MENTAL CAPACITIES AND THEIR IMPERFECT EXERCISES:

THE ESSENTIAL NORMATIVITY OF THE MIND

by

Kim Chandrasekhar Frost

Bachelor of Arts (Hons), University of Sydney, 2001

Submitted to the Graduate Faculty of

Philosophy in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Kim Chandrasekhar Frost

It was defended on

23rd July 2012

and approved by

Robert Brandom, Distinguished Professor, Philosophy

Michael Thompson, Professor, Philosophy

Peter Machamer, Professor, History and Philosophy of Science

Dissertation Advisor: Kieran Setiya, Associate Professor, Philosophy

MENTAL CAPACITIES AND THEIR IMPERFECT EXERCISES:

THE ESSENTIAL NORMATIVITY OF THE MIND

Kim Frost, PhD

University of Pittsburgh, 2012

I develop Anscombe's distinction between mistakes in judgment and mistakes in performance into a novel account of intentional action and the metaphysics of mind.

Anscombe's distinction is usually understood in terms of the "direction of fit" possessed by different kinds of mental states. In Chapter 1 I argue that direction of fit is a hopeless idea. Direction of fit is guided by intuitions of symmetry, but those intuitions are misguided: there are ineliminable asymmetries between the mind's theoretical and practical activity. I further argue that Anscombe's distinction is best understood not in terms of direction of fit, but in terms of mental activity that is partially constituted by norms.

In Chapter 2 I develop a theory of *fallible capacities*: capacities that sometimes issue in *mistakes*. Fallible capacities are essentially normative because norms are built into their logical structure. I argue for the essential normativity of the mind on the basis of the claim that the fallible capacity *to know* is essential to minds like ours. The primary dialectical opponent to the argument is a reductive naturalist, who accepts the appearance that we have a fallible capacity to know, but offers a reductive account of what it is to possess that capacity. I argue that such reductive accounts fail; the options are to reject the appearances outright, or to accept that the mind is essentially normative.

Chapter 3 solves a problem about action individuation. A basic action is one that is not performed by means of some other action. Basic action theorists say that all intentional actions decompose into basic actions; Michael Thompson says that none do. I argue that neither view is

right, because action individuation is up to individual agents themselves. I further argue that this independently plausible conception of intentional action is, in one key respect, best accommodated by the theory of fallible capacities outlined in Chapter 2, because that theory can explain why mistakes in performance fall into the logical category of *particulars*; something that traditional basic action theory and Thompson's view cannot easily explain.

TABLE OF CONTENTS

PREFACE

I have been uncommonly fortunate to study philosophy at the University of Pittsburgh, and I owe many debts of gratitude.

First, thanks are due to the members of my committee. I came to Pittsburgh in 2004 as a Visiting Scholar under the generous auspices of Bob Brandom. I had no clear ambition to make a career of the study of philosophy at the time. It is through conversations with Bob, and attention to his work, and the few months I spent in the department, that I changed my mind, and I am very happy that I did so. When I started as a graduate student, I also had little idea what philosophy of action was, or why it might be important. In this respect, I owe Michael Thompson a great intellectual debt, as will be obvious from the text of the final chapter in this dissertation. When I was at a low ebb, and wondering whether to pack it all in, it was some pointed challenges from Peter Machamer that spurred me to complete the account of fallible capacities, in the face of what seemed at the time to be insuperable difficulties. Each in different ways deserves a "without whom not".

Special thanks must go to my advisor Kieran Setiya. Kieran, through conversation, criticism and example, has taught me more than almost anyone else about *how* to do philosophy, and there really would be no dissertation were it not for his unfailing encouragement and support. I have, as best I can, tried to live up to the high standards he has taught me. I will take those standards with me when I move on, and I am not sure how to repay *that* debt, which is enormous, except by paying

it forward. (Should I have students of my own to advise some day, I have no doubt I will be saying to myself, *sotto voce*, "What would Kieran say?")

I would also like to thank the members of the philosophy faculty at Pitt more generally. You just can't get an education like this anywhere else in the world. (And they *pay* you for it!) In particular I would like to thank James Shaw, for teaching me a great deal about how to teach, and John McDowell, for his advice, help and philosophical example.

The graduate student community at Pitt is uniquely lovable. For conversation and camaraderie I would particularly like to thank Daniel Addison, Ben Breuer, Rachael Driver, Tim Greenfield, Dorcinda Knauth, Kathryn Lindeman, David Matthews, Brandi Neale, Sasha Newton, Stephen Makin, Evgenia Mylonaki, Jesse O'Brien, James Pearson, Ben Schultz, Greg Strom, and Tim Willenken. Tyke Nunez deserves special thanks for long, illuminating conversations about almost any aspect of philosophy. I hope I haven't left anybody out: if I have then I am a forgetful wretch, but not an ungrateful one.

Michael Cuccaro and Stacy Hoffman have been wonderful friends over the years. I am extremely lucky to have met them, and they have made me feel right at home in the 'burgh.

Thanks are also due to Paul Redding, my advisor in Sydney back in the day, who let me skip out of a Masters thesis with never a word of reproach and only words of encouragement.

My family's love and encouragement has been a blessing. I look forward to the day when we are not separated by an ocean and can see each other more often. (A year between visits is too long!)

I am not sure how to find the right words to thank Hille Paakkunainen. With her, I am *living well*: nothing has been plainer in my life. And she's a damn good philosopher too. Much of what is right and intelligible in what follows would be wrong and garbled were it not for her keen insight, loving support, and uncanny good judgment.

INTRODUCTION

What are mistakes? We are quite familiar with the experience of error: of making mistakes and being mistaken. Other higher animals seem to share in this liability to err. But God (presumably) doesn't make mistakes, and things without minds, like atoms, rocks, and pistons, don't make mistakes either. Mistakes seem to be coordinated with – perhaps even an essential feature of – minds like ours.

My dissertation consists in three loosely connected investigations of mistakes and their place in our mental lives. The investigations are united by a common theme: namely, that fallibility enters into the constitution of our powers of thought at a metaphysically fundamental level.

Chapter 1: On the Very Idea of Direction of Fit

My starting point is Anscombe's distinction between mistakes in judgment and mistakes in performance, as this distinction is applied to our powers of theoretical and practical thought. Anscombe's distinction is usually understood in terms of the "direction of fit" possessed by different kinds of mental states. According to direction of fit theory, some mental states (beliefs) have a *thetic* direction of fit, in that they "aim at truth" or "ought to fit" the world, whilst other mental states (desires, or intentions) have a *telic* direction of fit, in that they "aim at realization" or the world "ought to fit" them.

In Chapter 1 I argue that the very idea of direction of fit is a hopeless one. The two directions of fit are supposed to be determinations of one and the same determinable two-place relation, differing only in the ordering of favored terms. But there is no determinable of which the

two directions of fit are symmetrically related determinations, because there are ineliminable asymmetries between the way that beliefs "aim at truth" and the way that desires (or intentions) "aim at realization". The vast array of views in philosophy of mind and moral psychology that rest on a conception of direction of fit all rest on a mistake.

I further argue that Anscombe's distinction is best understood not in terms of direction of fit, but in terms of *standards of correctness*. Beliefs are partially constituted by a standard of correctness of truth (or knowledge) against which a *mistake in judgment* may be measured. Beliefs do not share this generic property with another kind of mental state, such as desire or intention, but rather with a kind of event: what I call *telic events*. Telic events are partially constituted by the standard of correctness of success (or practical knowledge), against which a *mistake in performance* may be measured. This shift in emphasis, from comparative analyses of kinds of mental state to comparative analyses of normatively-constituted states and events, accounts for the asymmetries that direction of fit theory cannot, and promises to deliver an argument for the essential normativity of the mind.

Chapter 2: Why the Mind Is Essentially Normative

In Chapter 2 I go on to develop an account of the nature of the mind that accommodates mistakes at a metaphysically fundamental level, and explains what mistakes are. Most contemporary theories of capacities (abilities, dispositions etc.) hold that any capacity to *A* has only perfect exercises, where a perfect exercise is a way for *A* to be wholly manifested in some state or episode. Taking inspiration from Aristotle's conception of rational capacities, I offer a theory of *fallible capacities*. Fallible capacities have perfect and imperfect exercises. An imperfect exercise of the fallible capacity to *A* is not a way for *A* to be wholly manifested in some state or episode, but is instead a *mistake*: a way of *precluding* perfect exercise on that occasion in that regard. For example, a mistake in judgment (e.g.

believing that not-p) precludes knowing that *p* on some occasion. The theory is quite general and provides a powerful and versatile explanatory tool.

Drawing on the theory, I argue for the essential normativity of the mind on the basis of the claim that the fallible capacity *to know* is essential to minds like ours. It certainly seems as if we can know, and that we can merely believe, and that these cognitive powers have some common source, and that the common source is oriented towards knowledge as cognitive *success*. Call these the *appearances*. The appearances would be well explained by positing the fallible capacity to know as essential to minds like ours. But fallible capacities are essentially normative, because they are inherently "positively valenced" towards one of a pair of mutually opposed exercises (for example: knowing, rather than merely believing). If the fallible capacity to know is essential to minds like ours, the mind is essentially normative.

The primary dialectical opponent to my argument is a reductive naturalist, who accepts the appearances, but offers a reductive, non-normative account of what it is to possess the fallible capacity to know, made out solely in terms of non-fallible capacities (or their functional-dispositional equivalents). I argue that any such reduction must fail to capture the appearances, and can only give us an alienated conception of the powers of thought of the thinking subject. The options are to reject the appearances, or accept that the mind is essentially normative.

Chapter 3: The Antinomy of Basic Action

Chapter 3 begins as an extended meditation on the mereological structure of intentional action. Traditional action theorists must choose whether to believe in *basic actions* or not, where basic actions are things that agents simply do intentionally, without the mediation of doing anything else intentionally as means to that end. The orthodox position is to accept basic actions. In *Life and Action,* Michael Thompson gives a powerful argument against the very idea of basic action, based on

little more than common sense considerations about the continuity of time: any purportedly basic action has temporal parts, and the parts are rationalized by the whole as compositional means, so the parts are (more basic) intentional actions in their own right. The disagreement between the orthodox view and Thompson's represents a kind of antinomy, where both poles of the antinomy have their own absurdity.

I argue against each pole of the antinomy, disputing the claims of necessity. It is neither the case that there must be basic actions in every expression of the will, nor is it the case that there must not be, but usually there *are* basic actions when the will expresses itself. This is because for the most part, and within certain limits, it is up to individual agents themselves to carve up the instrumental joints of what they have in mind to do, by conceiving of what they do in terms of potential successes or failures.

The connection of Chapter 3 to the earlier chapters only emerges in the second half of the chapter, once the importance of the individual agent's practical conception of the case has been brought into view. Intentional actions are discrete events in history: they fall into the logical category of *particulars*. The same seems to be true of mistakes in performance, because they are cases of doing *something* that rules out (precludes) doing what one has in mind to do. If we conceive of intentional actions as exercises of the fallible capacity to do what one has in mind to do, we can account for the fact that mistakes in performance fall into the logical category of particulars; something that traditional basic action theory and Thompson's view cannot easily explain.

Overview

Taken together, the three investigations lay the groundwork for what I take to be a coherent, explanatorily powerful and genuinely novel conception of the mind and its essential powers of thought. The investigations also expose some of the more tempting errors in theorizing about the

nature of the mind and intentional action. But most importantly, the three investigations offered here articulate a conception of the mind that helps to make sense of ourselves not as complicated machines, nor yet as little Gods imprisoned in the Cartesian Theater, but as the more or less bumbling, thoughtful monkeys that we are.

1.0 ON THE VERY IDEA OF DIRECTION OF FIT

1.1 INTRODUCTION

Many philosophers draw a distinction between practical and theoretical thought. Drawing the distinction precisely is difficult, but the rough idea is easy to grasp. Theoretical thought aims at reflecting the world as it is (was; will be; could be; etc.). Practical thought aims at getting things done, which may change how the world is (from how it is at the moment).

How can we draw the distinction between practical and theoretical thought more precisely? Some philosophers have looked to Anscombe for inspiration:

> Now let us consider a man going around town with a shopping list in his hand. Now it is clear that the relation of this list to the things he actually buys is one and the same whether his wife gave him the list or it is his own list; and that there is a different relation when a list is made by a detective following him about. If he made the list himself, it was an expression of intention; if his wife gave it him, it has the role of an order. What then is the identical relation to what happens, in the order and the intention, which is not shared by the record? It is precisely this: if the list and the things that the man actually buys do not agree, and this and this alone constitutes a *mistake*, then the mistake is not in the list but in the man's performance (if his wife were to say: 'Look, it says butter and you have bought margarine', he would hardly reply: 'What a mistake! We must put that right' and alter the word on the list to 'margarine'); whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record.[1]

[1] Anscombe (1957: 56)

Anscombe's story of the shopper and the detective has provoked several theories about the *direction of fit* (DOF) an intentional mental state might have with regard to its object. According to DOF theory, there are two and only two directions of fit (DOFs). The first, which we shall call the *thetic* DOF, flows inwards from the world to the mind, so that some states of mind (beliefs) "aim at the truth" about the world, or "ought to fit" how things are with the world. The second, which we shall call the *telic* DOF, flows outward from the mind to the world, so that some states of mind (pro-attitudes, or desires, or maybe intentions) "aim at realization" or are such that the world "ought to fit" them.[2] These slogans present an image of symmetry at work in the thetic and telic DOFs: whatever the thetic relation of mind to world is, the telic relation is somehow the "flipside" version of the same. The slogans differ primarily in the ordering of terms, suggesting that a developed theory of DOF could distinguish practical from theoretical thought not just by means of slogans (images, metaphors, stories, etc.) but by means of a deep, general and precise *logical* contrast – the mark of respectability in philosophy.

DOF was introduced to contemporary philosophy of mind as a normative notion, but there are both normative and descriptive theories of DOF.[3] Normative theories use normative terms like 'responsibility' and 'ought (to fit)' in the statement of what DOF consists in, whilst descriptive theories use terms that refer to causal relations or dispositions. Different theories of DOF have been employed in support of a number of influential and controversial views in philosophy of mind and moral psychology. The language of different and opposed directions suggests that the two DOFs

---

[2] Humberstone (1992) came up with the terms 'thetic' and 'telic'. I use these terms rather than those of Searle (1983) because Searle's 'mind-to-world' and 'world-to-mind' terminology is potentially confusing. Searle's terminology fits normative theories of DOF reasonably well, where the term 'mind-to-world' is, for instance, shorthand for 'the mind ought to fit the world' (the thetic DOF). But Searle's terms have everything backwards when applied to descriptive theories of DOF, where 'mind-to-world' is most naturally thought of as a direction of causation, and so would be a term for the telic rather than the thetic DOF.

[3] Searle (1979; 1983) is responsible for introducing the term 'DOF' to contemporary philosophy of mind. Austin (1953) used the term earlier than Searle, but Austin's distinction and purposes were somewhat different to those of most DOF theorists, so I do not deal with him in this chapter.

exclude each other somehow, so that we could use them to do some psychological taxonomy. Psychological taxonomy is not as boring as it sounds. On the basis of a theory of DOF the claim is sometimes made that no mental state could have both DOFs, or be both thetic and telic, and this has consequences for live philosophical debates about whether beliefs could independently motivate one to act (or whether a more Humean theory of motivation is true), whether virtue could be a form of knowledge, and exactly what it takes for a nearby event to count as one's own full-blooded intentional action.[4]

Apart from its use in support of substantive (and influential) contemporary philosophical views, and apart from its more modest use as a constraint on exercises in psychological taxonomy,[5] a true theory of DOF promises to explain the unity and difference of theoretical and practical *reason* and *knowledge*. We can think of reasoning as a formally structured movement in thought (i.e. from one set of thoughts to another, where the former require, or at least count in favor of, the latter), and of knowledge as non-accidental agreement of thought with its object. (Nothing hangs on these characterizations: I pick them up only in order to make a broad point about the stakes). *If* our thoughts are divided in kind by two fundamental, closely-related constitutive standards, functions or aims, such as aiming at truth and aiming at realization, then one could argue that this division is the basis of corresponding differences in kinds of reasoning and knowledge. For example, one could argue that the difference between theoretical and practical reasoning consists in whether the reasoning concludes in a *thetic* or *telic* thought, or one could argue that theoretical knowledge is non-accidental agreement of a *thetic* thought with its object, whilst practical knowledge is non-accidental

---

[4] For the relevance of DOF theory to the Humean theory of motivation see Smith (1994); for the relevance of DOF theory to the question of whether virtue could be a form of knowledge see Little (1997); for an argument for the claim that full-blooded intentional actions depend on the presence of attitudes with both DOFs see Velleman (2000). For a small list of other applications of the concept of DOF, see Searle (1979, 1983); Blackburn (1988); Schueler (1991); Aulisio (1995); Lenman (1996); Kriegel (2003). There are many more examples.

[5] For this more modest use of direction of fit as a (potential) constraint on psychological taxonomy, see e.g. Jacobson-Horowitz (2006), Setiya (2007).

agreement of a *telic* thought with its object.[6] According to such a strategy, to explain why there are two (and only two) DOFs would be to explain why there are two (and only two) fundamental kinds of reasoning and/or knowledge. Alternatively, indulging what Anscombe calls our "incorrigibly contemplative" attitude to reasoning and knowledge, a true theory of DOF, coupled with an independent account of necessary conditions on reasoning and knowledge, could explain why there just isn't anything for distinctively practical reasoning or distinctively practical knowledge to be, by explaining why telic thoughts cannot enter into reasoning or cognition in the way required.[7]

Given the widespread use of the concept of DOF, not to mention the promise of clarity about the unity and difference of practical and theoretical thought, reason and knowledge, it's important to work out which theory of DOF (if any) is correct. In this chapter I argue that no theory of DOF is correct. There is no clearly unified genus or determinable of which the two proposed DOFs are species or determinations. Whatever they are, the two proposed DOFs do not deserve treatment under a common heading. So philosophical views that rest on a theory of DOF all rest on a mistake; at very least, they require reformulation.

My argument does not merely make a methodological point about the use of an ill-conceived technical term. The failure of DOF theory represents a deeper failure to appreciate the real point of Anscombe's story. Despite what DOF theorists will tell you, Anscombe herself was not a DOF theorist. Her story of the shopper and the detective was not primarily a way of comparing two fundamental kinds of mental state (i.e. the thetic and telic kinds) in terms of some essential property that they share (i.e. possession of a DOF, of the thetic or telic determination). It was rather a way of comparing a kind of mental state to a kind of *event*. Anscombe's basic insight is that just as there is a

[6] This characterization does not preclude the idea that practical reasoning is reasoning that concludes in an action: see Rödl's talk of thought that *is* a movement in Rödl (2007).

[7] Another way of exploiting DOF to illuminate the unity and difference of practical and theoretical would be to explain how practical knowledge is a species of theoretical knowledge, because telic thoughts are a species of thetic thoughts. Arguably this is the view of Velleman (2000), for whom telic thoughts are a species of belief, differentiated by a special self-referential content and causal role, but justified in the same way that other thetic thoughts are justified.

kind of mental state (belief) that is partially constituted by a standard of correctness, against which a mistake in judgment may be measured, so too is there a kind of event – what I call a *telic event* – that is partially constituted by a standard of correctness, against which a mistake in performance may be measured. If Anscombe is right, then an understanding of telic events is prior to an understanding of telic attitudes. The failure of DOF theory represents a failure to appreciate the merit or possibility of this way of drawing the distinction between practical and theoretical thought, which promises a generality of account that DOF theory cannot match. It is the task of this dissertation to defend Anscombe's basic insight, and to develop it into a novel account of the nature of the mind and the nature of intentional action.

