capacity to feel pain. If I stub my toe, then it seems to me that the pain I feel is not merely caused by a disturbance of my stubbed toe, but is itself located there. While pain is a qualitative state, and so is subjective in one sense, it is nonetheless a state of a sensitive part of an animal body (cf. Brewer 1995, Hyman 2003).

Moreover, the traditional view is not applicable to every mode of bodily awareness. Consider proprioception, or our ability to know the relative positions and motions of our limbs and other bodily parts. If the traditional view is correct, then if I am proprioceptively aware that my leg is bent this is presumably because I feel a certain sensation and infer from this fact that it is bent. But as Anscombe points out, this inferential picture will not work. The reason is that there is no way to describe the sort of sensation in question except in terms of the way it feels when my body is positioned a certain way (1962; 1963, §8). If I close my eyes and bend my leg I do not feel something that could be described in terms of a non-located tingle or pressure which I notice is correlated with my leg's being bent. Rather, the only way I can describe the sensation I have is as the kind that that one has when one's leg is bent.

While the traditional view is unsatisfactory in these respects, there is a way in which the deliverances of bodily awareness and those of our perceptual capacities are analogous. Just as one's visual system purports to provide one with an awareness that some object is colored in a certain way, so too one's capacity to detect pain purports to make one aware that some part of

¹¹⁰ See, for example, Ryle 1949, 105.

one's body hurts and one's capacity for proprioception purports to deliver knowledge of the positions and motions of one's limbs. And just as one's purported awareness that something is a certain color can be illusory, so too can one's purported awareness that one's leg hurts or is bent. The possibility of illusion is shown by cases such as referred pain and the existence of phantom limbs, in which amputees appear to feel limbs that no longer exist.

I agree that the traditional model is inadequate and that bodily awareness shares some common features with perception. Proponents of the perceptual model such as Martin tend to slide from the good point that bodily awareness is a set of powers that purports to provide one with knowledge of one's body to the conclusion that this set of powers is perceptual.¹¹¹ But all that the commonalities show is that both bodily awareness and perception are epistemic powers, or *powers to know* how things stand with certain aspects of some object or objects. While perceptual powers are a species of epistemic powers, the fact that our modes of bodily awareness fall under the same genus does not entail they are also the same species. Again, the fact that states of awareness that are the upshot of a certain mode of bodily awareness can be either veridical or illusory does not show that that mode is a perceptual power. It shows only that it is a fallible epistemic power.

I have already explained why bodily awareness is not a perceptual power in Sections I and II above. Perceptual powers are essentially powers of receptive knowledge, and this means that any knowledge of oneself that they provide is liable to error through *wh*-misidentification. But the knowledge that we have of ourselves based on bodily awareness is immune to such error. The different modes of bodily awareness are epistemic powers that are self-conscious and non-perceptual in nature.

¹¹¹ O'Brien makes a similar slide (2007, ch. 10).

BIBLIOGRAPHY

Alvarez, Maria and John Hyman 1998. "Agents and their Actions", *Philosophy* Vol. 73, No. 284, pp. 219-45.

Anscombe, G.E.M. 1958. "Modern Moral Philosophy," in *Collected Philosophical Papers, Vol. III: Ethics, Religion and Politics*, Oxford: Basil Blackwell, 1981, pp. 26-42.

----- 1962. "On Sensations of Position", in Anscombe 1981, pp. 71-4.

----- 1963. *Intention*, 2nd Edition, Ithaca, NY: Cornell University Press.

----- 1974. "Practical Inference", in Anscombe 2005, pp. 108-47.

----- 1971. "Causality and Determination", in Anscombe 1981, pp. 133-47.

----- 1975. "The First Person", in Cassam, ed. 1994, pp. 140-59.

----- 1981. *The Collected Papers of G.E.M. Anscombe Vol. II: Metaphysics and the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press.

----- 1983. "The Causation of Action", in Anscombe 2005, pp. 89-108.

----- 2005. *Human Life, Action and Ethics: Essays by G.E.M. Anscombe*, ed. by Mary Geach and Luke Gormally, Exeter: Imprint Academic.