Before we turn to Anscombe and her basic insight, we had better make sure that the very idea of DOF is a hopeless one: this is the task of §§1.2–1.4. In §1.2 I summarize the extant attempts to say what DOF consists in, offer some preliminary criticisms, and give reasons to dismiss descriptive theories of DOF. In §1.3 I describe an internal tension between the logical structure and theoretical goals of an adequate normative theory of DOF. In §§1.4.1-1.4.3 I outline three problems that rely on the internal tension. Together the problems indicate that we should treat whatever interesting differences there are between belief and desire (or intention) directly, without supposing that the most important of these differences are somehow symmetrically reflected determinations of some common determinable (i.e. possession of a DOF). In §1.4.4 I diagnose what goes wrong with DOF theory. The opposition between mind and world that guides DOF theory is a bogus one, and in trying to maintain the opposition, DOF theory misapprehends the nature of the telic.

Having found reason to think that the very idea of DOF is a hopeless one, we may return to Anscombe's basic insight. In §1.5 I argue that Anscombe's view is not a DOF theory and explain her distinction between mistakes in judgment and mistakes in performance. In §1.6 I outline several puzzles that face an account of practical and theoretical thought based on Anscombe's basic insight.

The puzzles motivate a program of inquiry in the metaphysics of mind and philosophy of action that is pursued in subsequent chapters.

1.2 THEORIES OF DOF

DOF theorists, whether normative or descriptive, think that some deep *symmetry* structures the normative or causal relations characteristic of practical and theoretical thought; hence the imagery of two DOFs, pointing in opposite directions. In this section I summarize the extant attempts to cash out this image of symmetry in philosophically respectable terms.

1.2.1 Normative "No Theory" Theories (Searle and Platts)

Searle (1983) has done the most to popularize DOF in contemporary philosophy of mind. Claiming inspiration from Anscombe's story, he offers the following theory of DOF:

> …the idea of direction of fit is that of responsibility for fitting… If my beliefs turn out to be wrong it is my beliefs and not the world which is at fault, as is shown by the fact that I can correct the situation simply by changing my beliefs. It is the responsibility of the belief, so to speak, to match the world … But if I fail to carry out my intentions or if my desires are unfulfilled I cannot in that way correct the situation by simply changing the intention or desire. In these cases it is, so to speak, the fault of the world if it fails to match the intention or the desire …[8]

Searle says that 'DOF' is an irreducible, unanalyzed, primitive term.[9] I call his theory a "no theory" theory because although he makes several suggestive remarks that cleave close to the slogans with which we began, he never tries to explain in any detail what it is for a belief or the world to bear

---

[8] Searle (1983: 7-8). The suggestion that the idea of DOF is Anscombe's is found in Searle (1979: 3-4).

[9] Searle (1983: 173). The lack of explanation in Searle's account is a bit surprising. Searle's (enormous) project in philosophy of mind and philosophy of language is partly founded on his theory of DOF. One expects at least a *little* more explanation of what this foundational primitive amounts to, even if it cannot be analyzed in other terms.

"responsibility for fitting"; what particular kind of norm is at issue; what difficulty or qualification the words 'so to speak' mark; why and to what extent symmetry is present in the very idea of DOF; etc.

Platts (1997) offers a slightly more detailed theory that is very similar to Searle's:

> Miss Anscombe, in her work on intention, has drawn a broad distinction between two *kinds* of mental states, factual belief being the prime exemplar of one kind and desire a prime exemplar of the other… The distinction is in terms of the *direction of fit* of mental states with the world. Beliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in belief, and false beliefs should be discarded; beliefs should be changed to fit the world, and not vice versa. Desires aim at realisation, and their realisation is the world fitting them; the fact that the indicative content of a desire is not realised in the world is not yet a failing in the desire; the world, crudely, should be changed to fit our desires, not vice versa.[10]

Platts also cleaves close to the slogans with which we began, and once again we have a "no theory" theory. Platts offers more claims than Searle, made out in terms of decisive failings, what should be changed, what should or need not be discarded, and so on. But he never tries to explain how the claims are related to each other; what kind of norms are at issue; whether "aiming at truth" and "aiming at realization" are the same kind of "aiming"; etc. (Unlike Searle, Platts has doubts about the value of the distinction he finds in Anscombe, and no doubt this explains the lack of detailed explanation in the account.)[11]

Although they both claim Anscombe as progenitor, the distinction Platts and Searle make is not the one Anscombe made. Anscombe's talk of records is plausibly read as a way of talking about beliefs. But where Anscombe wrote of intentions and orders, Platts and Searle substitute or add in *desires*. Furthermore, although Anscombe is clear enough about mistakes ("decisive failings"), she has nothing to say about whether the world *should* be changed to fit one's desires or intentions, and nothing to say about the world being "at fault" when it fails to match one's desires or intentions. We will return to the question of how to properly understand Anscombe in §1.5. For now we may note

---

[10] Platts (1997: 256-7)
[11] Platts worries that "all desires involve elements of belief" and so are not purely telic. See Platts (1997: 257).

that Platts and Searle depart appreciably from what Anscombe wrote, and that it is not obvious that they have Anscombe right, or even close to right. The point is worth raising now because every DOF theorist follows this misreading of Anscombe.

Platts's account brings out a problem for normative theories that Searle's account obscures. Platts works hard to preserve symmetry of account, swapping "[relevant state of mind]" for "the world" as required, but he balks at the final hurdle. Beliefs should be changed to fit the world, but it is only in a "crude" sense that the world should be changed to fit our desires. Although the nature of the crudity is mysterious, the fact of it is not surprising. There is certainly no moral sense in which the world should be changed to fit someone's desire to murder out of mere curiosity (not even *prima facie*, if it is indeed murder that they have in mind). If the sense of 'should' at issue is that of a reason for action, whereby agents should do the things for which they have reasons, Platts's view may involve the unreasonably strong assumption that all desires, merely by their possession, give their possessors decisive reason to act so as to fulfill the relevant desire. Even a committed subjectivist would want to qualify this claim.[12] This seems like a general problem. Whilst one leaves the nature of the relevant norms in relative obscurity, nothing seems to hinder a thoroughly general symmetry of account. But for any precise statement of the norms involved, it seems that qualification is required on the telic side, where no corresponding qualification seems in order for the sense in which beliefs "ought to fit" the world. I will outline a generalized version of this problem in §1.4.2 below.

1.2.2 The Higher-Order Mental State Theory of DOF (Humberstone)

Humberstone (1992) thinks that a "no theory" theory of DOF is no theory at all. He cites with approval Anscombe's talk of mistakes, but expresses dissatisfaction that the "source of normativity"

[12] See e.g. Sobel (2009) for a brief discussion of subjectivists' attitudes to desires and reasons for action.

remains mysterious in her work, and suggests that an account of DOF grounded in occurrent episodes or states of the thinking subject's psychology will help dispel the mystery. Humberstone says the two DOFs are two different higher-order intentions that individuals adopt towards lower-order attitudes. These higher-order intentions are supposed to be partially constitutive of those lower-order attitudes' being the kinds of attitudes they are. According to the theory, it is constitutive of belief (the paradigmatic thetic attitude) that one intends that one does not believe what's false. It is also constitutive of a broad range of pro-attitudes (encompassing desire, intention, and a whole lot more besides as paradigmatic telic attitudes) that one intends that the objects of those lower-order pro-attitudes obtain.[13]

Humberstone produces some notation to exhibit the logical structure of these two fundamental, constitutive, higher-order intentions:

[Thetic DOF]: *Intend* ($\neg Bp$ / $\neg p$)

[Telic DOF]: *Intend* (p / Wp) [14]

The formulations are conditional in form: one intends to not believe that p, given that not p ($\neg Bp$ / $\neg p$), and one intends that p obtain, given that one wants, or desires, or intends that p, or has some other lower-order pro-attitude towards p obtaining (p / Wp).

We can be alienated from some of our desires. When we are alienated from a desire, it would be false to say that we intend that the desire's object obtains, so Humberstone's stated account of the telic DOF is false, and this casts doubt on his account of DOF in terms of higher-order telic attitudes. Humberstone acknowledges this problem, but his solution is odd: he says that he has used the wrong word, because intentions are obviously too committed for cases of alienation from lower-

[13] Humberstone is rather blasé about what counts as telic: he says that "intentions, desires or *whatever*" are telic attitudes, and later that all kinds of pro-attitudes are telic attitudes. See Humberstone (1992: 75, 81).
[14] Humberstone (1992: 75).

order attitudes.[15] He doesn't tell us what the right word is, so let us call these mysterious higher-order telic attitudes that (magically?) sidestep alienation 'humberstentions'. It is very unclear what humberstentions are supposed to be. They are not intentions. They are not desires or evaluative attitudes: Humberstone himself argues against the definition of the thetic DOF as *Desire* ($\neg Bp$ / $\neg p$) or *Value* ($\neg Bp$ / $\neg p$).[16] Presumably they are not affective attitudes like *Feel-Warm-And-Fuzzy* ($\neg Bp$ / $\neg p$). We are left with the vague idea that they are telic attitudes that set a normative standard, but the attitude is not a kind of telic attitude with which we are familiar (and for which we already have a word), and the standard is not a *committed* standard that would cause trouble for the possibility of alienation. Positing a largely empty theoretical term in order to fix a theory is an occupational hazard for the philosopher; we could be forgiven for thinking that 'humberstention' is a term of this kind.

The worst problem with Humberstone's account is that even if we could make sense of the mysterious humberstention, the account would still be *circular*. Appealing to higher-order telic attitudes, Humberstone attempts to ground the thetic DOF on a prior understanding of the telic DOF, but this depends on a prior understanding of telic attitudes, and it is not clear that the grab-bag of telic attitudes that Humberstone provides, or the mysterious humberstention itself, represents an improvement over the "no-theory" theory.[17] To avoid circularity, Humberstone could drop the individualistic aspect of his theory, so that instead of '*Intend*' he just writes '*Standard*', without presupposing that the constitutive normative standards must be grounded in the content of some occurrent episode or state of the subject's psychology. This would not explain the source of normativity in the only individualistic way that Humberstone seems to think possible, but then his

[15] Humberstone (1992: 81).
[16] Humberstone (1992: 66-68).
[17] Humberstone, honest philosopher that he is, acknowledges this problem with circularity. He does have a defense against *one* understanding of what is wrong with such circular explanations. See Humberstone (1992:67 n13).

own theory doesn't do that, so it looks to be a reasonable trade. We will consider, and find reason to reject, such a normative account in §1.4.

1.2.3 Causal-Dispositional Theories of DOF (Smith and Velleman)

Smith (1994) provides the most widely cited and criticized theory of DOF:

> … the difference between beliefs and desires in terms of direction of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that p and a desire that p on a perception with the content that not p: a belief that p tends to go out of existence in the presence of a perception with the content that not p, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that p. Thus, we might say, attributions of beliefs and desires require that different kinds of counterfactuals are true of the subject to whom they are attributed.[18]

According to Smith, the thetic and telic DOFs are not normative relations, but rather tendencies or dispositions possessed by belief and desire respectively. He rules out the possibility that a mental state could have both DOFs, as nothing could have both dispositions at once.

There are counterexamples to Smith's theory.[19] On the telic side, my desire that the car is red *now* tends to desist in the face of a perception with the content that it is not the case that the car is red now, where Smith's theory says it should tend to persist.[20] (I see that I cannot have what I want.) Of course, I might quickly adopt the desire that the car is red *very soon*, but that is a different desire, just like the belief that it will be red very soon is a different belief. On the thetic side, my belief in your innocence, sustained by a long history of trust, is thoroughly robust in the face of appearances

---

[18] Smith (1994: 115).

[19] Apart from raising similar counterexamples, Sobel and Copp (2001) also object that Smith's account is circular in an objectionable way. According to them, the perception that not *p* must function as a response to evidence if it is to play the role it is supposed to in Smith's theory of DOF. But as a response to evidence, it is something that itself has the thetic DOF, so no non-circular account of the underlying phenomenon has been offered.

[20] Setiya (2007: 49) makes essentially the same point.

to the contrary: my first and abiding impulse is to go looking for reasons to doubt the appearances rather than reasons to doubt you.[21]

Apart from such counterexamples, Smith's theory has a notably lopsided logical structure. DOF theory is guided by intuitions of deep symmetry. The disposition to *A* in *C* and the disposition to not-*A* in *C* exclude each other and so presumably satisfy the requirement of symmetry of account by exhausting the options in some interesting logical space. But there are *three* elements to relate in Smith's theory, not just two. These elements are: (a) the disposition to desist in the face of a perception that not-*p*, (b) the disposition to persist in the face of a perception that not-*p*, and (c) the power to dispose the subject to act (so as to bring it about that *p*). Whereas Smith's thetic DOF is simple, consisting of (a) alone, his telic DOF is complex, consisting of a conjunction of (b) and (c). Even allowing that (a) and (b) are incompatible, and so by implication that the thetic and telic DOFs that Smith describes are incompatible, Smith hasn't told us why there couldn't be a *third* DOF that combines (a) and (c).[22] Mental states that possess this third DOF might be bizarre – they are called 'besires' in the literature – but we should not think that just because Smith's two favored DOFs exclude each other that he has shown that besires are therefore impossible.[23] Indeed, the fact that it doesn't seem to matter to Smith whether a DOF is simple or complex raises the worry that for all he has told us about the genus DOF, one could invent an infinite number of further species of DOF, by coming up with more or less arbitrary additions to (a), (b) and (c) to cover one's favorite dispositional aspects of thought.

---

[21] We may speak of appearances here as Smith notes in correspondence with Humberstone that he does not mean factive perception by the word 'perception', where what one perceives must be true, but rather how things appear to one. See Humberstone (1992: 64 n10).

[22] If his thought is that nothing that was disposed to desist (etc.) could *hang around* long enough to dispose the subject to act, his thought is straightforwardly mistaken. When one has reason to doubt the appearances, a belief may hang around whilst being in the presence of a 'perception' to the contrary, so there is no reason why it couldn't hang around long enough to exert some power over the subject and her behavior.

[23] Zangwill (2008: 52 n1) makes essentially the same point.

Smith himself implies that something like this last suggestion is the right one. In the footnotes Smith recommends giving up on DOF talk altogether in favor of "speaking directly about patterns of dispositions," because talk of DOF is often too "slack" to do whatever work is required.[24] He goes on to say that when he uses the term 'DOF' we should understand him as referring to "whole packages of dispositions constitutive of desiring and believing" rather than his stated theory of DOF. So there is a relatively clear and self-attributed sense in which Smith is not after all a DOF theorist, and has no particular interest in cashing out an image of symmetry in terms of DOF. His theory of DOF, however widely cited and criticized, is really a theory about bundles of dispositions, and there *are* an infinite number of possible such bundles, pointing in an infinite number of directions, not just two pointing in opposite directions.[25]

Are there any other descriptive theories of DOF apart from Smith's? Velleman (2000: 105) seems to present an alternative. Velleman distinguishes two modes in which a propositional attitude might regard its propositional object – regarding-as-true and regarding-as-to-be-made-true – and says that those are the two DOFs. He also says that these modes of regard do not involve any commitment or disposition towards truth or action: beliefs, hypotheses, musings, fantasies etc. all regard their object "as true" (even if not as *really* true); desires, hopes, wishes, fantasies etc. all regard their object "as-to-be-made-true" (even if not as to be *really* made true).[26] This looks like an alternative descriptive theory of DOF, but it isn't. Velleman notes that his use of the term 'DOF' is

---

[24] Smith (1994: 209-210 n8)

[25] Of course Smith's interest is not in any old bundles, but in bundles of dispositions that are constitutive of the mental states relevant to the motivation of action. But for all Smith has shown, there are an infinite number of those, not just two. Certainly the claim that there are only two that are relevant to the motivation of action requires more argument than Smith has given.

[26] Fantasies appear in both lists because Velleman isn't very clear about where they fall. He introduces fantasies as cognitive attitudes that regard their propositional object "as-true" but within a few pages he talks of "fantasy wish-lists" that regard their propositional object "as-to-be-made-true". See Velleman (2000: 110, 120). If one has fantasies that one enacts, or considers enacting, perhaps they have both modes of regard in turn, or at once.

"somewhat different" from that found in the DOF literature.[27] This is an understatement. Modes of regard do not involve anything "fitting" anything else in *any* sense of fitting (normative or otherwise). To this extent, Velleman is talking past the concerns of DOF theorists when he uses the term 'mode of regard' synonymously with the term 'DOF'. We should ignore his stated theory as an unfortunate terminological confusion, however interesting modes of regard might be in their own right.[28]

As it happens, Velleman does have a secret theory of DOF of a more traditional form, whereby two paradigmatic kinds of mental state contrast in how they "fit" their objects, although he doesn't acknowledge it as such. Velleman thinks that belief and choice contrast in their "direction of guidance": beliefs are caused by what they represent, whilst choices cause what they represent.[29] "Direction of guidance" is Velleman's (cheeky? dimly cognizant?) term for his secret descriptive theory of DOF.

Velleman's secret theory seems too crude as stated. We may grant that when all goes well choices cause intentional actions that they also represent. But some mere beliefs also reliably cause what they represent: consider the belief that I am a failure, or the belief that this relationship isn't working. Presumably Velleman doesn't think these beliefs are telic attitudes. On the thetic side, the claim that beliefs are caused by what they represent only seems plausible when restricted to certain kinds of belief. It seems plausible enough to say that when all goes well, the object of a perceptual belief causes the belief and determines its content. But even when all goes well, beliefs about the future are not caused by what they represent: not unless effects precede their causes (a claim that

---

[27] Velleman (2000: 105 n14).
[28] Velleman doesn't need the term 'DOF' to do any work for him, unless he wants to make the objectionable suggestion that all the other DOF theorists unconsciously wanted to write about *his* distinction between modes of regard, but somehow didn't manage to do so. Velleman seems to make this objectionable suggestion in Velleman (2000: 250 n12).
[29] Velleman (2000: 25)

requires some defense). And it seems quite adventurous to say that e.g. when all goes well, beliefs about abstract objects are caused by abstract objects.

Velleman has a less adventurous gloss on what he means when he says that belief is caused by what it represents; he means that beliefs are regulated by "truth-tracking" mental mechanisms that aim to ensure that the believer believes what's true.[30] In some cases, such as perception, these mechanisms relate thought directly to its object, but in other cases, such as beliefs about the future, something else happens (something that isn't precognition). Given the gloss, there is no clear commonality and reflected symmetry that unites the two DOFs of the secret theory. One DOF relates thought to a broad range of regulatory "truth-tracking" mechanisms; the other relates thought to its object. The DOFs point in merely different directions to merely different things, not in *opposite* directions to the same (kinds of) things.[31] It may be for this reason that Velleman, who is perhaps conscious of the similarity between talk of DOF and his own talk of directions of guidance, does not acknowledge what we have called his secret theory of DOF as his own. For if one isn't trying to cash out an image of symmetry in terms of some deep and general logical contrast, one isn't really in the business of offering a theory of DOF, and the constraints on one's account will be quite different.

---

[30] Velleman (2000: 253).

[31] Could Velleman restore symmetry of account by saying that choices also relate thought to regulatory mechanisms, and not (more or less directly) to the object of thought? No. If Velleman takes this option, he has no reason to deny that the mechanisms that incline us to intentional action are just part of "truth-tracking" mechanisms more generally. A mechanism that produces an object of thought seems like just one more way to track the truth for certain actionable contents, the way that perceptual mechanisms are just one more way to track the truth for certain perceptual contents. So there is no special reflected symmetry to be found here, but only mere difference.

1.2.4 Preliminary Conclusions

There is something wonderfully farcical about the preceding literature survey. Apart from Searle, who doesn't really have a theory, Humberstone is the only unabashed DOF theorist in the world, and his account is circular. Platts, Smith and Velleman (implicitly at least) acknowledge in the footnotes that they have their doubts and would like to change the subject. The consensus seems to be that on close examination, the very idea of DOF is slightly embarrassing.[32]

I think it would be a mistake to dismiss the very idea of DOF on the basis of this consensus of embarrassment. The image of symmetry is still with us, and it is a powerful one. There aren't many philosophers who have proposed a third kind of thought (or reasoning, or knowledge) to lay alongside the practical and theoretical kinds with equal dignity. The image of symmetry hints at an explanation of why this is. There are only two ways to order terms in a two-place relation between mind and world: either mind comes first and world second or vice versa. If there is such a schematic relation that delimits the bounds of thought in an interesting way (i.e. DOF *as such*), that would explain why practical and theoretical thought are the only two fundamental kinds in this particular logical space. We ought to at least try to see if we can do justice to this intuition (supposing we have it) before giving up and deciding that we are chasing shadows. So in what follows, I will broaden the focus and investigate whether it is possible for there to be an adequate theory of DOF, regardless of the relative merits of the extant theories.

That said, I propose that we dismiss descriptive theories from further consideration. Smith and Velleman have no particular interest in cashing out the image of symmetry that guides DOF theory: that is one reason why they want change the subject (to bundles of dispositions in Smith's

---

[32] Unfortunately, this consensus of embarrassment doesn't seem to stop non-DOF theorists from employing the concept of DOF in the exposition of their views. If I am right that the very idea of DOF is a flawed one, such philosophers need at very least to reformulate their views in other terms.

case, and to modes of regard in Velleman's case). More generally, it is very difficult to see what advantage or deep insight a symmetrical descriptive account of DOF could afford us. The central explanatory concept available to the descriptive theorist is that of the causal disposition. But the great advantage of the causal disposition is its flexibility. Mental states seem to be caught up in *all kinds* of causal dispositions that point in *all kinds* of different directions, and not just in two that point in opposite directions. We may grant that some mental states (e.g. intentions) tend to be caught up in causal relations where they contribute to an intentional action that matches their content in some special way. We may also grant that some other mental states (e.g. beliefs) can be caught up in perceptual causal relations, where the belief is more or less directly caused by what is the case, and its content determined by what is perceptually available to the perceiver. But this opposition doesn't mark any *deep* symmetry. As noted above in response to Velleman: there are some beliefs that have a (reliable) role in causing what they represent without doing so via intentional action, and there are many beliefs for which the invocation of a direct causal relation to the object of belief as cause of the belief is strained at best. To try and force the relevant causal relations into the mold of *deep* symmetry for two particular paradigmatic kinds of mental state (the thetic and telic kinds) seems pointless, as it puts a straitjacket on the very flexibility that is supposed to be one of the great advantages of theorizing about the structure of the mind by reference to causal dispositions.

This is not to say that there cannot be a descriptive account of DOF that cashes out an image of deep symmetry in an illuminating way. But I have no idea what such an account would look like, or what its interest would be. It doesn't seem like an accident that the actual theorists who tried for a descriptive theory of DOF ended up changing the subject. So rather than trying to construct an ill-motivated straw-man descriptive account of DOF to criticize, I shall proceed on the assumption that if there is an adequate theory of DOF it will be a normative one.

1.3 THE LOGICAL STRUCTURE AND THEORETICAL GOALS OF AN ADEQUATE THEORY OF DOF

What would make for an adequate theory of DOF? On the one hand the slogans with which we began dictate a particular logical structure for an adequate theory of DOF. On the other hand, 'DOF' is a technical term, so we will not get a good grasp on what would make for an adequate theory of DOF in advance of understanding what DOF is *for*, theoretically speaking. Let us deal with each point in turn.

The opening slogans present an overwhelming impression of deep (geometric) symmetry at work in the very idea of DOF. The point is best shown, then said:

Thetic DOF: Mind ⟵ World

Telic DOF: Mind ⟶ World

The DOF theorist must explain what the arrows represent, but the arrows must represent the same kind of relation; that is why they deserve treatment together under a common heading. Here instead of 'Mind' and 'World' we might put 'aspect of Mind' and 'aspect of World', but the aspects must be the formally the same across the thetic and telic cases: suitable candidates are e.g. a mental state and what the mental state refers to, or a propositional attitude and the propositional object of the attitude. If the aspects weren't formally the same for both DOFs, we would have good reason to think that the arrows are not symmetrically reflected versions of the same kind of relation, precisely because of the formal difference in the relevant *relata*.

What logical structure of account would capture both the intended unity and reflected (geometric) symmetry of the two DOFs? A two-dimensional spatial direction can be represented without a diagram by means of an ordered pair of coordinates. So represented, spatial directions that

point in opposite directions differ only in the ordering of the coordinates they share. Similarly, we can expect an adequate theory of DOF to have the following schematic logical structure:

Thetic DOF: *Ought-to-Fit* < (aspect of) Mind, (aspect of) World>

Telic DOF: *Ought-to-Fit* < (aspect of) World, (aspect of) Mind>

Here '*Ought-to-Fit*' is a schematic term for the normative relation that captures the unity of practical and theoretical thought in terms of DOF. We can see from the schema why there are two and only two DOFs.[33] DOF as such is supposed to be a determinable normative relation with two terms, one standing for (some aspect of) the mind and the other standing for (some aspect of) the world, where the determinations of the determinable differ only in the ordering of favored terms. There are only two ways to order terms in a two-term relation, and that is why there are two and only two DOFs. An adequate theory of DOF must reveal DOF as such to be a determinable relation of this kind.[34] Call this the *basic criterion of adequacy* for a theory of DOF.

It is because of the basic criterion of adequacy that no one has ever tried to mix a normative and descriptive account of DOF, so that e.g. the thetic DOF is a matter of truth being a norm for belief whilst the telic DOF is a matter of desires conspiring (with beliefs etc.) to cause intentional actions. Arbitrarily mixing and matching normative and causal relations between mind and world won't help to display the unity of practical and theoretical thought. If DOF is to live up to its

---

[33] No theorist explicitly claims that there must be two and only DOFs. But it has never occurred to any theorist to deal with some third, fourth, *n*th DOF, nor has occurred to one to mention the possibility of there being some third, fourth, *n*th DOF. The only exception is Zangwill (1998) who treats the idea as a replacement for traditional DOF theories (that is: to propose some third, fourth, *n*th DOF would be to change the subject).

[34] Couldn't an account of DOF just point out some differences between belief and desire? No. Suppose the theorist says "Beliefs *X* and desires *Y*; the thetic DOF consists in *X*-ing and the telic DOF consists in *Y*-ing." We have no clue here as to why *X*-ing and *Y*-ing deserve treatment under a common heading, or how they cash out the image of symmetry that justifies talk of DOFs. Couldn't an account of DOF just point out a property that beliefs have that desires do not have, as Zangwill (1998: 177) suggests that Stalnaker (1987: 80) does? (Arguably, this is what Smith tried to do.) Again, no. We do not say that an object that is not male is therefore of the female gender; similarly, we do not say that a mental state that fails to have the thetic DOF therefore has the telic DOF.

promise (and its name) then the two DOFs must be shown to be the same kind of normative relation between mind and world, although "pointing" in opposite directions.

One might object that the basic criterion of adequacy is too strong. The basic criterion of adequacy requires the two DOFs to differ *only* in the ordering of favored terms in one and the same schematic determinable relation. But surely it's permissible to add qualifications on one side of the account and not the other, so long as the core '*Ought-to-Fit*' relation is recognizably the same in both cases. Platts, for example, qualified the telic side of his account with the term "crudely", the idea presumably being that not every telic attitude ought to be realized, in the relevant sense of 'ought'. What's wrong with a mildly lopsided account like Platts's?

Such qualifications confuse our understanding of the underlying genus or determinable. Suppose, for example, that the thetic DOF is a matter of a mental state actually standing in some determination of the '*Ought-to-Fit*' relation, whilst the telic DOF rather consists in a mental state's *liability* to stand in some determination of the '*Ought-to-Fit*' relation in certain circumstances (and not others). To say that these are both DOFs in the same sense of DOF as such would be like saying that being male and being liable to be female are both determinations of gender. The DOF theorist claims that there is a clearly unified genus or determinable – DOF as such – of which there are two and only two species or determinations. Either the genus or determinable is of the form of a relation, or it is of the form of a liability to stand in a relation, but the DOF theorist cannot have it both ways. As it happens, DOF theorists agree that every belief has the thetic DOF; beliefs are not merely liable to stand in a normative relation to the world, but always do so. We may conclude that the telic DOF is also a normative relation, and not a mere liability to stand in a normative relation in

certain circumstances. The basic criterion of adequacy stands, and lopsided accounts like Platts's are ruled out.[35]

We can work out what the concept of DOF is for by examining DOF theorists' judgments as to which attitudes are thetic or telic. On the thetic side, there is resounding agreement: "Belief, belief, and again we say belief" say the various theorists.[36] But when we turn to the telic side, there is widespread disagreement. Anscombe (if she has a theory of DOF) says that intention is the paradigmatic telic attitude; Platts says desire is; Searle says desire and intention are; Smith and Humberstone say a wide range of pro-attitudes are; Velleman's secret theory says that choice is (and Velleman thinks that choices are a species of *belief*).[37] What explains the agreement on the thetic side and the disagreement on the telic side?

On the thetic side, the agreement can be explained by a common interest in what is usually called the 'aim of belief'. Belief seems to have a kind of internal axiology – "aiming at truth" – which merits investigation and explanation.[38] On the telic side, one's preferred philosophy of action makes all the difference to what one says. Searle, Smith and Velleman go on to use their reflections on DOF in the service of illuminating the nature of intentional action, according to their preferred theories of it, and Platts and Humberstone's choices of telic attitude are plausibly read as a gesture towards the Humean and Davidsonian traditions of action theory. None of the theorists acknowledges an interest in a broader theory of DOF according to which, say, *merely* affective or evaluative attitudes (i.e. ones that do *not* have an essential connection to intentional action) end up classified as telic. So let us proceed on the assumption that illuminating the internal axiology of belief ("aiming at truth") is one theoretical goal of the DOF theorist, and illuminating the internal

[35] The basic criterion of adequacy does not preclude there being many asymmetries that *follow* from the account of DOF. It just limits the asymmetries admissible in the core account to an asymmetry in the ordering of terms in one and the same determinable relation.
[36] Here we are ignoring Velleman's theory of modes of regard, as it is not a theory of DOF as such.
[37] Velleman (2000: 25-26).
[38] For interpretations of the "aim of belief" see e.g. Velleman (2000); Wedgwood (2002); Shah (2003).

axiology of telic attitudes, in terms of their essential role in *intentional action* ("aiming at realization"), is the other theoretical goal of the DOF theorist.

1.4 PROBLEMS WITH ASYMMETRY OF ACCOUNT

There is an internal tension between the logical structure and theoretical goals of an adequate theory of DOF. The basic criterion of adequacy pulls towards symmetry of account, whilst the theoretical goals pull towards asymmetry of account. In this section I will outline three problems for normative theories of DOF that rely on this internal tension, and which together give reason to think that the very idea of DOF is a hopeless one.

1.4.1 Problem One: Kind of Normativity

The first problem for any normative theory of DOF is in specifying what kind of norm or normativity is at issue. Different kinds of norm have very different properties. If the proposed account of DOF is to meet the basic criterion of adequacy then the kind of norm or normativity must be made out to be of the same kind in the thetic and telic cases. It won't do, for instance, to say that one DOF is a matter of individualistic-evaluative normativity, and the other is a matter of biological-teleological normativity, unless one has a *general* theory of normativity that unites these two (a tall order). This simplifies our methodology. If a kind of norm doesn't work on the thetic or telic side then we have reason to discard it. Let us run through the options.

We cannot appeal to teleological norms, for that is what we started with in the "metaphorical" slogans. We want an *account* of what it means to say that beliefs 'aim at truth' and

telic attitudes 'aim at realization' that reveals these 'aims' as somehow the same kind of relation to the world, although of symmetrically reflected determinations.

We saw in Platt's case that the telic norm cannot be a *moral* norm, for some of our desires are morally reprehensible no matter the circumstances. In any case that would make the thetic norm an implausibly moralistic proposal concerning the ethics of belief, reminiscent of Clifford's talk of the "sins against Mankind" involved in believing without evidence.[39]

The thetic norm cannot be a non-moral evaluative norm consequent on an individual's or group's (actual or ideal) evaluation, for as Humberstone argues, it might be wonderful (for oneself, or everyone) to believe something false, and this would not mean that the false belief is not *mistaken* in the way that an adequate theory of DOF tries to explain.[40]

For similar reasons we cannot appeal to a teleological norm in the form of a biological imperative drawn from theories of natural selection: it might be wonderful for survival to believe something false (on some, or many occasions) and it might be wonderful for survival to not realize some of one's desires or intentions (on some, or many occasions).[41]

Perhaps we could appeal to the normativity characteristic of *reasons*. Suppose we define thetic attitudes as those to which reasons for belief apply, and telic attitudes as those that supply reasons for action, on the assumption that one *should* believe what one has reason to believe and do what one has reason to do. This would risk endorsing the suggestion that is (perhaps) implicit in Platts's account: namely that there is a kind of attitude (e.g. desire) that by its very nature supplies reasons for action *whenever* one has an attitude of that kind. The suggestion seems too strong: it needs qualification on the telic side, which makes for an asymmetry between the thetic and telic DOFs that flouts the basic criterion of adequacy. Apart from this asymmetry, this 'reasons-based' approach to

---

[39] Clifford (1999).
[40] Humberstone (1992: 68).
[41] I will discuss the reasons to reject this view of the relevant kind of normativity in more detail in Chapter 2.

the normativity characteristic of DOF seems circular. The thetic DOF is defined in terms of reasons for thetic attitudes, for instance. If we have an independent grasp on the unity and difference of reasons for action and reasons for belief, we won't need a DOF theory to illuminate the unity and difference between them. If we don't, then adding the word 'reason' to the account won't help to relieve our ignorance.

Perhaps we could appeal to norms of *rationality*. According to this suggestion, beliefs ought to fit the world because having true beliefs constitutes, or at least greatly promotes, (maximal) rationality on the part of the believer, and desires or intentions ought to be fulfilled for the same reason. But again, circularity threatens. Two kinds of rationality must be made out if we are to make sense of the unity and difference of the two DOFs. If we have an independent grasp on the unity and difference of practical and theoretical rationality, we won't need a DOF theory to illuminate the unity and difference between them. If we don't, then adding the word 'rationality' to the account won't help to relieve our ignorance.

We are fast running out of options. Having found reason to discard bare teleological norms, moral norms, non-moral evaluative norms, evolutionary-teleological norms, norms consequent on possession of a reason, and norms of rationality, all that seems left is something akin to Anscombe's own suggestion: the kind of normativity involved is that of a bare *standard of correctness*. According to this suggestion, considering a mental state and its object, if the mental state has the thetic DOF, then the mental state is correct iff it matches its object, and if the mental state has the telic DOF, then the object is correct iff it matches the mental state, and either of these (mental state or object, respectively) is incorrect otherwise.

We might wonder, with Rosen, whether standards of correctness are norms at all.[42] The fact that there is a correct way to play the *Hammerklavier* sonata (where this standard allows of many artistic interpretations, but does not, for instance, allow rearranging the notes) does not by itself seem to have any consequences for whether I ought to play it correctly here and now, or what I ought to do if I happen to play it incorrectly. Sometimes I may have reason to play it incorrectly: Rosen's example is playing incorrectly to amuse someone. Certainly, having botched the opening, there is no obvious requirement to immediately start over. This makes it seem as if standards of correctness are merely descriptive features of the world. Call the difficulty of making plausible the claim that standards of correctness are genuine norms "Rosen's challenge".

The correct thing to say in response to Rosen's challenge is, I think, that standards of correctness are a kind of limiting case of normativity. They are extremely bare with regard to their intrinsic connection to reasons for action or robust norms of conduct.[43] But they do articulate something more than a mere description: there is some difference between false sentences and false beliefs that is expressed by the claim that false beliefs are (in themselves) incorrect, whilst false sentences are neither (in themselves) correct nor incorrect. Rosen's example itself seems to involve a standard of correctness of this bare type – namely doing what's required to succeed at amusing one's audience – even though this requires flouting whatever standard of correctness is embodied in the score. The standard of correctness implicit in the performance is *internal* to it in some way: it might well conflict with what one ought to do when all things are considered (e.g. perhaps one ought not to take *these* means to the end, given the lack of respect they show to the great composer). But there

---

[42] Rosen (2001).

[43] Rosen himself is in two minds about whether to call standards of correctness norms. Although they do not have any obvious connection to what one ought to do (where what one ought to do is a paradigmatic case of normativity), they do represent what he calls "operative standards", which, like the laws of etiquette, can provide one with reasons for action, although on a contingent basis.

is no obvious reason to think that this internal standard of correctness must therefore be a merely descriptive feature of one's will.

Rosen's challenge would perhaps be difficult to meet if we had to assimilate all forms of normativity to norms of conduct. Although our clearest understanding of norms (arguably) comes from norms of conduct, there is no obvious reason why a norm must have direct consequences for what one ought to do, where 'do' carries the sense of intentional action, if it is to be called a norm at all. The term 'norm' is somewhat fungible. That said, I do think there is something right about Rosen's challenge, and we will return to this point in §1.4.4. For now it is enough to note that if a normative theorist of DOF is to meet the basic criterion of adequacy, then the most plausible candidate for the kind of normativity involved is that of a standard of correctness. If Rosen's challenge is a good one, and standards of correctness are not norms, then hopes for an adequate normative theory of DOF are dim indeed.

1.4.2 Problem Two: Asymmetry of Application

The second problem for normative theories of DOF is that the proposed thetic and telic DOFs will fail to have the same *universality* of application. This gives us reason to think that they are not reflected determinations of one and the same determinable normative relation.

Humberstone notes that the DOF theorist's interest on the thetic side is in the sense in which *every* belief "ought to fit" the world.[44] There may be instrumental reasons to have a false belief, but the DOF theorist has no particular interest in those.[45] There may also be evidential reasons to

---

[44] The thetic norm could be stronger: perhaps every belief ought to be a case of *knowledge*. Nothing in the present chapter turns on this difference, so I will deal with the weaker case of truth, which is a standard for belief to which all the actual normative DOF theorists agree. In Chapter 2 I argue that if mere beliefs are imperfect exercises of the fallible capacity for knowledge, then the standard of correctness for belief is knowledge, or something very much like it.
[45] Humberstone (1992: 68)

adopt a false belief, as when a false conclusion is what all the available evidence happens to point to, but the DOF theorist has no particular interest in such evidential reasons either. Her particular interest is in the normative sense in which all beliefs "ought to fit" the world (i.e. be true) regardless of how this norm figures in more comprehensive calculations of what one ought to believe (in some more comprehensive sense of 'ought').

All thetic attitudes ought to fit the world, in the relevant sense of 'ought', but the telic norm does not have the same universality of application. Anscombe's example of this is the man who wants to buy shark tackle in Oxford – no matter what he does, he won't buy shark tackle, because there aren't any shark tackle shops.[46] Truly predicating an actual normative failing (or success) of some object requires that the object exist. According to Anscombe, the shark tackle man's intention has a kind of action as its object. But although he does various other things, in such a case the man never even *starts* to buy shark tackle. His "action" of buying shark tackle is akin to the person who intends to walk to the edge of the world and sets off in some (any) direction: such a person is not walking to the edge of the world, she merely thinks she is. Without delving into theories of reference, we might call this a failure of reference. As far as the shark-tackle man's attitudes are concerned, it makes no sense to say that whatever is represented by the attitude is as it ought to be or not, because whatever is represented doesn't exist.[47]

The problem here is just the ordinary fallibility of our states of mind. The world makes sense and does not tolerate falsehood or contradiction, so given some intentional mental state with the thetic DOF and a corresponding truth or falsehood regarding the object of that mental state, the thetic norm always applies and it always makes sense to apply it. Our psychology, on the other hand,

---

[46] This isn't true for all times: http://replay.web.archive.org/20060518011854/http://bobstackleandbait.co.uk/

[47] Loose talk helps to cover this problem over. DOF theorists sometimes say that "the world" is assessed as having a failing in it iff it does not match the content of my desire, yet if one tries to narrow in on *that part* of the world that ought to match the determinate content of an impossible-to-fulfill desire, it is nowhere to be found. The problem can perhaps be sidestepped by countenancing the existence of states of affairs defined by means of negated propositions, but no DOF theorist supposes that their view commits them to such wildly adventurous ontology.

is notoriously fallible: we can desire or intend to do things that are impossible, and we can have various telic attitudes that conflict with each other, producing an impossible goal by implication. Nothing guarantees that the objects of our mental states exist, or can exist, so the telic norm does not always apply, and it does not always make sense to apply it.[48] Given this asymmetry, we have reason to think that the basic criterion of adequacy can't be met, because the proposed DOFs differ in more ways than the ordering of favored terms in one and the same determinable normative relation. Our psychology is just too feeble to fund a symmetrical telic norm that measures up to the Almighty Norm of Truth.