Aristotle 1984. *The Complete Works of Aristotle*, 2 vols., ed. by Jonathan Barnes, Princeton, NJ: Princeton University Press.
DA: De anima; DM: De motu animalium, HA: Historia animalium NE: Nicomachean Ethics.

Armstrong, D.M. 1984. "Consciousness and Causality", in *Consciousness and Causality*, Armstrong, D.M., and Malcolm, N. eds., Oxford: Basil Blackwell.

Bargh J. A., M. Chen, and L. Burrows 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action", *Journal of Personality and Social Psychology*, Vol. 71, pp. 230-44.

- Beaney, Michael, ed. 1997. *The Frege Reader*, London: Blackwell.
- Bermúdez, José Luis 1998. *The Paradox of Self-Consciousness*, Cambridge, MA: The MIT Press.
- Bermúdez, José Luis, Anthony Marcel, and Naomi Eilan, eds., 1995. *The Body and the Self*, Cambridge, MA: The MIT Press.
- Bishop, J. 1989. *Natural Agency: An Essay on the Causal Theory of Action*, Cambridge: Cambridge University Press.
- Boethius 518. *De persona et duabus naturis*, *Patrologia Latina* 64.
- Boyle, Matthew and Douglas Lavin, 2010. “Goodness and Desire”, in Tenenbaum, ed. 2010, pp. 161-201.
- Bratman, Michael E. 1991. “Cognitivism about Practical Reasoning”, in *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge: Cambridge University Press, 1999, pp. 250-64.
- 2001. “Two Problems about Human Agency” in *Structures of Agency: Essays*, Oxford: Oxford University Press, 2007, pp. 89-105.
- Brewer, Bill 1995. “Bodily Awareness and the Self”, in Bermúdez, Marcel, and Eilan, eds., pp. 291-310.
- Broome, John 2002. “Practical Reasoning”, *Reason and Nature: Essays in the Theory of Rationality*, ed. by José Bermúdez and Alan Millar, Oxford: Oxford University Press, pp. 83-102.
- Cassam, Quassim, ed. 1994. *Self-Knowledge*, Oxford: Oxford University Press.
- 1997. *Self and World*, Oxford: Clarendon Press.
- Castañeda, H-N 1967. “On the Logic of Self-Knowledge,” *Noûs*, 1, pp. 9-21.
- Chang, Ruth 1997. “Introduction”, *Incommensurability, Incomparability, and Practical Reason*, ed. by Ruth Chang, Cambridge, MA: Harvard University Press, pp. 1-34.
- Chisholm, Roderick 1964. “Human Freedom and the Self”, in *Free Will*, 2nd edition, ed. by Gary Watson, Oxford: Oxford University Press, 2003, pp. 26-37.
- 1968. “Comments and Replies”, *Philosophia* 7, 597-636.
- Clark, Philip 2001. “Velleman’s Autonomism”, *Ethics*, 111, pp. 580-93.

- Cohon, Rachel 2008. *Hume's Morality: Feeling and Fabrication*, Oxford: Oxford University Press.
- Cooper, John 1986. *Reason and the Human Good in Aristotle*, Indianapolis, IN: Hackett Publishing Company.
- Davidson, Donald 1963. "Actions, Reasons, and Causes", in Davidson 2001, pp. 1-19.
- 1969. "How is Weakness of Will Possible?", in Davidson 2001, pp. 21-42.
- 1971. "Agency", in Davidson 2001, pp. 43-62.
- 1973. "Freedom to Act", in Davidson 2001, pp. 63-82.
- 1978. "Intending", in Davidson 2001, pp. 83-102.
- 1987. "Problems in the Explanation of Action", in *Problems of Rationality*, Oxford: Clarendon Press, 2004, pp.101-16.
- 2001. *Essays on Actions and Events*, Oxford: Clarendon Press.
- Descartes, René 1974-98. *Œuvres de Descartes*, 11 vols., ed. by Charles Adam and Paul Tannery, Paris: Vrin.
- 1984-1991. *The Philosophical Writings of Descartes*, 3 vols., John Cottingham, Robert Stoothoff, and Dugald Murdoch, trans., Cambridge: Cambridge University Press.
- Donnellan, Keith 1963. "On Knowing What I'm Doing", *Journal of Philosophy* 60, pp. 401-9.
- Eliot, T.S. 1991. *Collected Poems, 1909-1962*, New York, NY: Harcourt Brace & Company.
- Enç, Berent 2003. *How We Act: Causes, Reasons, and Intentions*, Oxford: Clarendon Press.
- Euclid 2007. *Euclid's Elements*, trans. by T. L. Heath, Santa Fe, NM: Green Lion Press.
- Evans, Gareth 1982. *The Varieties of Reference*, ed. by John McDowell, Oxford: Clarendon Press.
- Falvey, Kevin 2000. "Knowledge in Intention", *Philosophical Studies* 99, pp. 21-44.
- Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person", in Frankfurt 2005, pp. 11-25.
- 1987. "Identification and Wholeheartedness", in Frankfurt 2005, pp. 159-76.