Could one restore the required symmetry of account by restricting the telic DOF to mental states that have objects that exist? The idea here would be to compensate for the asymmetry of application by drastically narrowing the range of telic attitudes. Only desires or intentions whose objects exist, for instance, would have the telic DOF and be telic attitudes; many desires and intentions (like the shark-tackle man's intention) would fail to be telic attitudes. Whether this suggestion is workable or not seems to depend on the kind of object at issue. If the objects of telic attitudes are e.g. states of affairs individuated by the content of the relevant telic attitude, then the telic norm would always be trivially satisfied on this view. In order for the telic norm to not be trivial, it would have to be possible for the object of a telic attitude to exist without matching the content of the telic attitude, so that it can sometimes be judged as not being as it ought to be relative to that telic attitude. The *attempt* or *performance* seems like a good candidate object to play this role, given the DOF theorist's interest in intentional action. Suppose that actual existing attempts or performances are the objects of telic attitudes. Telic attitudes would then set the determinate content of the normative standard for actual attempts or performances, and many desires and intentions

[48] I take it that a kind of failure of reference is what Anscombe is gesturing towards when she talks of the order to clench one's teeth "falling to the ground" when it is given to the man with no teeth. See Anscombe (1957: 56-57).

would fail to be telic attitudes, because the possessor of the attitude is not actually engaged in an attempt or performance.

This response to the problem of asymmetry of application is extremely distant from traditional DOF theory. The deep symmetry we expected to see between two paradigmatic kinds of mental state has now been transferred to a deep commonality shared by thetic attitudes and performances: namely the way that they can both be as they ought to be, or not. An understanding of performances now seems importantly prior to an understanding of telic attitudes, as whether or not a mental state possesses the telic DOF depends on the existence of an actual performance. The view is in fact Anscombe's, and I do not think it is a DOF theory at all. We will consider the view in §1.5.

1.4.3 Problem Three: Asymmetry of Form or Content of Thetic and Telic Attitudes

The third problem for normative theories of DOF is that even if the asymmetry of application is superable, thetic and telic attitudes *themselves* display asymmetries that cause trouble for the basic criterion of adequacy. These asymmetries make for a kind of dialectical progression, where what is distinctive about telic attitudes is progressively weakened under pressure from the basic criterion of adequacy, until what is left is useless for the theoretical goals of the DOF theorist. Let us walk through the dialectical stages.

We begin with the thought noted in §1.3: that one theoretical goal of the DOF theorist is to illuminate the essential role of telic attitudes in intentional action. The theorist must then decide on the logical form of telic attitudes as well as deciding which attitudes (e.g. desires, intentions, whatever) are telic. Anscombe is straightforward in this regard. She takes intentions to be essentially involved in intentional action and takes intentions to have kinds of action as their objects. For

Anscombe, the basic logical form of intention is an intention *to A*. It would a mistake to render this thought in terms of an intention that-*p*, for kinds of action are not propositions.[49] If the DOF theorist follows Anscombe in this, then the theory will not meet the basic criterion of adequacy. A formal difference in *relata* is good reason to think that the thetic and telic DOFs are not reflected determinations of one and the same determinable relation.

A (philosophically) natural response is to make sure that the logical form of telic attitudes is that of a *propositional attitude*. There will be no intentions to *A*, but only telic attitudes that-*p*, and in this regard telic attitudes will be just like thetic attitudes. This is the route taken by all the DOF theorists in §1.2. The theorist then faces a difficult choice. In principle, beliefs seem to be able to range over any content that-*p*. The theorist must then choose whether to say that telic attitudes similarly range over any content, or whether to say that their contents are essentially restricted in some way.

There are good theoretical reasons to restrict the contents proper to telic attitudes. Intentional action involves distinctive features that require an account. For instance, as Anscombe (and others) have noted, when one acts intentionally one seems to know without observation or inference what one is doing.[50] It is also a traditional view that practical thought involves representing oneself as cause of the object of practical thought. Given that telic attitudes are supposed to be essentially involved in intentional action, and that the account of DOF is supposed to illuminate this feature of the telic, it seems quite natural to say that a telic attitude that-*p* must include reference to bearer of the attitude (e.g. "I") or must include reference to a kind of action (e.g. "that I do *A*") or otherwise include some content that represents the bearer of the attitude as cause of what is to happen, so as to help explain how telic attitudes are essentially involved in these distinctive features

[49] See also Baier (1970, 1977) on this point.
[50] Anscombe (1957: 11).

of intentional action. Of the theorists in §1.2, only Velleman explicitly endorses this route (and only in his secret theory): he says that telic attitudes must be self-referential, and suggests that they always include reference to what one is going to do.[51] The problem is that if the theorist claims that the contents of telic attitudes are *essentially* restricted in some way, then the theory will not meet the basic criterion of adequacy. An essential difference in range of possible contents provides good reason to think that the thetic and telic DOFs are not reflected determinations of one and the same determinable relation.

A further (philosophically) natural response is to drop the claim that it is essential to telic attitudes to have any essential connection to intentional action, by virtue of form or content. Some philosophers say that God can intend that 2+2=4, which makes it seem as if it is not essential to intentions to refer to the subject or a kind of action: why should telic attitudes be any different? It may then be allowed that for *us* telic attitudes have (or tend to have) contents that refer to the subject or a kind of action, but this would not follow from the nature of the telic DOF. In formal terms, this move could be accomplished by containing the account of DOF within the scope of the consequent of a conditional. The account of DOF would begin with a phrase such as "For all propositional attitudes *A* and all propositions *p*, *if* the subject bears *A* to *p* then …" The fact that the antecedent of the conditional is only fulfilled by creatures like us for some contents on the telic side (such as those that refer to the subject or a kind of action) will then be a contingent fact, whose explanation will be found in some other aspect of the case, such as what we tend to think about when we think a telic thought.

The problem with this final stage of the dialectic is precisely that it gives up on one of the theoretical goals of the DOF theorist identified in §1.3. According to the suggestion, telic attitudes can (in principle) range over any content. If so, then there is no reason why even we feeble creatures

[51] Velleman (2007: 109).

cannot have a telic attitude with the content that e.g. 2+2=4. Such an attitude might not have anything to do with intentional action, but that is now beside the point. (Recall that we are dealing with *normative* theories of DOF: appeal to the causal role of telic attitudes in intentional action is irrelevant to their classification as telic.)

The DOF theorist may refine her conception of telic attitudes so as to satisfy the basic criterion of adequacy, but only at the cost of jettisoning what is distinctive about the telic – namely some essential connection to intentional action, by virtue of form or content – and so jettisoning one of the theoretical goals of DOF theory. And if the DOF theorist gives up on illuminating the essential connection of telic attitudes to intentional action, she has no reason to bother with an account of DOF. She may as well change the subject.

1.4.4 Where DOF Theory Goes Wrong

The problems above provide good reason to think that the very idea of DOF is a hopeless one. We should not try to make out the Almighty Thetic Norm of Truth as of the same kind (though a reflected determination) as the feeble telic norm of satisfaction. We would do better, and learn more, were we to leave the intuitions of deep, exhaustive symmetry between the two DOFs behind, and attend directly to our subject matter: namely, practical and theoretical thought.

Where does DOF theory go wrong? I think the answer is found in the slogans with which we began. There doesn't seem to be anything *prima facie* odd in saying that beliefs ought to fit the world, but there is something *prima facie* odd about saying that the world ought to fit one's telic attitudes just because one has them: namely that this is a markedly *petulant* way of talking. Suppose that I want (intend, etc.) to fit into last year's jeans, but sadly this is no longer possible, and I exclaim "These jeans ought to fit me, but they don't!" Let us suppose further that I make clear that I intend

to pass normative judgment on the jeans by saying this, rather than merely to describe them and to express my disappointment about, or aesthetic reaction to, this state of affairs. What would make sense of such a judgment as legitimate, and not petulant? We do not suppose the jeans were *designed* to fit my end, however large it may get, so we cannot make sense of the judgment as one of malfunction, deterioration or poor design or construction. Yet the only other way to make sense of the judgment seems to be to attribute to the jeans an underlying *capacity* to potentially fulfill the relevant normative requirement. They would have to be made of magical animate lycra, and have a capacity to change themselves to fit my (enlarged) end, not to mention a capacity to be sensitive to when this is required, if we were to treat them as a normative subject, of whom we could legitimately make this kind of normative demand. This is why my judgment is petulant. The judgment involves treating the jeans as if they were the kind of thing that could do and ought to do (or could have done and ought to have done) something about the relevant failure, based on some actual or potential cognition of the (imminent) failure. The jeans are not that kind of thing.

What we have exposed in this example is a mild form of an "ought implies can" principle. The norms characteristic of DOF must conform to this principle, on pain of petulance on the telic side. The sense of 'can' here does not simply mean possible development from one state into another state by some definite time in the future, for it is surely possible that the jeans could have had a history whereby they ended up fitting my enlarged end, just as the world could have had a history that ended up fitting most of my desires. The sense of 'can' at issue is one that attaches to a *normative subject* and is distinctively cognitive: it applies to something that can, in some sense, be sensitive to the requirements of the norm and conform itself to them.[52]

---

[52] It may well be impossible for a normative subject to recognize and fulfill the relevant normative demand in some particular case. But it would be thoroughly ridiculous to apply this kind of norm to some subject who *in general* lacks the capacity to recognize and fulfill the kind of demand in question. I leave open the analogous possibility that, for instance, a moral subject might be subject to various moral demands by virtue of their general capacity to recognize moral

If the norms characteristic of DOF are not petulant, then in order to make this clear, on the telic side the DOF theorist would do well to follow Anscombe and require telic attitudes to represent *action*, because states of affairs and such-like do not have the required capacities. That said, actions don't do anything themselves, so they can't be the subject of the normative demand either. The subject of whom the demand is made on the telic side is *the agent*, who fulfils the demand in her action, by virtue of her capacity to act. Similarly on the thetic side: although we said above that there is nothing *prima facie* odd about saying that beliefs (plural) ought to fit the world, there is something *prima facie* odd about saying that *this* false belief ought to fit the world. Beliefs are individuated by their content: *this* belief can't change to meet the requirements of the norm, but can only be discarded. Mental states do not possess capacities to change themselves so as to fit the world. So the normative subject of whom the demand is made on the thetic side is ultimately *the believer*, who fulfils the demand in her belief concerning whether *p*, by virtue of her capacity to be sensitive to the truth and adjust her beliefs accordingly.

Once we have identified the proper subject of the normative demands, we see immediately that the opposition between mind and world, found in the opening slogans in §1.1 and the diagrams in §1.3, is a bogus one. On the thetic side it makes some sense to talk about the mind on one side (as representing) and the world on the other (as represented), because the failures to meet the requirements of the norm are all tallied up on the side of a normative subject with a certain capacity (the capacity to be sensitive to the truth). But on the telic side, the opposition between mind and world breaks down. The part of the world that is to be assessed on the telic side is not some non-mental chunk of the world (nor some mental chunk treated 'as other', or as external to the subject) that may or may not have the capacity to be sensitive to and meet the requirements of the telic

---

requirements, without thereby recognizing all the requirements (all at once), or being able to fulfill them all in some morally nasty situation.

norm: that would risk petulance. The part of the world that is to be assessed on the telic side is *the mind in action*, because it is the thinking subject who has the requisite capacities. In no case is the world considered as external to the mind the object of the telic norm. This then is the deeper diagnosis of why the very idea of DOF is a hopeless one: in trying to treat intentional action as something *external* to the mind, so as to line up symmetrically with the sense in which the objects of belief are external to the mind, DOF theories evict half their subject matter from their theories.

The sense in which ought implies can for the thetic and telic norms explains how Rosen's challenge from §1.4.1 is on the right track. The relevant kind of norms, if they are not to be petulant, must have some direct connection to *activity* or *action*. The sense of "action" here need not be that of intentional action. A broader sense of activity or action is available and appropriate: namely that of the exercise of a capacity to be sensitive to and meet the requirements of a norm. (I will legitimate this broader sense of 'capacity' in Chapter 2.)

1.5 ANSCOMBE'S BASIC INSIGHT

1.5.1 Anscombe Was Not a DOF Theorist

One of the main attractions of DOF theory is its promise to tell us something about the essence of thetic and telic attitudes. Given the asymmetries outlined in §§1.4.1-1.4.3 we should not expect DOF to play the same kind of constitutive role in the thetic and telic arenas. While it may make sense to *define* beliefs as those mental states that are correct or as they ought to be iff they are true, nothing nearly as strong and helpful to our understanding will be forthcoming with regard to telic attitudes, because of the qualifications required on the telic side.

It is worth returning to Anscombe at this point, because she is entirely explicit about the relevant asymmetries, and accepts them. This gives us good reason to think that Anscombe was not, after all, a DOF theorist. Anscombe says of the shark-tackle man that we would not say he had made a mistake in performance, precisely because there are no shark tackle shops in Oxford. She accepts the asymmetry (of application of the relevant norm) because she was not trying to spell out an essential difference between belief and intention by appeal to the general application a third kind of thing – a DOF – which those two states of mind have in *common*, although in symmetrically reflected determinations. Rather she was simply trying to point out an essential difference between belief and intention.

If one looks for something like symmetry of constitutive roles in Anscombe, one will find it in a different place. Anscombe's basic insight is that what beliefs have in common with another thing is a standard of correctness against which a distinctive kind of mistake may be measured. Beliefs do not have this in common with another kind of state of mind, but rather with an *event*: namely a performance. If one is tempted to define beliefs in terms of a constitutive normative standard, then the lesson to draw from Anscombe is that one might define some *events* in a similar way. Those events would be actions that would be intentional actions were they to succeed, and the constitutive standard would be success (doing what one has in mind to do). Let us call these events *telic events*. Starting with an account of telic events, one could then work backwards to define telic attitudes: telic attitudes are those attitudes that set the determinate content of the standard of correctness for *actual performances that are occurring* (or have occurred, or will occur) on particular occasions. In the light of this reading of Anscombe, there is good reason to think that desires cannot play the role that telic attitudes are supposed to perform. Desires do not set standards of correctness for performances directly. There is no mistake, for instance, if I fail to fulfill my grisly and unwelcome desire to murder out of curiosity (it's not like getting stage fright). It is only if I decide,

or otherwise intend, to fulfill the grisly desire that a mistake might be in the offing. The DOF theorists were, for this reason, wrong to follow Platts in supposing that Anscombe was writing about the difference between belief and desire.[53]

One might object: "Didn't Anscombe say that intentions and orders have one relation to what they represent that records (beliefs) do not share, and didn't she say that records (beliefs) have another relation to what they represent that intentions and orders do not share, and didn't she imply that the relations have *something* in common?" Yes, she said and implied all that. But she never said that these relations cash out an image of symmetry. The image of symmetry was an unfortunate imposition on her view that (I believe) first occurred in Searle's reading of Anscombe, and was then picked up by Platts, and through Platts, the rest of the DOF theorists. Mind and world do seem to exhaust the options for *relata* in some deep logical space, even if that logical space is hard to specify. But in the surrounding passages, Anscombe is very clear that *her* opposition is between judgment and performance; she never puts the opposition in terms of mind and world. It is far from obvious that judgments and performances *exhaust* the options for fundamental and distinctive kinds of mistake (thought, reason, knowledge etc.), and Anscombe never claimed that they do, or that they do by virtue of some kind of deep reflected symmetry.

The possibility that an understanding of telic events might be prior to our understanding of telic attitudes explains why Anscombe herself (uncommonly) thinks we should start with what someone actually does (the telic event as such), rather than with some mental something-or-other, when examining the nature of practical thought.[54] Yet no latter-day DOF theorist takes seriously the idea that an event might be constituted as the kind of event it is by the application of a norm to it. Instead they focus on the idea that while one kind of attitude (belief) is constituted by being *subject* to

---

[53] Of course, intentions do not set standards of correctness *just* by themselves either: it is only if there is a corresponding telic event (past, future or present) that they do.
[54] Anscombe (1957: 9).

the authority of the relevant state of affairs, another kind of attitude (the telic attitude) is constituted by its authority to *set* a norm for a state of affairs, or its *ability* to generate a state of affairs that the agent (in some sense) likes. The task then is to account for the authority or to describe the ability, and to come up with clever enabling conditions clauses to cover the obvious (and as we have seen, only to be expected) failures of authority or ability. Anscombe's alternative promises a kind of universality of account that DOF theory cannot match. If one focuses on actual telic events, there will be no failures of reference on the telic side: for all telic events, a constitutive standard of correctness is applicable, just as for all beliefs, a constitutive standard of correctness is applicable. So one can offer a general constitutive account of these aspects of the mental in terms of the potential for mistakes, without the litter of qualifications, caveats and escape clauses that can make it seem implausible that one has uncovered something both unitary and fundamental.

Anscombe's basic insight is subtle, startling and original. We are quite happy in thinking that theoretical thought primarily manifests itself in a kind of state: belief. Anscombe's view suggests that practical thought primarily manifests itself in a kind of event: the telic event. In what follows I will explain Anscombe's distinction between mistakes in judgment and mistakes in performance, and then consider how it might be put to use in an account of these two elements of the nature of the mind. In doing so, I give up on intuitions that guide DOF theory, but that is as it should be.

1.5.2 Mistakes in Judgment and Mistakes in Performance

Anscombe's story of the shopper and the detective is supposed to illustrate the difference between mistakes in judgment and mistakes in performance: the detective makes a mistake in judgment, whilst the shopper makes a mistake in performance. But what exactly is the difference between the two, and how should it be understood?

Here we must be careful. Judging seems to be something one does, so there is a risk of conflating the two kinds of mistake by subsuming mistakes in judgment under the category of mistakes in performance. Anscombe's example is particularly unhelpful in this regard. The detective's aim is to investigate and record the contents of the shopping cart, so many or all of his mistakes in judgment will presumably explain, or be explained by (or perhaps even *be*) mistakes in his performance of investigating and recording. Coming from the other side, there is some temptation to subsume mistakes in performance under the category of mistakes in judgment. The etymology of 'mistake' itself suggests that mistakes must involve taking one thing to be another, which looks to be a purely theoretical error, as it were. Certainly the shopper's mistakes in performance seem to be explained by his mistakes in judgment (e.g. about whether this is butter); they might even be constituted by them, as he may have done everything else with perfect grace.

To avoid the potential confusion involved in thinking of judgment as a performance, I will change Anscombe's terminology and use the term 'mistaken belief' instead of 'mistake in judgment'. The difference between the two kinds of mistake can then be clarified by attending to a difference between being and doing. Being mistaken about whether p, by virtue of one's false belief that p, is something that one is. Fouling up the shopping trip, by virtue of picking up margarine instead of butter, is, by way of contrast, something that one does. *Acquiring* a false belief that p is something that can just happen to you (e.g. in idle observation), or it can be something one does in the progress of a performance with an explicitly cognitive end, as when one adds up sums incorrectly. In either case one ends up being mistaken, where the state of being mistaken is to be distinguished from the process of coming to be in the state. Mistakes in performance made during inquiry are then to be distinguished from the state of being mistaken itself just by the fact that mistakes in performance are not states, but rather activities or actions.

It would not threaten the distinction between mistaken beliefs and mistakes in performance if they always accompanied each other, but it would help if we could get a clear grasp of cases where one mistake is present independently of the other kind. We seem to be able to make sense of mistaken beliefs in isolation from mistakes in performance, because it seems quite possible to acquire false beliefs without any explicit performance being implicated in the acquisition. Idle observation and being raised to believe something false seem like plausible candidate cases here. Can mistakes in performance be made sense of in isolation from mistaken beliefs? Humberstone says that all mistakes in performance are attributable to false belief.[55] Anscombe does not agree with this common view. Consider the following example she gives:

> But is there not possible another case in which a man is *simply* not doing what he says? As when I say to myself 'Now I press Button A' – pressing Button B – a thing which can certainly happen. This I will call the *direct* falsification of what I say. And here, to use Theophrastus' expression again, the mistake is not one of judgment but of performance.[56]

The case is a *direct* falsification because it is not explained by a mistaken belief. This point is perhaps obscured by Anscombe's description of the case. If she pressed Button B instead of Button A, one might think this cries out for explanation in terms of e.g. a false belief that Button A is Button B. But all Anscombe needs for a clear example of the direct falsification of what one says is a case in she does not press Button A but does something else instead. A case of mere clumsiness or lack of grace – as when she stubs her thumb on the side of the phone booth machine (or on Button B) – can be one where she directly falsifies what she said she was doing whilst having no false beliefs about means to her ends, where the Button A is, where her thumb was when she began etc. As Anscombe says, this is "a thing which can certainly happen." Call a mistake in performance that is not explained by false belief a *basic mistake in performance*. (I will legitimate the concept of a basic mistake in performance in Chapter 3.)

---

[55] Humberstone (1992: 68).
[56] Anscombe (1957: 57).

There is an important difference between the temporality of mistakes in performance and mistaken beliefs. When one believes something false, one is mistaken about whether *p* for the entire duration of one's belief that *p*. Because mistakes in performance are things one does, they must be, in some sense, complete and irrevocable, because at some stage one must have done them. But whilst one is still doing something, one need not have made a mistake. Having fouled up the first shopping trip, the shopper (who still needs butter) must *try again*. But there is no mistake (yet) when he is halfway through his ordeal: he is, as it were, *still making* the mistake.[57]

We now have some idea of the difference between mistaken beliefs and mistakes in performance; but what unites them? Anscombe introduces the shopper and the detective and her example of a "direct falsification" of what is said immediately before her claim that modern philosophy is blind to the possibility of there being two kinds of knowledge.[58] Later she makes a cryptic remark to the effect that mistakes in performance help to resolve an apparent paradox concerning whether practical knowledge can be knowledge of what is not the case.[59] Given these passages, it is plausible that Anscombe thinks that the unity of mistaken beliefs and mistakes in performance is somehow bound up with the unity of practical and theoretical knowledge.

I confess that I do not know how Anscombe thinks that mistakes in performance help to resolve the paradoxical idea that practical knowledge can be knowledge of what is not the case. But we do not need to work this out in order to exploit her basic insight. The distinction between mistaken beliefs and mistakes in performance provides us with a general picture of mental activity as partially constituted by norms, against which distinctive kinds of mistake may be measured, coordinate with distinctive kinds of activity (where 'activity' includes e.g. being of a certain state of

---

[57] There are, of course, limits to the plausibility of claims that a 'mistake' is not yet an irrevocable mistake because one and the same performance is still in progress. It would be a bad joke if, having left the store without butter, the shopper were to respond to his wife's criticism by saying "What are you going on about? I haven't finished yet."
[58] Anscombe (1957: 57).
[59] Anscombe (1957: 82).

mind, and doing what one has in mind to do). Anscombe does not herself develop this picture into a general view of the metaphysics of mind required to make sense of such mistakes, nor she does she address difficult questions about how such a development might proceed. So let us turn to the prospects and puzzles involved in developing Anscombe's basic insight into a general account, on the understanding that the resultant view, though inspired by Anscombe, is not an exposition of her view, but something new.

1.6 PROSPECTS AND PUZZLES

I have suggested that Anscombe's basic insight can be used to construct an account of certain fundamental aspects of the mind – beliefs and telic events – in terms of (partially) constitutive standards of correctness, against which a distinctive kind of mistake may be measured. This promises a powerful general argument for the claim that the mind is essentially normative, and a distinctive and novel account of the nature and structure of intentional action. Let me conclude by summarizing some puzzles that such a project faces.

The first puzzle is in explaining precisely what it means to say that fundamental features of the mind, like beliefs and telic events, are partially constituted by norms. Beliefs seem like a paradigmatic exercise of a power of thought. So do telic events: they are 'the mind in action' (as we put it in §1.4.4 above). It is difficult to see how norms could enter into the constitution of a power or its exercises. If we are to develop Anscombe's insight into a general account of the nature of the mind, and to argue that the mind is essentially normative, we need a general account of a kind of *power* that is essentially normative.

The second puzzle is in explaining what mistakes in performance and mistaken beliefs have in common. As mentioned above in §1.5.2, there is some danger of conflating the two kinds of mistake, or subsuming one under the category of the other. We need to keep them apart if we are to remain true to Anscombe's basic insight, yet we also need to explain what they have in common, else the project will be open to the same sort of objection as that faced by DOF theory: namely, that there is nothing unitary and deep here to discover. We cannot easily appeal to Anscombe's cryptic remarks about knowledge to explain what unites the mistakes: those remarks are (for me at least) too cryptic. So we need a general account of what a mistake is that can make sense of both the unity and difference of mistaken beliefs and mistakes in performance.

The third puzzle is in determining the relationship between the standards of correctness for belief and telic events and what we might 'external accord' with those standards. For ease of presentation I have assumed that the standard of correctness for belief is truth and the standard of correctness for telic events is success. But merely true beliefs may be accidentally true, and merely successful performances may be accidentally successful performances, and these seem like degenerate cases that are in merely external accord with the relevant standards. If we say that the standards of correctness are *non-accidental* truth and success, for which there is no possibility of 'merely external' accord, then we come close to saying that the standard of correctness in the case of belief is *knowledge*, and in the case of telic events, *practical knowledge*. A satisfactory development of Anscombe's basic insight would clear up this question of whether the standards against which mistakes are measured allow for external accord or not, and if not, whether beliefs and telic events do in fact 'aim at knowledge'. (Clearing up this question will not amount to an account of practical knowledge, but it will in some sense be a propaedeutic to such an account, and a propaedeutic to deciphering Anscombe's cryptic remarks about the relationship between practical knowledge and mistakes in performance.)

In Chapter 2 I offer solutions to these puzzles. I solve the first two puzzles by offering a general account of fallible capacities: powers of thought that issue in perfect and imperfect exercises, where the imperfect exercises are mistakes. This account will be applied in a general argument for the essential normativity of the mind, given the pre-theoretically plausible claim that we make mistakes. I solve the third puzzle by showing why the relevant standards of correctness do not allow for external accord, so that beliefs and telic events do aim at (something like) knowledge.

2.0 WHY THE MIND IS ESSENTIALLY NORMATIVE

2.1 INTRODUCTION

One of the most familiar aspects of our mental lives is the fact that we make mistakes. Mistakes are things we make, not things that happen to us. In this chapter I develop these two thoughts into an argument for what I call the Normativist Claim:

The Normativist Claim Norms are essential to having a mind (like ours).[60]

I will call those who believe this claim 'normativists'. If the Normativist Claim is true, then the large number of naturalist views that aim to reduce or otherwise do without norms in their accounts of the nature of the mind must fail to adequately capture their subject matter.[61] That's a startling conclusion. For this reason alone, any contemporary philosopher of mind ought to be interested in whether and why the Normativist Claim is true or false.

---

[60] I should note at the outset that the Normativist Claim is implicitly restricted to minds *like ours*. Most normativists think that the Normativist Claim applies to the minds of imagined rational Martians; some (like me) suspect that it also applies to the minds of many non-rational animals. But most contemporary normativists don't think that the Normativist Claim applies to e.g. God's mind. I will be more specific about what I mean by "minds like ours" when we get to my argument for the Normativist Claim.

[61] Who aims to do without norms? Pretty much everybody. Analytic functionalists like Jackson (1998) and Shoemaker (2003) aim to do without norms. Computational theorists of mind like Fodor (2008) and Rey (1997) also aim to do without norms. Even philosophers of mind with teleofunctionalist sympathies like Dretske (1995), Millikan (1984) and Papineau (1993) could be said to aim to do without norms at the relevant (metaphysically fundamental) level of analysis. The list goes on and on. Searle (1983) counts as an exception, if only because he treats direction of fit as an unanalyzable, primitive and apparently normative term.

The chapter has the following structure. In §2.2 I identify a promising argument schema for the Normativist Claim. Applications of the argument schema require an account of an essentially normative power of thought. In §2.3 I outline an account of an essentially normative power of thought – a *fallible capacity* – and compare it to Aristotle's conception of a rational capacity. In §2.4 I argue that the mind is essentially normative, because the fallible capacity to know is essential to minds (like ours), and the fallible capacity to know cannot be reduced to something non-normative. In §2.5 I consider how the account of fallible capacities solves the puzzles raised in Chapter 1.

2.2 AN ARGUMENT SCHEMA FOR THE ESSENTIAL NORMATIVITY OF THE MIND

Normativists tend to argue for the Normativist Claim by applying the following argument schema:

***S**chematic **A**rgument for **N**ormative **E**ssentialism*

[or ***SANE*** for short]

P1 *Y* is essential to having a mind (like ours).

P2 A true and adequate statement of the essence (or metaphysically fundamental nature) of *Y* must use normative terms.

C1 So norms are essential to having a mind (like ours).

Y could be a mental property like *concept possession*, but it could also be a property of *Y-ity*, such as *conceptuality*, or it could be a property of engaging in the *Y* kind of activity, such as *conceptual activity*.[62] For example, Nick Zangwill (1995) has argued that *propositional attitudes* are essential to having a mind and that propositional attitudes have normative essences that cannot be explained in non-normative terms. Ralph Wedgwood (2007a, 2007b) has argued that *concepts* are essential to intentional thought and that concepts are possessed by virtue of possession of rational dispositions that cannot be specified in non-normative terms.[63] [64]

SANE depends on some understanding of the difference between normative and non-normative terms. Roughly, normative terms directly invoke a conception of how something ought to be, over and above mere description of how it is. Drawing the distinction between normative and non-normative terms more precisely than this is notoriously difficult to do. I will follow the bulk of the literature in listing paradigmatic normative terms rather than trying for a general account of the normative / non-normative divide. Terms like 'ought' and 'correct' count as paradigmatic normative terms. Terms like 'is disposed to', 'belief' and 'desire' do not.[65]

There is a dialectical problem for *SANE*: it risks begging the question in P2. Consider a reductive naturalist philosopher of mind who is loathe to accept the Normativist Claim. She might

---

[62] What a "true and adequate statement of essence" amounts to I don't really know. I am not primarily a metaphysician, although I have been forced to dabble. In what follows I am dodging the questions of whether there's no such thing as essence, or no such thing as true and adequate statement of essence. See Kit Fine (1994a, 1994b, 1995) for helpful (preliminary) work on the ground rules for statements of essence.

[63] *SANE* is found in the work of other normativists apart from Zangwill and Wedgwood. Brandom (2000), for instance, is often cited as a key contemporary normativist, and his work can be read on the model of an application of *SANE*. Brandom argues that participation in a socially-mediated practice of *commitment and entitlement* is essential to having a mind, and that commitment and entitlement are normative statuses that cannot be reduced to something non-normative. Commitment and entitlement, and the kind of activity appropriate to them, are Brandom's substitutions for *Y*. But Brandom doesn't do metaphysics in any straightforward sense, so the words "essential to having a mind", as *he* would employ them (were he to ever use them) would require a difficult non-metaphysical interpretation. I won't address Brandom directly in this paper, as it would take us far afield.

[64] Wedgwood (2007a: 172) might object to the second premise of *SANE*, but only because of qualms about the difference between normative terms and "mentioning normative properties". I ride roughshod over some of the subtleties here in order to bring out the common theme.

[65] So even if Zangwill is right that belief and desire have normative essences, the normativist cannot presuppose that 'belief' and 'desire' are normative terms. If *SANE* is to be even mildly persuasive, the normative terms that figure in P2 must be *clearly* normative, in a sense that is relatively uncontested.

accept that minds like ours are subject to mental norms, or assessable in terms of mental norms, where mental norms are epistemic norms, or perhaps norms of practical rationality, or perhaps norms of proper mental function more generally. She might even accept that minds like ours are *necessarily* subject to, or assessable in terms of, mental norms. Where she balks is at the claim that true and adequate statements of the essence of minds like ours *must* include normative terms. There are a number of understandable reasons why one might balk in this way. For example, one might think that norms are not causally efficacious, and also think that only terms that refer to causally efficacious things should figure in true and adequate statements of essence. (Many physicalists think that something like this is true, and have related doubts about mental causation, given the causal closure of the physical.)[66] Such a reductive naturalist might well accept, at least for the sake of argument, that *SANE* is a valid argument schema.[67] She might well accept a version of P1: suppose she thinks that *representation* is essential to minds like ours. But she is likely to find particular versions of P2 question-begging. If *Y* (e.g. representation) is indeed essential to having a mind, then explaining *Y* in non-normative terms is precisely what the reductive naturalist aims to do.[68]

Applications of *SANE* wouldn't beg the question if there were support for P2 that were independent of accepting the Normativist Claim. But so far, no normativist has been able to come up with the required independent support.[69] Wedgwood provides a good example of the dialectical

---

[66] See for instance Kim (1993).

[67] In what follows I assume that *SANE* is a valid argument schema. This requires a very strong reading of P2. P2 does not merely make an epistemological point about us. It's not just that we (stupid monkeys) must use normative terms in stating the essence of *Y*, because of some idiosyncratic epistemological or linguistic limitations. Rather anyone (even God) who aimed to truly and adequately state the essence of *Y* would have to use normative terms in doing so.

[68] My use of the term 'reductive naturalist' is somewhat non-standard. Reductive naturalists in philosophy of mind typically aim to explain the nature of the mind without appeal to *psychological* terms. They don't aim to explain the nature of the mind in non-normative terms, except insofar as that is a consequence of explaining the nature of the mind in non-psychological terms. I explain the particular sense of reduction in more detail in §2.4.2.

[69] Zangwill (1998) argues that analytic functionalist accounts of the nature of the mind fail to account for systematic irrationality and causal deviance, and concludes that we need to accept the Normativist Claim. But he never actually argues that the failure of analytic functionalism implies the truth of his version of P2. Other reductive projects might succeed where analytic functionalism fails, in which case we don't need to accept Zangwill's P2, and so don't need to accept the Normativist Claim. Zangwill indirectly acknowledges this point. He notes that his attack of analytic

difficulties here. He argues that rational dispositions must be specified in normative terms, and that concepts are possessed by virtue of possession of particular rational dispositions. But there's a big difference between saying that concepts *can* be possessed by virtue of possession of a particular rational disposition, and saying that they *must* be possessed this way. Even granting that rational dispositions must be specified in normative terms, and that they can ground concept possession, one might still wonder: why can't irrational, or non-rational, dispositions ground concept possession? When Wedgwood considers this kind of possibility, he says the following:

> …it seems to me doubtful that one's possession of a concept can rest on an irrational disposition… concept possession is a cognitive *power* or *ability*, not a cognitive defect or liability.[70]

Wedgwood says this because he thinks that when we employ a concept in thought, we exercise the very same kind of cognitive power that a perfectly rational subject would exercise in employing the very same concept. But he doesn't actually have a good argument for the claim that being like the perfectly rational subject in this way is the *only* way to possess a cognitive power, or to possess a concept.[71] Georges Rey has a direct response to Wedgwood that brings out very well how the move to cognitive powers is no help, dialectically speaking. Rey says this:

> …it does rather beg the question to suppose that concept possession must be the kind of "power" or "ability" that couldn't be a "defect" or "liability". The question is why we should think that it is a power that needs *intrinsically* to be described in any normative terms at all.[72]

I think that Rey's question is an excellent one, and it deserves a direct answer. If the normativists are to move the dialectic along, they need to explain how there could even be a kind of power that

---

functionalism doesn't touch teleofunctionalism or Bill Lycan's 'homunctionalism'. Given that Millikan's earlier work on teleofunctionalism sketches a reduction of mental norms to a kind of evolutionary advantage, Zangwill hasn't given independent support for the view that normative terms *must* be used in true and adequate statements of the essence of the mind. In his most recent work Zangwill says that he just wants to speculate on the consequences of accepting the Normativist Claim, rather than arguing that it must be accepted. See Zangwill (2005, 2010).

[70] Wedgwood (2007b: 168-9).

[71] Wedgwood (2007a: 172) does say that a perfectly rational subject could explain our errors in concept-use to us. But this does not show that our powers of thought must be of the same metaphysical kind as those of the perfectly rational subject. The perfectly rational subject could explain our errors precisely in terms of a metaphysical difference in the kind of causal power at work, if concept possession can in fact be grounded in many ways, and not only by (approximating to) manifesting a rational disposition.

[72] Rey (2007: 77).

"needs intrinsically" to be described in normative terms, and then argue that the mind's essential powers are of that kind. It's not enough to gesture towards an idealized perfectly rational subject and say that our powers of thought are "like that". There's a clear sense in which we're obviously not "like that". *We* bumble about like idiots. It's a substantive claim, and one that's hard to defend, to say that the bumblings of an idiot *must* be of the same metaphysical kind as the perfectly rational thoughts of an angel, or the perfectly rational deeds of a *phronimos*.[73]

In what follows I will try to turn our familiar bumbling into a virtue of a normativist account, so as to answer Rey's question directly, so as move the dialectic along. The answer goes as follows. There are indeed powers that need intrinsically to be described in normative terms. These powers are what I will call *fallible capacities*: capacities that admit of both perfect and imperfect exercises, where (some of) the imperfect exercises are *mistakes*. The mind's essential capacities are of this kind, and this is why the mind is essentially normative.

2.3 CAPACITIES, FALLIBLE CAPACITIES AND RATIONAL CAPACITIES

2.3.1 Capacities

Thinking is something that minds, by their very nature, *do*. "Capacity talk" is, for me at least, just a way to talk about such essential causal powers of minds, like the capacity to think. Let me say a few words about "capacity talk" before proceeding to the account of fallible capacities.[74]

---

[73] In the context of a discussion about practical thought, Setiya (2007: 64) makes essentially the same point: "Why accept that the motivation of imperfectly rational beings, like us, must be explained in each case by its resemblance to good practical thought, as though our failures are mere perturbations of a system that is otherwise ideal?"

[74] I use the terms 'power', 'causal power', 'ability' and 'capacity' interchangeably. Sometimes the word 'disposition' might also do. Many views of dispositions are indistinguishable from views about capacities etc. Views about dispositions differ

Capacities, as I understand them, are possessed by individuals, have conditions of exercise, and are individuated by what they are capacities to do.[75] We can both be moved to tears, and in that sense we share a capacity, but what moves you to tears need not be what moves me to tears (the conditions of exercise differ for you and me).[76] Capacities explain their exercises as non-accidental (in some sense of "no accident"). Capacity-specifications have an infinitive form: one has a capacity *to A*. But no hefty metaphysical lessons should be read off from the grammatical form of these specifications: some capacities are capacities to engage in *activities* or *processes* whilst others are capacities to be in *states*.

Determinative exercises of the capacity to *A* are exercises that are determinations of *A*: cases or ways for *A* to be wholly manifested in some state or episode. Walking jauntily to the left is a determination of walking, for example. Capacities exhibit a distinctive unity of explanatory power in their determinative exercises. We do not explain the hot knife's action on nearby heatable objects by citing one capacity to heat butter; another capacity to heat bread a bit on Mondays; and so on. Arbitrarily isolating a subset of determinative exercises of a capacity in this way does not automatically serve to individuate a distinct capacity; neither does arbitrarily isolating subsets of conditions of exercise, such as being held on one or other side of the butter. That said, it is harmless to talk profligately about the number of capacities that something possesses, and to sort them into kinds according to arbitrary modes and conditions, so long as it is remembered that this is not a good guide to individuation. I shall engage in such profligacy often in what follows, on the assumption that some happy middle ground is available between the view that we each have one

---

from views about capacities to the extent that they include some notion of comparative quantity (e.g. if I am disposed to *A*, then *A* is something I can do, but moreover I tend to *A* more often than not.)

[75] By 'individuals' I mean to encompass e.g. individual groups, associations etc.