- 2005. *The Importance of What We Care About*, Cambridge: Cambridge University Press.
- Frege, Gottlob 1892. “On *Sinn* and *Bedeutung*”, in Beaney ed., 1997, pp. 151-71.
- 1918. “Thought”, in Beaney ed., 1997, pp. 325-45.
- Goldman, Alvin 1970. *A Theory of Human Action*, Princeton, NJ: Princeton University Press.
- Grice, H.P. 1961. “The Causal Theory of Perception”, *Proceedings of the Aristotelian Society, Supplementary Volume*, Vol. 35, 121-52.
- 1971. “Intention and Uncertainty”, *Proceedings of the British Academy* 57, pp. 263-79.
- Hampshire, Stuart 1959. *Thought and Action*, Notre Dame, IN: University of Notre Dame Press.
- Harman, Gilbert 1997. “Practical Reasoning”, *The Philosophy of Action*, ed. by A. Mele, Oxford: Oxford University Press, pp. 149-77.
- 1999. “Rationality”, in *Reasoning, Meaning and Mind*, Oxford: Clarendon Press, pp. 9-45.
- Harré, R. and Madden, E.H. 1975. *Causal Powers: A Theory of Natural Necessity*, Oxford: Basil Blackwell.
- Hegel, G.W.F. 1977. *Phenomenology of Spirit*, trans. by A.V. Miller, Oxford: Oxford University Press.
- 1991. *Elements of the Philosophy of Right*, ed. by, Allen W. Wood, trans. by H.B. Nisbet, Cambridge: Cambridge University Press.
- Hornsby, Jennifer 1980. *Actions*, London: Routledge and Kegan Paul.
- 1997. “Action and the Mental-Physical Divide”, in *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind*, Cambridge, MA: Harvard University Press, pp. 83-92.
- 2004a. “Agency and Actions”, in Hyman and Steward, eds., 2004, pp. 1-24.
- 2004b. “Agency and Alienation”, in *Naturalism in Question*, ed. by Mario De Caro and David Macarthur, Cambridge, MA: Harvard University Press, pp. 173-87.
- n.d. “Actions in their Circumstances”, ms.

- Hume, David 2000. *A Treatise of Human Nature*, ed. by D.F. Norton and M.J. Norton, Oxford: Clarendon Press.
- Hyman, John 2003. "Pains and Places", *Philosophy*, Vol. 78, No. 1, pp. 5-24.
- 2006. "Three Fallacies about Action", *Proceedings of the 29th International Wittgenstein Symposium*, Hölder-Pichler-Tempsky, pp. 137-63.
- Hyman, John and Helen Steward, eds. 2004. *Agency and Action*, Cambridge: Cambridge University Press.
- Irwin, T.H. 1992. "Who Discovered the Will?", *Philosophical Perspectives*, Vol. 6, *Ethics*, ed. by James E. Tomberlin, Atascadero, CA: Ridgeview, pp. 453-73.
- Johnston, Mark 1988. "Self-Deception and the Nature of Mind", in *Perspectives on Self-Deception*, ed. Brian P. McLaughlin and Amélie Oksenberg Rorty, Berkeley, CA: University of California Press, pp. 63-91.
- Kant, Immanuel 1999. *Critique of Pure Reason*, trans. and ed. by Paul Guyer and Allen Wood, Cambridge: Cambridge University Press.
- 2005. *Critique of the Power of Judgment*, ed. and trans. by Paul Guyer and Eric Matthews, Cambridge: Cambridge University Press.
- Korsgaard, Christine 1998. "The Normativity of Instrumental Reasoning" in *Ethics and Practical Reasoning*, ed. by Garrett Cullity and Berys Gaut, Oxford: Clarendon Press, pp. 215-54.
- Lowe, E.J. 1981, "All Actions Occur Inside the Body", *Analysis* Vol. 41, No. 3, pp. 126-9.
- 2008. *Personal Agency*, Oxford: Oxford University Press.
- MacIntyre, Alasdair 1999. *Dependent Rational Animals*, Chicago, IL: Open Court.
- Martin, M.G.F. 1995. "Bodily Awareness: A Sense of Ownership", in Bermúdez, Marcel, and Eilan eds., pp. 267-90.
- McCann, Hugh J. 1975. "Trying, Paralysis, and Volition", in *The Works of Agency: On Human Action, Will, and Freedom*, Ithaca, NY: Cornell University Press, 1998, pp. 94-109.
- McDowell, John 1982. "Criteria, Defeasibility, and Knowledge", in *Meaning, Knowledge, and Reality*, Cambridge, MA: Harvard University Press, 1998, pp. 369-94.
- 1996. *Mind and World*, Cambridge, MA: Harvard University Press.