[76] If I can do a two-meter long jump on Earth, but you can only do one on the Moon, isn't it true that I have a capacity that you don't? In the sense of potential range of determinative exercises, you can do something that I can't, but in both cases the exercise of the capacity to jump two-meters is an exercise of the capacity to jump (as one wills) and that is something that we share. "Capacity talk" is often used to compare ranges of potential determinative exercises, but it is important to note that not every colloquial difference corresponds to a difference in capacity.

capacity *to do stuff* (with an infinite number of more or less specific determinative exercises) and the view that we each have an infinite number of more or less specific capacities (one for each halfway-plausible verb phrase, such as "to crawl doggedly across the road in response to heartbreak on a Saturday night").

Is it prejudicial to conceive of mental activity in terms of the exercise of capacities? Talk of capacities and their exercise fits seamlessly with most accounts of the nature of the mind. It does conflict with the view that, strictly speaking, there are no causal powers possessed by objects, but only instantiations of properties and relations at various times according to exceptionless laws of nature. There is a considerable body of literature supporting the view that we cannot do without the concept of objects possessing and exercising causal powers.[77] I defer to such literature in defense of the very idea of a causal power. Capacity talk is just a way to talk about such causal powers.

2.3.2 Fallible Capacities

All capacities have perfect (i.e. determinative) exercises.[78] Fallible capacities have *perfect* and *imperfect* exercises. Perfect exercises of the fallible capacity to *A* are exercises that are determinations of *A*. Imperfect exercises of the fallible capacity to *A* are exercises that are *not* determinations of *A*.

Are there any such imperfect exercises of capacities? Consider a baseball player practicing catches. The ball flies high, she positions herself under it, it's an easy catch, and she catches it smoothly. This seems ground enough to say that she has the capacity to catch balls of that type. Consider the next case. The ball flies just as high, she positions herself under it the same way, it's the

---

[77] For example, coming from the direction of philosophy of science, see Cartwright (1994), Machamer, Darden and Craver (2000). Coming from direction of literature on dispositions and abilities, see Mumford (1998), Molnar (2003).
[78] The phrasing here is correct, but slightly misleading. If there are non-fallible processual capacities, then they will have perfect determinative exercises as well as grammatically-imperfect exercises, even if they have no normatively-imperfect exercises. See below for a description of processual capacities.

same kind of easy catch, and yet she fumbles it. She was in the right conditions of exercise for her capacity, given that the conditions were the same as for the first catch, and she didn't lose the capacity momentarily, only to magically regain it again when she makes the next catch. So on the face of it, the fumbled catch presents us with an imperfect exercise of her capacity to catch balls of that type. Such things happen all the time.

There will of course be an explanation of the failure: perhaps she dropped her glove-hand a little at the last moment whilst checking third base, and then couldn't bring it back up in time when she looked back.[79] (A coach might point this out: it could be the basis for improving her technique.) The case is nevertheless distinct from other forms of failure to do something one has a capacity to do because the *source* of the failure is found in the catcher herself and her exercise of her powers, such as they are.

There are other cases where failure to catch a ball may be attributed to *external or internal interference*, as when someone pushes the catcher (external interference), or a sudden heart attack or hiccup perverts the progress of what otherwise would have been a smooth catch (internal interference). Cases of interference provide a merely *grammatical* sense of imperfection: the same sense in which anything that takes time to do can be interrupted by something alien and so left incomplete. (It is grammatical imperfection because one can truly say: "*X* was *A*-ing but did not *A*.") Failure may also be attributed to *prevention*, as when someone prevents the catcher from getting near the ball at all. But cases of interference and prevention aren't cases of an imperfect exercise of a capacity to catch high-flying balls, where the imperfection has a *normative*, rather than merely grammatical, significance. In cases of interference, it is not a normatively-imperfect exercise if the

[79] Shouldn't this imply that the conditions of exercise weren't the same? No. Conditions of exercise for a capacity are external to the exercise itself, but catching a ball is an activity that includes certain kinds of preparation within its scope, so those preparatory activities cannot be considered conditions of exercise. Otherwise no one could catch a ball until the conditions of exercise were such that they had already caught it.

catcher couldn't be expected to know about the interference, so as to try and do something about it. We might say that the exercise was (rudely) *interrupted*, but it seems false (and mean) to say that it was *imperfect*, in a sense that traces the imperfection to some failing in the catcher and the exercise of her powers.[80] Cases of prevention are defined by the fact that they prevent exercise of the relevant capacity. So normatively-imperfect exercises of capacities represent a third category of failure to do something one has the capacity to do, distinct from both *prevention* and *interference*. Let us label this third category that of *mistakes*.

Mistakes have not been acknowledged as a distinct category of failure by most contemporary views about dispositions or abilities.[81] Fara (2008) comes close: he classifies the case of a seasoned golfer failing to make an easy putt as a case of a "masked" ability, where some sudden gust of wind or distraction explains the failure. The case of distraction looks similar to the second of our ball-catching examples.[82] But Fara says that masked abilities are abilities that the agent *fails to exercise* for some reason, despite being in appropriate circumstances for the exercise, and despite the agent trying to exercise the relevant ability. This commits Fara to the view that whatever the golfer does is *not*, in any sense, an exercise of his ability to sink easy putts, and perhaps to the view that abilities are only exercised when they are exercised perfectly.[83] This latter commitment is found in most contemporary views about dispositions and abilities: according to contemporary views, every

---

[80] It may be a different matter if the catcher saw the person coming, or if catching whilst dodging, or fighting, (or hiccupping) is a part of exercising the relevant skill.

[81] See for example Armstrong, Martin & Place (1996), Lewis (1997), Manley & Wasserman (2008), Mumford (1998), and Ryle (1963). See Bird (2012) for some indirect reflections about the relationship between linguistic/conceptual analysis and the relevant metaphysical thesis concerning perfect (determinative) exercises.

[82] The case of a sudden gust of wind employs the merely grammatical sense of imperfection, for the gusts of wind are *sudden*, and so are unknowable ahead of time. Things are different with prevailing winds: golfers are expected to account for them, and doing so is a test of their skill. The same is true of (most) distractions: a certain degree of steady care and attention is part of golfing (well). This issue receives brief further discussion towards the end of Chapter 3.

[83] In Fara's official definition of masking, the crucial condition is that the agent fails *at whatever she is trying to do*, rather than that she fails to exercise the relevant ability. This definition could accommodate mistakes as cases of "masking". But elsewhere Fara is relatively clear that he does not think the ability has been exercised in such cases: the failure to achieve the end *is* a failure to exercise the relevant ability when the ability is masked. Fara's identification of these two failures would be explained by adherence to the claim that abilities are only exercised when they are exercised perfectly.

exercise of the disposition or ability to *A* must be a *perfect* (i.e. determinative) exercise.[84] If we countenance mistakes, we move beyond contemporary views in this regard. There are cases of failure where the baseball player's capacity to catch easy balls *is* exercised in the circumstances – but *imperfectly* so – for missing an easy ball is *not* a case or a way of catching an easy ball.

The very idea of capacities with imperfect exercises may seem outrageous and outlandish, especially as it flies in the face of contemporary wisdom about dispositions and abilities. We can ease our way into the idea by considering the merely grammatical sense of imperfection I mentioned before. Consider capacities whose exercises take time to resolve into their proper end state: for example, the capacity to walk across the road. Until one has crossed the road there is no state or episode that is a determination of what the capacity is a capacity to do, for the capacity is not merely a capacity to be walking with a certain goal in mind, but to actually get to the other side. Yet considered at some time before the proper resolution has been reached, it would be bizarre to credit some *other* capacity with the progress to date. So if there are capacities that are capacities to do something that takes time to reach its proper resolution – call these *processual* capacities – then there are capacities that have exercises that are not determinations of what the capacity is a capacity to do.

So far I have characterized both fallible and processual capacities in negative terms: as capacities that have imperfect exercises that are *not* determinative exercises. Having eased our way into the idea of the general class of capacities with imperfect exercises, we need a positive way to distinguish processual from fallible capacities.[85] We cannot rest content with the mere labels "grammatical imperfection" and "normative imperfection". We need to understand what the labels mean.

---

[84] I think commitment to this claim is the *source* of problems about "masking", "finking" etc. The importance of such cases stems from the fact that the logical category employed in the analysis of dispositions is too narrow.

[85] Note that a capacity can be both fallible *and* processual, as the capacity to catch balls is. The terms 'fallible' and 'processual' mark possibilities for exercise of a capacity, not wholly distinct classes of capacity.

A logical relation is required to make sense of which episodes and states count as exercises of a capacity and which do not. When we deal with perfect exercises of capacities, the logical relation is easy to specify: perfect exercises will be species of some genus, or determinations of some determinable, where the genus or determinable is whatever the capacity is specified as a capacity to do. For example, the chameleon's change of its skin-color to blue, so as to blend into a blue environment, is a perfect exercise of its capacity to change its skin-color to blend into its environment. When we deal with the grammatically-imperfect exercises of processual capacities, the logical relations found in *mereology* seem appropriate. Walking halfway across the road is not a species or determination of walking all the way, but when all goes well it is a part of walking all the way. We need to sketch a similar logical relation that is characteristic of normatively-imperfect exercises of fallible capacities.

The answer I have hit upon (with help from Aristotle) is this: the logical relation is one of *preclusion*, in the sense of mutual temporal non-compossibility.[86] Normatively-imperfect exercises of fallible capacities are determinate ways of falling short of a perfect exercise, where falling short in that determinate way is not attributable to interference, and falling short in that determinate way *precludes* a perfect exercise on that occasion and in that regard. For instance, in the case of the fumbled catch, an inadequate ordering of preparatory activities precludes perfection. (Remember how the catcher dropped her glove hand, so that she *couldn't* bring it back up in time to catch the ball.) Supposing that false belief or inadequately justified true belief is an imperfect exercise of the fallible capacity to *know*, an inadequate ordering of epistemic grounds precludes knowing the relevant truth on that occasion and in that regard. *Every* mistake consists in an inappropriate or inadequate ordering of some kind of activity, in a suitably broad sense of activity. We know this

[86] Apart from Aristotle (1995), credit is also due to Kimhi (forthcoming), McDowell (2011), and Rödl (2007), who all develop Aristotle's conception of a 'two-way' rational capacity to a greater or lesser degree.

because in each case of such imperfection, one can in principle give an explanation of *why* the exercise goes wrong that consists in pointing out *how* the relevant ordering of activity precludes perfection on that occasion and in that regard. If we couldn't explain how the ordering of activity precluded perfection, we would have good reason to suppose that the failure wasn't with the bearer of the capacity in the way relevant to attribution of a mistake; it would rather constitute some problem with interference or conditions of exercise.[87]

Preclusion is local to an individual *occasion* and *regard*. Consider a pedestrian who walks halfway across the road, dances back from an approaching bus, and then completes her journey, "all in one go" (as we say). What this pedestrian did need not be a mistake because it did not preclude her getting all the way across "all in one go". It could have been a mistake, had she been squashed, or (less gruesomely) driven to squander a genuine opportunity by retreating to the curb, so that she had to try again. (What is meant by 'try again' here is part of the subject matter of Chapter 3.)

I have invoked Aristotle. We should see what the great man had to say about capacities that admit of two kinds of exercise, if only to help illuminate the nature of the current proposal.

2.3.3 Aristotle on Rational Capacities

The first philosopher to have developed the idea of capacities that admit of two genuinely different kinds of exercise is Aristotle.[88] In Metaphysics Theta Aristotle distinguishes non-rational capacities from rational capacities in terms of the number and kinds of exercise of which they admit. The

---

[87] There is a difficulty about omissions. Some mistakes consist in, for instance, not paying enough attention or care to what one is doing given one's knowledge of circumstances, and it is hard to see how an omission could be *part* of an inappropriately-ordered activity. Chapter 3 sketches a solution to the problem for the case of intentional action.

[88] See Beere (2010) and Makin (2006) for excellent commentaries on Aristotle's distinction between one-way non-rational capacities and two-way rational capacities. The exposition of Aristotle in this section is my own, although I have benefitted from both Beere and Makin's work in coming to this interpretation.

current proposal is inspired by Aristotle's distinction, but it is not identical with it, so it will help to point out the similarities and differences.

According to Aristotle, non-rational capacities are "one-way" capacities in that there is only one kind of exercise of the capacity. Aristotle's example is the capacity hot rocks have to heat heatable things, like cold water. Should the bearer of an active non-rational capacity meet with a suitable patient in the conditions of exercise for that capacity it invariably produces the associated change in the patient. Something similar, suitably transformed, holds for passive capacities, such as the capacity to be heated. The claim about invariable affection is not as crazy as it may seem. Aristotle thinks that a proper understanding of what the capacity is a capacity to do already includes understanding of the appropriate conditions for the relevant exercise of the capacity – the capacity is to heat under certain conditions, not to heat whilst wrapped in an insulating bag – so there is no need to qualify claims about whether the capacity will be exercised with clauses such as "provided nothing external hinders it from acting" (see 1048a5 and 1048a20).[89]

Rational capacities are "two-way" capacities in that there are two opposed kinds of exercise of the capacity. Aristotle's examples are arts such as building and medicine. The one who can build or heal knows how to produce a building or health, but such knowledge consists in grasp of a *logos* or rational account, and so is by implication knowledge of the privation of structural integrity or health. If you can heal, for instance, you know the order of health, and so know the steps towards

[89] The insulating bag is a bad example. A hot thing in an insulating bag will heat the insulating bag (a little bit) even if it fails to heat further things via the material mediation of the insulating bag. This suggests the interesting conclusion that inanimate things never fail to do something they have a capacity to do when in the conditions of exercise for that capacity. The hot thing *heats* even if it doesn't heat this heatable object, or heat this heatable object sufficiently for *X* (e.g. up to 25 degrees). If this is right, then no inanimate object has a *processual* capacity, for these could be interrupted by something alien so as to result in a failure. In fact, the last claim seems plausible. Processual capacities must have a proper end point for their exercises. It is not in the nature of a rock to hang around near a heatable object long enough to accomplish some arbitrarily limited change, so in that sense whether it does or not is an accident as far as the rock and its powers are concerned. It *is* in the nature of animals to hang around in an environment until they have accomplished some (more or less arbitrarily) limited change. Nothing hangs on whether inanimate objects have processual capacities or not, so I do not pursue the matter in this chapter.

health, and by reversing an order of reasoning towards health you know the steps away from health as well. Should the bearer of a rational capacity meet with a suitable patient in suitable conditions of exercise for the capacity we cannot tell what will happen unless we know something more about the bearer of the capacity: namely, what she desires or chooses to do with her knowledge. Should the healer desire to heal, healing will ensue; should she desire to harm, illness (in the sense of privation of health) will ensue; should she desire neither to heal nor to harm, no exercise of the capacity will ensue. Aristotle also says that rational capacities are more properly capacities to produce one of the opposed contraries, and only "in a way incidentally" capacities to produce the other.[90] His reason for saying that rational capacities are "positively valenced" towards one exercise rather than the other is that rational understanding is of some unified thing in itself (e.g. health) and the knowledge one has of the contrary is by means of negation or privation of the relevant unity. We cannot understand, for instance, illness in all its variety except by relating it to health, but we can understand health on its own terms as the flourishing of some organism.

For present purposes, the most important structural feature of rational capacities is that they are defined by their "positive valence" towards an ideal unity. The capacity to heal is *a capacity to heal* (*simpliciter*), not a capacity to heal or harm. It was mentioned above that we do not postulate distinct capacities willy-nilly because that would do violence to the unity of explanatory power that capacities exhibit. In the case of rational capacities, that unity of explanatory power is of a special kind. One does not have a capacity to harm (through medical knowledge) that is different from one's capacity to heal (through medical knowledge), because to suppose that one does would be to violate the special unity of explanatory power that rational capacities exhibit: the *single* unified order of reasoning involved, and the *privilege* of one of a pair of contraries (health, not illness) in delimiting

---

[90] Aristotle (1995: 1046b13).

the nature of that order of reasoning and associated capacities that apply cognitive grasp of that order in action.

Fallible capacities are similar to Aristotle's rational capacities in that they defined by their "positive valence" towards an ideal unity, where the ideally unified activity is contrary to indefinitely many ways of falling short of that ideal unity.[91] But Aristotle's conception of rational capacities is different to the present proposal in a number of ways. The first difference is that our rational capacities presuppose the presence and exercise of fallible capacities. The surgeon who desires to harm may *fumble* his vicious scalpel incision, making his exercise of his capacity to heal doubly incidental: incidental by the lights of the kind of exercise of his knowledge in play (harming, not healing) and incidental by the lights of the fallibility of this capacity (fumbling, not cutting). The second difference is that it would be quite wrong to say that e.g. whenever a thinking subject believes what's false she *desires* to believe what's false. The element that determines what went wrong or right in some exercise of a fallible capacity will be internal to the bearer of a fallible capacity in an interesting way (consider again the baseball catcher's inappropriate ordering of her preparatory activity), but usually the internal explanation need not invoke desire. The third difference is that non-rational animals seem to have fallible capacities (as witnessed by many *America's Funniest Home Videos* entries) and at least as far as Aristotle's official definitions go, non-rational animals could not have rational capacities.[92]

Differences aside, Aristotle's analysis of kinds of capacity provides us with a model for how to resist the idea that there is a way to render the operations of one kind of capacity in terms of the operations of some number of capacities of a different kind. The interesting aspect of Aristotle's views about the metaphysics of capacities is that no combination of one-way capacities could

---

[91] Here it is important to stress that 'activity' does service for activity, process, state, etc.
[92] Makin (2006) suggests that this is merely an oversight on Aristotle's part, and fills in an account on his behalf.

account for the unity of explanation found in the explanation of someone's healing or harming as the exercise of her rational two-way capacity to heal. It is by exploiting an analogous unity of explanation that we will resist the urge to reduction in the case of fallible capacities that are essential to having a mind (like ours).[93]

2.3.4 Two Important Features of Fallible Capacities

Let me highlight two features of fallible capacities that will be important for my argument for the Normativist Claim.

2.3.4.1 Fallible Capacities Set Standards of Correctness

The first important feature is that fallible capacities set standards of correctness. It follows from the fact of an imperfect exercise that there are grounds for criticism of what was done. It would be unintelligible for someone who understands the criticism to deny its relevance to what she ought to have done on that occasion in that regard, even if the mistake doesn't register significantly on some other normative or evaluative measure (e.g. the measure of justice, or glamorousness). I will mark this normativity by saying that a fallible capacity sets a *standard of correctness* for the bearer of the

[93] Why only analogous? Apart from the differences noted, I am not sure that Aristotle, were he introduced to the contemporary terminology, would agree that rational capacities are essentially normative. *Pace* Hippocrates, there need be nothing *wrong* or incorrect with the incidental exercise of a surgeon's capacity to heal in some case. Aristotle is very clear that something else – namely desire – dictates what one does with a rational capacity. I suspect Aristotle would say that correct desire also dictates what one ought to do with a rational capacity. Certainly there is no indication in Metaphysics Theta that desire, or exercise of the capacity, ought to tend towards promotion of the relevant unity (e.g. promoting health) rather than its privation (not even *prima facie*). The standards for correct desire may come completely apart from the ideal unities of which we can have knowledge that can be applied in action.

capacity in her exercises of that capacity. The exercise of a fallible capacity is correct iff it is a perfect exercise and it is incorrect iff it is a normatively-imperfect exercise.[94]

There are no intermediate degrees of partial correctness or incorrectness for the standard of correctness set by a fallible capacity. This may seem counterintuitive. If I make some minor mistake and fail to catch a ball, but very nearly catch it, isn't that better than making a more drastic mistake where I do not even come close to success? If I have some evidence for a false belief, isn't that better than having no evidence at all? The answers to these questions are to one side of the question of whether the relevant fallible capacity was exercised imperfectly or not. The standard of correctness set by a fallible capacity concerns only this question. As noted in Chapter 1, a false belief is *mistaken* regardless of one's evidence for it.[95] Similarly, a fumbled catch is a failure, regardless of how close one came to success. The criticism relevant to imperfect exercises of fallible capacities is, first and foremost, concerned with what is *required* for perfect exercise and what *precludes* perfect exercise, not with makes for better imitation of, or approximation to, perfect exercise.

This does not mean that there couldn't also be a standard of excellence that admits of many degrees of partial fulfillment that applies to the relevant exercises. But any such standard of excellence would be distinct from the standard of correctness, precisely because it allows of many degrees of partial fulfillment. Consider my fallible capacity to write a neat letter 'g' for instance.[96] I can write a letter 'g' that is very messy; suppose this is a normatively-imperfect exercise of the fallible capacity to write a neat letter 'g'. If I write a *slightly* messy letter 'g', that seems to approach closer to a

---

[94] If the future is genuinely open, such that the Law of the Excluded Middle fails for some things that are (at the moment) still developing in time, then fallible processual capacities allow a middle ground that is neither correct nor incorrect whilst the exercise is still in progress, because the occasion hasn't yet resolved one way or the other.

[95] In the present context we are also assuming that an accidentally true belief is also mistaken in the relevant sense of 'mistaken'. I will justify this classification of accidentally true beliefs in §2.5.

[96] Particular thanks are due to Kieran Setiya for the example and for pushing me to address this point in more detail.

normative standard of writing a neat letter 'g' than writing the very messy letter 'g' did.[97] The nature of the capacity itself may seem to set a standard of excellence that can be met or fulfilled to varying degrees in these imperfect exercises.[98] If it does, the standard of excellence must be distinct from the standard of correctness set by the capacity. The standard of correctness is not met or fulfilled in either case, because in neither case is a neat letter 'g' produced: I didn't do what was required. The purported standard of excellence is met or fulfilled in both cases (to some degree, but met or fulfilled nonetheless). Based on the property of being met or fulfilled, Leibniz's Law dictates that the standard of correctness and the standard of excellence must be distinct.

2.3.4.2 Fallible Capacities are Essentially Normative

The second important feature is that fallible capacities are themselves essentially normative. Although preclusion is a symmetrical relation, fallible capacities are *not* indifferently related to their mutually incompatible exercises. Suppose we have a fallible capacity to know, where cases of false or inadequately justified belief (cases of "mere belief") are the imperfect exercises that preclude knowing on some occasion in some regard. This fallible capacity to know, by its very nature, is not indifferently related to its mutually incompatible exercises. It is a fallible capacity to *know*, not a fallible capacity to merely believe. If it were a fallible capacity to *merely believe*, then mere belief would be the perfect correct exercise, and knowing would be the imperfect incorrect exercise, for which the bearer of the capacity would be liable for criticism. ("You ended up knowing, you fool! You're supposed to merely believe! Now let's see what you did wrong…") The structure of a fallible capacity is such that it already includes within itself a positive normative assessment of one of a pair

---

[97] The standard need not be an *ideal of perfection*, such as the Form of Neat 'G' that is unrealizable by mere mortals, because it requires a perfect circle as a part. There may be many neat 'g's that I have produced in my life.

[98] In fact, although the neat letter 'g' provides a model to which one may approximate by degree, I think there is no standard of excellence here. Neatness only matters as subordinated to some further end. Doctors know this; that is why they sign prescriptions the way they do. A better example would be approximating to the behavior of the *phronimos*.

of mutually opposed contraries, because *that* one of the pair (e.g. knowing) is what the capacity is individuated as a capacity to do, not the other. Similarly, when we explain a state or completed episode as an exercise of the fallible capacity to *A*, and there is no interference, we already assess the state or episode as correct or incorrect depending on whether or not it is a determination of *A*.

When I say that fallible capacities are essentially normative I mean to invoke the same gloss on "essential normativity" that *SANE* invokes: I mean that a true and adequate statement of the essence of a fallible capacity must use normative terms. This invites a direct objection. Someone might grant that norms necessarily follow from possession of a fallible capacity, and yet still wonder why we can't state the essence of a fallible capacity without using normative terms. Consider the fallible capacity to know, for instance. The terms 'correct' and 'criticism' seem like normative terms, but the terms 'know,' 'mere belief,' 'determinative exercise' and 'preclusion' seem like epistemological or logical terms, rather than distinctively normative ones. So why can't we just use epistemological and logical terms to truly and adequately state the essence of the fallible capacity to know, without going on to talk about standards of correctness, possibilities for criticism and so on?[99]

The normative aspect of a fallible capacity cannot be so easily separated from its nature. Consider the difference between the fallible capacity to know and the disjunctively-specified capacity to know or merely believe. The disjunctively-specified capacity is, in itself, *indifferent* to its mutually incompatible exercises, but its exercises are co-extensive with those of the fallible capacity to know. If we had reason to, we could perhaps haul in some principle from elsewhere that serves to privilege one type of exercise of this disjunctively-specified capacity over the other (i.e. privileging knowing over merely-believing). But we couldn't say that any normative privilege flows from the nature of the disjunctively-specified capacity itself.

---

[99] Horwich (1998) makes this kind of objection with regard to the sense in which truth is normative for belief.

If we do not think of fallible capacities as *essentially* normative, then I think we cannot tell the difference between a fallible capacity that explains the relevant normative privilege by reference to its own nature, and a disjunctively-specified capacity that does not explain any normative privilege, but requires supplementation from elsewhere. The individuation of a fallible capacity is not clearly prior to the determination of which of the two classes of exercise should be normatively privileged. Rather, the individuation of a fallible capacity *is* the normative privileging of one class of exercise over the other. The difference between fallible capacities and disjunctively-specified capacities consists in the essential normativity of the former. Another way of putting the same point: the normative term one must use in stating the essence of a fallible capacity is not 'correct' or 'criticism', but rather 'privilege' or 'valence'.

Schroeder (2003) anticipates the kind of (very abstract) normativity at work here. He argues that a genuinely normative account of the mind must not only divide cases of mental activity into exclusive classes, but also provide a "force-maker": a normative valuing of at least one set of cases over the others. He calls this giving "normative *oomph*" to the merely logical division of cases. He also notes that sometimes, such as when one is considering the division of people into courageous and cowardly, the distinction between division of cases and normative force-maker may collapse. Fallible capacities are like that too. The "normative *oomph*" is not just conjoined with the individuation of perfect and imperfect exercises; it *is* the individuation.

2.4 THE MIND IS ESSENTIALLY NORMATIVE

2.4.1 An Application of *SANE*

Now we are in a position to argue for the Normativist Claim. It is part of an ordinary conception of the nature of our minds that we have a capacity to know that *p*, even if for many substitutions for *p*, individual knowing subjects are not or cannot be *in a position* to know that *p*. It's also part of an ordinary conception of the nature of our minds that we can have mere beliefs, as when one adds up sums incorrectly, or jumps to a conclusion, thus precluding knowledge in oneself (whilst one continues to maintain the mere belief on the relevant bad epistemic grounds). Knowledge and mere belief seem to have a common source, in term of the power(s) of thought of which they are manifestations. When one adds up sums correctly and comes to know, and when one adds up sums incorrectly and comes to merely believe, it certainly does not seem as if one suddenly switches from employing one set of powers (the cognitive ones) to another set of powers (the ones liable to error). What's more, our cognitive powers seem normatively-oriented towards knowing rather than merely believing: in *every* case, knowledge seems like a cognitive *success*, and mere belief seems like a cognitive *failure*.[100] These features of our minds don't seem like mere accidents of our nature as minded beings; they are pervasive in our mental lives, and central to our mental lives. To that extent they seem essential to what it is to have a mind like ours. Call these features collectively 'the appearances'. If we accept the appearances, then a capacity to know, a liability to merely believe, a connection between these two in terms of some common source, and some kind of (apparently)

[100] Of course, one might not care that one has failed, cognitively speaking, and sometimes it might be right not to care. The present point concerns the sense in which our cognitive powers are oriented towards knowledge regardless of our (contingent) cares and obligations. This marks a point of continuity with the concerns of direction of fit theorists in Chapter 1.

normative orientation of the common source towards knowing must appear in an adequate account of the nature of the mind.

It will help to give names to some of the key elements of the appearances:

Capacity	The capacity to know is essential to minds like ours.
Unity	Our powers of knowing and merely believing have a common source.
Privilege	Our powers of knowing and merely believing are normatively-oriented towards knowledge as cognitive success.

Let's go back to *SANE*. I would like to substitute the fallible capacity to know for *Y* and say that *that* is essential to minds like ours. Substituting the fallible capacity to know for *Y* is my way of doing justice to the appearances. Here is the proposed application of *SANE* with the appropriate substitution:

Kim's version of SANE, for Bumbling Epistemic Agents

[or ***Kim-SANE***, for short]

P3 The fallible capacity to know is essential to having a mind (like ours).

P4 A true and adequate statement of the essence (or metaphysically fundamental nature) of the fallible capacity to know must use normative terms.

C2 So norms are essential to having a mind (like ours).

Kim-SANE is intended as just one example of a general argumentative strategy. The overarching theme of the strategy is this: if the potential for mistakes (whether epistemic mistakes or not) is essential to minds like ours then the Normativist Claim is true. Knowledge is here used as an

example, but other substitutions, such as representation, or self-movement, or non-accidentally true belief (supposing that is different from knowledge) could work too. Because knowledge is used here as an example of a quite flexible argumentative strategy, I will for the moment ignore opponents to *Kim-SANE* who would reject the claim that the capacity to know (whether fallible or not) is essential to minds like ours. (I will return to this point in §2.4.3.1.)

Given this flexibility, what exactly do I mean by 'minds like ours'? I mean this: *complicated* minds that have a complex internal order and structure to their mental activity, which can potentially be interfered with. On many views, God's mental activity is (blessedly) *simple*, and cannot be interfered with, so God's mind is not a mind like ours. Some less divine minds might also be (blessedly) simple. Fodor (1987) has argued for the metaphysical possibility of a 'punctate mind' whose mental life consists in manifesting one atomic representation (e.g. DOG). Although punctate minds can presumably be messed with somehow, they have no internal order or structure to their representational activity, so they are not minds like ours.[101] I leave it open how far the phrase 'minds like ours' extends when we consider simpler and simpler non-rational creatures.

I assume that *Kim-SANE* has a valid argument form. As I argued in §2.3.4.2, fallible capacities are essentially normative, so the defense of *Kim-SANE* looks to rest on P3 as the crucial premise rather than P4. In one sense that's right, and in another sense that's wrong. P3 is intended as an intuitive and ecumenical way of capturing the appearances. My account of fallible capacities provides one reading of P3. For the sake of argument I leave open the option of a reductive reading of P3, where the fallible capacity to know is explained (away) in terms of non-fallible capacities. In

[101] Incidentally, no punctate mind makes a mistake, given that all mistakes consist in inappropriate ordering of some kind of activity.

what follows my primary dialectical opponent will be just such a reductionist, who accepts the appearances, and so accepts P3, but disputes its meaning by rejecting P4.[102]

2.4.2 Prospects for Reductionism

Reductionism is my name for the view that accepts the appearances but claims to be able to explain the appearances in non-normative terms, within an explanatory framework of non-fallible capacities. (My use of the term 'capacities' here and in what follows is intended to apply to *any* way of talking about the mind's powers of thought: it covers talk of capacities, abilities, functions, bundles of dispositions etc.) The reductionist accepts P3, but offers a reductive account of what it is to possess the fallible capacity to know, and so rejects P4. I should stress that reductionism accepts *Privilege*. The reductionist merely proposes to explain *Privilege* in non-normative terms at some metaphysically fundamental level.[103]

That's a very general description of reductionism. For the sake of ease of presentation, it will help to have a more specific example. Some non-reductive physicalists count as reductionists in my sense of the term.[104] Although non-reductive physicalists reject any direct reduction of the mental to the physical, they do tend to explain the nature of the mind by reference to mental activities that have a purpose, or proper function, and they also explain proper function in non-normative terms.

---

[102] One might reject P3 outright because one thinks that it isn't essential to minds to do *anything*, let alone fall into error. I suspect that some medieval substance dualists think this: according to them, to have a mind is to be a mental substance, and *no* capacity, let alone a fallible capacity, is essential to what it is to be a mental substance. (*Caveat*: I am wary of speculating on who actually subscribed to this position given my limited knowledge of medieval philosophy of mind, so perhaps this is a straw man view.) I don't have any arguments against this kind of view: I just assume it's false. As I said before, thinking is something that minds, by their very nature, *do*. ("By their very nature" here means: according to an essence; *not* a *mere* consequence, however necessary, of something's essence.)

[103] 'Reduction' here does *not* refer to the reduction of the mental to the physical, as it usually does in philosophy of mind, but to the reduction of normative to the non-normative.

[104] For example, Millikan (1984); van Gulick (2002, 2006); Dretske (1991). Davidson (2001) doesn't count for present purposes as his views on the constitutive ideal of rationality are often read as implying the normativity of the mind. See Schoeder (2003) for a dissenting opinion on this last point.

Millikan's earlier work provides a good sketch of how to explain proper function in non-normative terms.[105] She says that norms of thought are biological norms of proper function, and sketches a reduction of proper function to historical-evolutionary explanation: a thing's proper function is something that explains why that kind of thing is here today.[106] In the present context, the idea would be that mere beliefs can be explained as failures of proper function. According to the reductionist, mere beliefs are *malfunctions*, where malfunctions can be explained in non-normative terms as states or episodes that are ultimately not conducive to the promotion of some non-normative measure that generates proper functions (e.g. evolutionary advantage, or a kind of stability, or perhaps maximal desire satisfaction). The overall reductive strategy promises to be extremely powerful, dialectically-speaking. The two-step process – of explaining mere beliefs as malfunctions, and then explaining malfunctions in non-normative terms – promises to account for the appearances at the same time as explaining why we might have been led by the appearances to think that the mind is essentially normative, when in fact it isn't.

Let me outline two problems that give reason to think that reductionism fails. The problems concern *Unity* and *Privilege,* respectively.

2.4.2.1 Problems Explaining *Unity*: a Conjunctive Account of Knowledge is Required

The reductionist accepts the appearances, so she owes us an explanation of *Unity*. The theory of fallible capacities explains the common source of knowledge and mere belief by tracing them both to the fallible capacity to know. But the reductionist is committed to reducing fallible capacities to an explanatory framework of non-fallible capacities, so she must be able to explain *Unity* within a

---

[105] Millikan (1984).

[106] That's not the only way to reduce proper function to non-normative terms. See van Gulick (2006) for a sketch of how to reduce proper function to a contribution to stability that meets certain non-normatively-specified criteria, with less of an emphasis on history. I will focus on Millikan's sketch rather than van Gulick's, as it directly exploits intuitively familiar ideas like evolutionary advantage.

framework of non-fallible capacities alone. Could she trace knowledge and mere belief to a *disjunctively*-specified capacity, such as the non-fallible capacity to know or merely believe? That won't help. Disjunctively-specified capacities do not presuppose or explain any unity of their disjuncts.[107] It seems the only halfway-plausible option for the reductionist is to posit a common denominator. For instance she could trace knowledge and mere belief to the non-fallible capacity to *believe*, and then analyze knowledge as perfect exercise of the non-fallible capacity to believe that is conjoined with extra conditions, such as presence of justification and truth of the belief (plus reliability, safety, stability, or what-have-you).

Williamson (2000) has argued against such conjunctive accounts of knowledge. Williamson's basic claim is that knowledge is a mental state, but belief plus extra conditions is not: it is rather a "metaphysical hybrid" of mental state conjoined with extra non-mental condition(s).[108] I agree with Williamson, but I won't repeat his arguments here: it is enough to note that those who aim to do without norms *must* fight Williamson as well as me. (Good luck to them, and may the best philosopher win.)

Williamson's worries aside, I think that there is a relatively clear sense in which a conjunctive account of knowledge, even if true, could not explain *Unity*. *Unity* is not just the claim that mere belief and knowledge share a common element. It is rather the claim that mere belief and knowledge share a common *source*, in terms of the power(s) of thought manifested in both. But the source of knowledge in us is not the non-fallible capacity to believe. The non-fallible capacity to believe is indifferent to the further conditions proposed by the conjunctive analysis of knowledge.

---

[107] For example, my capacity to joke or be poisoned does not have a common source. If it did, and we were trying to account for that common source, we would at very least want to know what the common denominator is, over and above the (parasitic) posit of a disjunctively-specified capacity to joke or be poisoned.
[108] The term 'metaphysical hybrid' comes from Williamson (1995).

I do not think this is a mere quibble about the meaning of the terms 'common element' and 'common source'. The proposed further conditions required for knowledge may be accidentally conjoined with belief in given cases, as Gettier (1963) brought to our attention. Gettier cases suggest that the proposed elements required for knowledge (on the simplest analysis e.g. justification and truth) must be non-accidentally unified with belief on given occasions if the would-be knower is to know, rather than to merely believe. Various proposals have been put forward as to how explain what "non-accidental unification" amounts to here, in terms of being the product of reliable dispositions, manifesting certain counterfactually robust properties, being caused "in the right way" etc.[109] Now whatever the reductionist proposes to respond to Gettier cases, in order to explain how the elements are non-accidentally unified on given occasions when a thinking subject knows that *p*, it deserves to be called a *capacity* in our (extended) sense of the term. Capacities just are the things that explain their perfect exercises as non-accidental. For capacities that unify various elements into a whole of a quite different order, this includes explaining how the elements are bound up with one another in a non-accidental fashion on a given occasion of perfect exercise of that capacity. But positing a capacity to unify-the-required-elements-so-as-to-know (or a capacity to-be-in-states-where-one-is-unified-so-as-to-know) does not explain *Unity* in terms of a common source; it rather explains why one might have (mistakenly) thought there was a common source, when knowing has one source – the non-fallible capacity to unify-the-required-elements-so-as-to-know – and mere believing has another – the non-fallible capacity to believe.[110]

Now rejecting *Unity* and explaining why one might have thought *Unity* was true is of course an option for the opponent to *Kim-SANE*. The line between explaining something in other terms

[109] See for example Goldman (1967); Dretske (1989); Nozick (1981).

[110] The alternative is to just accept that whatever qualifies a belief as knowledge does not deserve to be called a capacity in our (extended) sense of the term. According to this position, it is an accident, as far as or powers of thought are concerned, whether our beliefs are blessed with the (honorific) title of 'knowledge' on individual occasions or not. I have not been able to determine whether anyone in the existing literature unblushingly subscribes to this position. As far as I can tell, there is nothing to distinguish this position from skepticism.

and explaining it *away* is often not very clear when considering various self-styled "reductive" proposals in philosophy of mind. But I think that taking this option would make a real (metaphysical) mystery of why mere beliefs – just as such – are cognitive errors, given that they are exercises of a capacity that is – just as such – indifferent to cognitive error. In any case, in the present context, reductionists in my sense of the term aim to explain the appearances, not to explain them away: the appearances form a real constraint on what the reductionist says, the way that the phenomena of chemistry form a real constraint on proposed reductions of chemistry to fundamental physics.[111] So the conjunctive analysis of knowledge is bad news for the reductionist: it makes it thoroughly mysterious how we could make sense of knowledge and mere belief as having a common source, rather than being (more or less distantly) related by common elements.[112]

2.4.2.2 Problems Reducing Privilege: Extensional Adequacy

The reductionist owes us an account of *Privilege*. Now the norms at work in *Privilege* are discrete and perfectly general: every case of knowing is a cognitive success, every mere belief a cognitive failure. There isn't a middle ground with regard to the judgment of cognitive success or failure. Put in terms of my own vocabulary: a particular exercise of the fallible capacity to know, whether perfect or imperfect, may well register on some *other* normative measure that allows partial degrees of fulfillment, but the standards of correctness set by fallible capacities themselves don't allow any intermediate stages between divine perfection and abysmal imperfection. Relative to *this* norm, there

---

[111] In terms of real constraint on reduction in philosophy of mind, mental causation is a good example of something that is not treated as a real constraint. Most folks think that chemical reactions are perfectly real: that is why the phenomena of chemistry form a real constraint on proposed reductions of chemistry to fundamental physics. But in philosophy of mind, reductionists about mental causation typically think it is a *myth*. The primary goal then is not to explain what mental causation *is*, but to explain it away.

[112] By way of analogy: consider the appearances that raw opium and heroin are coming into the USA in large quantities and seem to have a common source. Suppose one is told that the opium is supplied by warlords from Afghanistan, but in certain circumstances Colombian drug-lords buy choice shipments of opium from the warlords, refine it into heroin in Colombia, and then smuggle it into the USA. This is not an explanation of the appearance of a common source; it is at best an explanation of why the appearance is misleading (perhaps because of the common element).

are no prizes for second (third, *n*th) place, even if one has e.g. partial evidence for a true belief that doesn't constitute knowledge.