- 1998. "Referring to Oneself", in Hahn, ed. *The Philosophy of P.F. Strawson*, Chicago, IL: Open Court Press, pp. 129-45.
- McGinn, Colin 1983. *The Subjective View: Secondary Qualities and Indexical Thoughts*, Oxford: Clarendon Press.
- Moran, Richard 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.
- 2004. "Anscombe on 'Practical Knowledge'", in *Agency and Action*, John Hyman and Helen Steward, eds., pp. 43-68.
- Moravcsik, Julius 1994. "Essences, Powers, and Generic Propositions", in T. Scaltsas, D. Charles, and M.L. Gill, eds., *Unity and Explanation in Aristotle's Metaphysics*, Oxford: Oxford University Press, pp. 229-44.
- Nagel, Thomas 1970. *The Possibility of Altruism*, Princeton, NJ: Princeton University Press.
- 1986. *The View from Nowhere*, Oxford: Oxford University Press.
- O'Brien, Lucy 2007. *Self-Knowing Agents*, Oxford: Oxford University Press.
- Olson, Eric T. 1997. *The Human Animal: Personal Identity without Psychology*, Oxford: Oxford University Press.
- O'Shaughnessy, Brian 1974. "Trying (as the Mental 'Pineal Gland')", *Journal of Philosophy*, Vol. 71, 365-86.
- 1980. *The Will: A Dual Aspect Theory*, 2 Vols., Cambridge: Cambridge University Press.
- Paul, Sarah K. 2009. "How We Know What We're Doing", *Philosophers' Imprint*, Vol. 9, No. 11.
- Perry, John 1979. "The Problem of the Essential Indexical", in Cassam, ed. 1994, pp.167-83.
- Prichard, H.A. 1945. "Acting, Willing, and Desiring", in *Moral Obligation*, Oxford: Clarendon Press, 1949, pp. 187-98.
- Pryor, James 1999. "Immunity to Error through Misidentification", *Philosophical Topics*, Vol. 26, No. 1&2, pp. 271-304.
- Raz, Joseph 1999. "Agency, Reason, and the Good," in *Engaging Reason: On the Theory of Value and Action*, Oxford, Oxford University Press, pp. 22-45.
- Rödl, Sebastian 2007. *Self-Consciousness*, Cambridge, MA: Harvard University Press.

- Rozemond, Marleen 1998. *Descartes's Dualism*, Cambridge, MA: Harvard University Press.
- Ryle, Gilbert 1949. *The Concept of Mind*, Chicago, IL: University of Chicago Press.
- Sartre, Jean-Paul 1984. *Being and Nothingness*, New York, NY: Washington Square Press.
- Schopenhauer, Arthur 1966. *The World as Will and Representation*, 2 vols., trans. by E.F.J. Payne, New York, NY: Dover Publications.
- Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.
- Setiya, Kieran 2007. *Reasons without Rationalism*, Princeton, NJ: Princeton University Press.
- 2008. "Practical Knowledge", *Ethics* Vol. 118, pp. 388-409.
- 2010. "Sympathy for the Devil", in Tenenbaum, ed. 2010, pp. 82-110.
- n.d. "Knowledge of Intention", ms.
- Shah, Nishi 2008. "How Action Governs Intention", *Philosophers' Imprint* Vol. 8, No. 5, pp. 1-19.
- Shoemaker, Sydney 1968. "Self-Reference and Self-Awareness", in Cassam, ed. 1994, pp. 80-93.
- 1986. "Introspection and the Self", in Cassam, ed. 1994, pp. 118-39.
- Smith, A.D. 1988. "Agency and the Essence of Actions", *The Philosophical Quarterly* Vol. 38, No. 153, pp. 401-21.
- Smith, Michael 1987. "The Humean Theory of Motivation", *Mind*, 96, pp. 36-61.
- 1994. *The Moral Problem*, Oxford: Blackwell.
- 2004. "The Structure of Orthonomy", in Hyman and Steward, eds., pp. 165-93.
- Stocker, Michael 1979. "Desiring the Bad: An Essay in Moral Psychology", *Journal of Philosophy* 76, 738-53.
- Strawson, P.F. 1963. *Individuals: An Essay in Descriptive Metaphysics*, Garden City, NY: Doubleday and Company, Inc.
- 1966. *The Bounds of Sense*, London: Methuen.

- Tenenbaum, Sergio 2007. *Appearances of the Good: An Essay on the Nature of Practical Reason*, Cambridge: Cambridge University Press.
- ed., 2010. *Desire, Practical Reason, and the Good*, Oxford: Oxford University Press.
- Thompson, Michael 2008. *Life and Action: Elementary Structures of Practice and Practical Thought*, Cambridge, MA: Harvard University Press.
- n.d. "Anscombe's *Intention* and Practical Knowledge", ms.
- Vallacher, R. R. and D. M. Wegner 1987. "What Do People Think They're Doing? Action Identification and Human Behavior", *Psychological Review*, Vol. 101, pp. 44-52.
- Velleman, J. David 1992a. "The Guise of the Good", in Velleman 2000b, pp. 99-122.
- 1992b. "What Happens When Someone Acts?", in Velleman 2000b, pp. 123-133.
- 2000a. "Introduction", in Velleman 2000b, pp. 1-31.
- 2000b. *The Possibility of Practical Reason*, Oxford: Clarendon Press.
- 2007. *Practical Reflection*, Stanford, CA: CSLI Publications.
- Von Wright, G.H. 1963. *Norm and Action*, London: Routledge and Kegan Paul.
- Wallace, R. Jay 1999. "Addiction as Defect of the Will: Some Philosophical Reflections", in Wallace 2006, pp. 165-89.
- 2000. "Caring, Reflexivity, and the Structure of Volition", in Wallace 2006, pp. 190-211.
- 2006. *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason*, Oxford: Clarendon Press.
- Watson, Gary 1975. "Free Agency", in Watson 2004, pp. 13-32.
- 1999. "Disordered Appetites: Addiction, Compulsion, and Dependence", in Watson 2004, pp. 59-87.
- 2004. *Agency and Answerability: Selected Essays*, Oxford: Clarendon Press.
- Wedgwood, Ralph 2008. "Review of *Reasons without Rationalism*", *Mind*, Vol. 117, No. 468, pp. 1130-5.
- Wiggins, David 2001. *Sameness and Substance Renewed*, Cambridge: Cambridge University Press.

Wittgenstein, Ludwig 1958a. *Philosophical Investigations*, trans. by G.E.M. Anscombe, Oxford: Basil Blackwell.

----- 1958b. *The Blue and Brown Books*, Oxford: Basil Blackwell.

----- 2001. *Tractatus Logico-Philosophicus*, trans. by D.F. Pears and B.F. McGuinness, London and New York, NY: Routledge.

Wright, Crispin 1998. "Self-Knowledge: The Wittgensteinian Legacy", in Wright, Smith, and Macdonald, eds., *Knowing Our Own Minds*, Oxford: Clarendon Press, pp. 13-46.

Yaffe, Gideon 2000. *Liberty Worth the Name: Locke on Free Agency*, Princeton, NJ: Princeton University Press.