By contrast, the norms of proper function to which the reductionist typically appeals do admit of partial degrees of fulfillment. A heart can pump blood more or less well; a memory system can function more or less well; a belief can contribute more or less to flourishing or desire satisfaction or an overall world-view. If the norms characteristic of knowledge and mere belief are to be reduced to norms of proper function, the reductionist needs to derive extensionally-adequate sharp dividing lines from norms of proper function, so that the malfunctions she identifies will capture all and only the mere beliefs. It's very hard to see how any such derivation would go.

To derive the requisite sharp dividing lines, the reductionist will have to say that mere beliefs are exercises of a capacity that have a position below a certain threshold on the relevant non-normatively-specified measure that admits of many degrees of partial fulfillment. Presumably, given the requirement to account for *Unity* within an explanatory framework of non-fallible capacities alone, the capacity here is the non-fallible capacity to *believe*. If it could be carried out, securing the claim of co-reference between mere beliefs and exercises of the non-fallible capacity to believe that figure below the relevant threshold would be the first step in an adequate and respectable reduction. But I think this first step can't be carried out. It isn't difficult to come up with particular examples where mere belief is more conducive to promoting some non-normative measure than knowledge is. Sometimes it serves our goals and purposes very well to believe something false, or to believe something for bad reasons. The same goes for evolutionary advantage. There are even identifiable stable classes of such exceptions.[113] So the reductionist will misclassify many mere beliefs as cases of

[113] Here one might consider false or badly grounded beliefs about the Gods, or about the Neighbors, that serve to unite a community, where systematic mere belief confers an advantage on members of that community that knowledge would not confer. It could even be quite normal for a species of animal to represent something as basic as *food* very inaccurately, so as to save time and resources that would be wasted in pursuit of more complete, reliable or accurate

proper function, rather than malfunction, because, as a matter of fact, they have a position *above* the relevant threshold on the relevant non-normatively-specified measure that admits of many degrees of partial fulfillment.

One might object that this is only true of impossibly crude reductionist accounts. Evolutionary advantage (or desire satisfaction, or whatever) is the *ultimate* non-normative measure that generates proper functions. But proper function is supposed to be distinct from the ultimate measure, and it is surely possible that on a suitably sophisticated account our capacity to believe could have e.g. the proper function of issuing in knowledge, even if in many cases mere belief would better serve the relevant end. The first thing to note about this kind of response is that it doesn't touch the underlying problem about how to derive the requisite sharp dividing lines from a measure that admits of degrees of partial fulfillment; it just assumes that this can be done. The second thing to note is that in the present context, the claim that knowledge is the proper function of our capacity to believe is an *empirical* claim, and one whose best support seems to come from armchair-bound *a priori* speculation.[114]

Comfortably nestled in my own armchair, I would like to offer the following (bold) counter-hypothesis. Nothing dictates that a true and adequate statement of proper function must have a logically simple form. Evolution has no special fetish for logical simplicity: most aspects, parts and subsystems of a creature have evolved to serve many functions, rather than just one function specifiable in a logically simple term. Given that knowledge helps in some cases and not others, a

---

representation. (Great White sharks apparently eat most anything that is in front of them: no need for accurate representation there. The same goes for us and fast or genetically modified food. If we had to *know* what we eat in any but the barest detail, we would soon perish.)

[114] Consider for example Velleman (2000) who claims that belief aims at truth, as it has been designed to do so by evolution. Velleman never bothers to consider the most obvious rejoinder to this empirical hypothesis: that belief aims at truth only in certain conditions, and at falsity in others. Or consider Papineau (1993) who says that belief has the normal function of being true, even though it may have the special function of being false. Papineau never tries to specify how one justifies the claim that being true is the normal function, or what the difference between normal and special functions amounts to.

more plausible empirical hypothesis is that the proper function of our capacity to believe is to issue in knowledge in those cases where knowledge helps *and* to issue in mere belief in those cases where mere belief helps. Perhaps it makes sense for evolution to produce a crude mechanism of belief that only has one goal in the short term. But over time, we should expect a mechanism initially designed to aim at knowledge to *develop* and become sensitive to those stable cases where mere belief would better serve the end of e.g. reproductive success. So it just isn't true that our capacity to believe is simply oriented towards knowledge, in the non-normative sense of orientation to which the reductionist appeals. And so the reductionist will still end up misclassifying some mere beliefs as cases of proper function.

A variation on the last objection draws on parallels between the present logical difficulties and those that emerge from the literature on the value of knowledge. Pritchard (2007) describes primary, secondary and tertiary value problems that emerge from considering the value of knowledge. The primary problem is to account for why knowledge is more valuable than mere true belief. The secondary value problem is to account for why knowledge is more valuable than any epistemic standing that falls short of knowledge. The tertiary value problem is described as follows:

> … one could … argue for a *tertiary* value problem which demands that the difference in value between knowledge and that which falls short of knowledge must be one of *kind* and not merely degree. The rationale for the tertiary value problem is that a response to the secondary value problem leaves it open whether the difference in value between knowledge and that which falls short of knowledge is merely a matter of degree. If the difference is merely one of degree, however, then this leaves it unclear why it is *knowledge*, specifically, that is of distinctive value to us. That is, why is it the point on the continuum of epistemic value that knowledge marks that is of special interest to us?[115]

A further challenge to the reductionist could then be put in the form of Pritchard's final question, where "special interest to us" is replaced with "special importance relative to the reductionist's non-normative measure". The problem would not be to justify the claim that specifications of the proper function of belief must be logically simple, but to justify the claim that *knowledge* is the proper

[115] Pritchard (2007:105n5).

function of belief rather than, say, maximally justified belief relative to cost-benefit analyses of time and resources spent inquiring. I can see no empirical grounds of the kind to which the reductionist appeals that could justify such a claim.[116]

2.4.3 Reasons for Rejecting P3

The problems outlined in §2.4.2.1 and §2.4.2.2 give us good reason to think that reductionism fails. One cannot accept the appearances and offer a satisfactory account of the fallible capacity to know made out solely within an explanatory framework of non-fallible capacities. Now I should stress that I expect that most opponents to *Kim-SANE* would simply accept that no reduction of the kind *I* am imagining could possibly succeed. They will accept this because they also reject P3. Let me consider what I take to be the two main reasons for rejecting P3.

2.4.3.1 Rejecting P3 on the Basic of Rejecting Capacity

One might reject P3 because one thinks it sets the bar for mindedness too high. The capacity to know may be an idealistic high standard imposed on us by dead Germans; perhaps the power to believe things that are non-accidentally true is a more appropriate standard for something that is essential to (contemporary Anglo-American) minds like ours. Or perhaps even that is too high: perhaps the power to represent a bare aspect the world is a more realistic standard for identifying something essential to minds like ours. It doesn't really matter where we choose the set the bar for mindedness, because *Kim-SANE* can be reformulated to accommodate the "new appearances".

---

[116] It may be noted in passing that some questions in the literature on the value of knowledge may mistake "value to us" for the kind of abstract normativity that is characteristic of a fallible capacity, and that this is the source of some of the puzzlement in the area. I will not develop this thought here.

All the plausible proposed substitutes have what seem to be deformed instances. Contrasting to the case of non-accidentally believing the truth is the case of accidentally believing the truth, or believing the false. Contrasting to the case of representation is the case of misrepresentation. As with knowledge and mere belief, the new appearances will be that (a) we can exercise the relevant power(s) successfully; (b) we can exercise the relevant power(s) unsuccessfully (c) the successful and unsuccessful exercises have some common source; and (d) there is a normative orientation of the power(s) towards one kind of exercise (the successful kind) rather than the other. We can then construct a new version of *Kim-SANE* that substitutes the fallible capacity to believe the truth, or to represent (or whatever) in place of the fallible capacity to know to capture the new appearances. Corresponding to *Capacity*, *Unity*, and *Privilege* will be *Capacity**, *Unity**, and *Privilege**, which are formed by substitution of the appropriate perfect and imperfect exercises (e.g. representation and misrepresentation) for knowledge and mere belief.

I think that the prospects for reduction are no better once the bar for mindedness has been lowered. The problems raised against reduction of *Unity* and *Privilege* are quite general and do not depend on anything specific to the nature of knowledge or mere belief. Even Williamson's basic point could be transformed to apply to the fallible capacity to represent, for example, if we allow that representing (successfully) is a mental state, and tokening a mental content conjoined with extra conditions is not.

One might think that there is a damaging disanalogy that blocks generalization of *Kim-SANE*, because knowledge implies rationality, and simpler substitutions don't seem to require rationality (that's part of the reason why e.g. representation lowers the bar for mindedness). But the arguments against reductionism don't exploit any particular claims about the nature of rationality, or how it relates knowledge and mere belief. All that is required for the style of argument to go through is a certain unity of cognitive activity that applies at the level of the whole being. This is just the kind

of unity according to which it makes sense e.g. to say that the whole being is representing or misrepresenting some aspect of its environment, but not doing both at the same time in the same way in the same regard. Now facts about rationality presumably do explain (at least partly) the conditions in which mere belief precludes knowledge and vice versa. But the coordinate account of the conditions in which misrepresenting precludes representing (and vice versa) is by-the-by: we can ignore the conditions, so long as it is granted that representing and misrepresenting *do* (sometimes) exclude each other in a way that validates the associated appearances. And this surely has to be granted for any recognizable account of what it is to be a thinking being.[117]

What I have offered here is of course a mere sketch of a generalization of *Kim-SANE*. But I think it's clear that opponents to the Normativist Claim cannot *easily* avoid the conclusion of *Kim-SANE* simply by rejecting *Capacity*. For if any mental capacity that has (apparently) deformed exercises is essential to minds like ours, then a revised version of *Kim-SANE* will give reason to think that the mind is essentially normative. And it is generally accepted that the possibility of *misrepresentation* is something that any adequate account of the mind would have to accommodate and explain.[118] Given the generality of problems with reductionism outlined in §2.4.2.1 and §2.4.2.2, any deep resistance to *Kim-SANE* will have to rest on something other than the truth of *Capacity*.

2.4.3.2 Rejecting P3 on the Basis of Non-Fallibilism about Capacities

At this stage, an opponent to *Kim-SANE* might be tempted to simply back up and insist on the generally accepted philosophical wisdom that there are only non-fallible capacities, so there can't be any fallible capacities in the sense I have described. This would be a way of rejecting P3 directly, on my reading of it: call it non-fallibilism about capacities.

---

[117] Fodor (1987) rejects this claim, but as far as I can tell he seems to be alone in doing so.
[118] See e.g. Dretske (1986, 1981, 1995); Fodor (2008); Neander (1995); Millikan (2004). For an excellent article demolishing Fodor's attempts to avoid troubles with misrepresentation see Baker (1991).

The non-fallibilist about capacities claims that there is only one kind of causal power: the non-fallible kind. Many philosophers assume this, but I haven't yet seen a really compelling argument for the claim. Appeal to parsimony won't help motivate non-fallibilism about capacities. It's true that we shouldn't multiply entities beyond necessity, but Ockham's Razor only applies here if it isn't necessary that we explain mistakes. It's also worth noting that nothing I have said turns on a distinction between mental and physical capacities. I am quite happy to say that fallible capacities are, in some sense, physical capacities, so no simple appeal to physicalism can be used to motivate non-fallibilism about capacities.[119] Nor can a simple appeal to modern conceptions of laws of nature be used to motivate non-fallibilism about capacities. There are obviously many fascinating questions about the relationship between capacities and laws of nature, where laws of nature are conceived of as equations relating quantities and distributions (over time and space) of fundamental physical forces. But laws, so conceived, do not obviously attribute causal powers to *anything*. The question of whether fallible capacities are compatible with such laws is posterior to the question of whether *any* capacity is compatible with such laws.

Non-fallibilism about capacities is recognizable as the main source of resistance to other elements of the appearances. In §2.4.2.1 I said that there is a sense in which the reductionist who offers a conjunctive account of knowledge rejects *Unity* rather than explaining it, for instance. But the reductionist was forced to the conjunctive account *precisely* because she was committed to non-fallibilism about capacities.[120] Similarly, we find resistance to *Privilege* that is recognizably based on

---

[119] It is for this reason that it doesn't matter if the reductive naturalist proposes to reduce mental fallible capacities to some set of physical fallible capacities. If there are physical fallible capacities, then they are essentially normative, because the normative term 'valence' or 'privilege' must be used in a true and adequate statement of their essence. The lines of debate here cut across more traditional lines of debate between physicalists and their opponents.

[120] Rejecting *Unity* is, I think, an extremely popular position, for without the conceptual resources made possible by the concept of a fallible capacity, it is very difficult to see how *Unity* could possibly be true. This point is forcefully made by Rödl (2007). It is because philosophers do not realize that fallible capacities provide a conceptual tool that is different from a merely-disjunctive capacity that McDowell, for instance, is usually classified as a disjunctivist, when his own view is closer to present proposal than might otherwise be apparent. See McDowell (2011) for an endorsement of the claim

commitment to non-fallibilism about capacities. Consider Dretske's remarks about the normativity implicit in deformed instances of mental activity:

> The only *fault* with fallacious reasoning, the only thing *wrong* or *bad* about mistaken judgments is that we don't like them… This, though, leaves the normativity of false belief and fallacious reasoning in the same place as the normativity of foul weather and bad table manners – in the attitudes, purposes, and beliefs of the people who make judgments about the weather and table behavior.[121]

Dretske's remarks here are, I take it, representative of a fairly common attitude to norms in contemporary philosophy of mind. The attitude is this: norms of any kind must be founded on the intentions and desires of individual agents in some way.[122] But Dretske is not like Aristotle in thinking that we all desire to know; treating norms as founded on desire is a way of rejecting the generality of *Privilege*. In adopting this attitude Dretske ignores the concerns of e.g. the direction of fit theorists from Chapter 1, who aimed to account for the perfectly general normativity apparently present in the operation of our capacity for belief, one of whose marks was *precisely* the way it could come apart from contingent individual or collective desire or endorsement. I think we should take a short line with Dretske's kind of (flippant?) objection to the idea that there is some perfectly general norm at work in the case of knowledge and mere belief, such that knowledge is a cognitive success and mere belief is a cognitive failure. The view that the only kind of normativity is that based on desire or intention is only plausible if one takes for granted that there are no fallible capacities, in the sense I have described. It is on this ground that the normativist ought to make her stand, at the level of claims about the metaphysics of capacities, and the possibility of a capacity that allows normatively-imperfect exercises.

---

that perception is an exercise of a fallible capacity for knowledge. McDowell's characterization of fallible capacities is more abstract than mine: for him, what I have called a processual capacity would fall into the same class (i.e. capacities that have non-determinative exercises).

[121] Dretske (2000: 248).

[122] If the view of intentional action proposed in Chapter 3 is correct, then intentions *presuppose* the more abstract normativity characteristic of fallible capacities.

Now I don't have a direct response to non-fallibilism about capacities. I don't myself have an argument for the modally-strong claim that there *must be* fallible capacities. There is extra work to be done here. But I have already shown how the appearances give us reason to think that there *are* fallible capacities, because the appearances can't be accounted for within an explanatory framework of non-fallible capacities alone. And I think there is a little more to say on the issue. Non-fallibilism about capacities can only give us an alienated conception of the causal powers of the thinking being.

All malfunctions of things that possess only non-fallible capacities are attributable to prevention or (external or internal) interference.[123] But as I have described them, no mistake is attributable to prevention or interference. That's the *point* of mistakes: they are a third category of failure, distinct from both prevention and interference. So mistakes cannot be *accommodated* by the non-fallibilist about capacities. They can only be eliminated.

Eliminating mistakes from one's ontology flies in the face of what I described at the opening of this chapter as one of the most familiar aspects of our mental lives. I think the best the non-fallibilist about capacities can do to accommodate our experience of error is to say that the concept of a mistake, as I have described it, is a more or less defective folk concept that picks out a range of cases of internal interference by cognitive subsystems. This is the sort of route that an analytic functionalist might take, when puzzled as to where to find something in the system of non-fallible capacities (the realm of causal law, narrowly construed) that realizes the platitudes enshrined in the folk concept of a mistake. The relevant cases of internal interference by cognitive subsystems will presumably be distinguished from cases like the hiccup by the fact that the cognitive subsystems

---

[123] Malfunctions of things that possess only non-fallible capacities are usually cases of loss of capacity, as when some part snaps, splits, melts or otherwise loses its form, or else they are cases of functioning poorly (e.g. doing the same thing, but slower, or less accurately, or less effectively). Losses of capacity are attributable to external or internal interference: e.g. excessive heat melted the capacitor; excessive pressure snapped the bridge support; etc. Poor function is attributable to external or internal interference or being in inappropriate conditions of exercise for the relevant capacity: e.g. the computer loads slowly because weighed down with malware; the drill drills slowly because of intermittent electrical current and an excess of rust, etc.

involved are normally enlisted in the service of some task, like catching a ball or producing knowledge, and hiccups (and so on) are not. But apart from being normally enlisted in the service of some task, the relevant cases of internal interference that (supposedly) roughly correspond to the folk concept of a mistake will be *just* like hiccups.

I am doubtful about an appeal to internal interference in this context. In one innocuous sense, to say that one cognitive subsystem coordinates badly with another is a way of explaining a mistake. We often trace the source of a mistake to inattentiveness or overexcitement, for example. But at the level of the whole being, to say that one's attention system or excitement system interferes with one's powers, or prevents one from being in the conditions of exercise for one's powers, is a bit like a craftsman blaming her tools. If mistakes are a real phenomenon, as I think they are, then an appeal to interference is inappropriate when we explain the cat's mistake when it sizes up the jump it can make, wiggles its tail, leaps, and falls short. To think that the cat's powers have been *interfered* with by its cognitive subsystems in such a case risks conceiving of the cat as a kind of God imprisoned in a mechanical body, such that if it could only be free of the relevant interference and limitation, it would do what it set its mind to do. This is an alienated picture of the causal powers of the cat, where its *way* of doing something interferes with its (imagined?) power to do something. Similarly for our cognitive powers: to suppose that, whilst a knowing subject is at her leisure, her attention system (for instance) *interferes* with her capacity to know is to think of the knowing subject as a kind of God, imprisoned in a mechanical mind, such that if she could only be free of the relevant interference and limitation, she would know rather than merely believe. This is an alienated picture of the causal powers of the knowing subject, where her *way* of doing something interferes with her (imagined?) power to do something.

Non-fallibilists about capacities think that, when it comes down to it, no one really makes a mistake in the sense I have described. It must be admitted that the claim that no one makes mistakes

has a lot going for it, rhetorically speaking: consider how it can be exploited by religious gurus or motivational speakers. But I just don't think it's true. Of course, what I think is plainly true not going to impress an opponent to *Kim-SANE*, but I think it does shift the burden of proof onto the right issues, and away from the ones that made Georges Rey cry foul against Wedgwood. Now that we normativists have a conceptual tool with enough structure to it to support the Normativist Claim without begging the *wrong* questions against naturalists like Georges Rey and company, we can get on with the more important project of finding ways to stop begging the *right* questions against them.

2.5 CONCLUSION

Kim-SANE is based (rather unblushingly) on the appearances, but it does offer a simple and elegant explanation of how it is possible to make mistakes in one's mental life. It is worth noting that *this* question has, in one form or another, been the bugbear of the dominant positions in philosophy of mind for at least the last few decades.[124] The phenomenon of misrepresentation has persistently resisted satisfactory analysis within the confines of an explanatory framework of non-fallible capacities. I offer the following (predictable) diagnosis: the problem is with commitment to non-fallibilism about capacities, and a poverty of explanatory resources.

My aim in the chapter was to argue for the essential normativity of the mind, but there are some recognizable ways in which I have fallen short of that goal. In particular, although fallible capacities explain how mistakes are possible, I have not addressed the legitimation project of showing how fallible capacities *themselves* are (metaphysically) possible. Part of the difficulty is that

[124] See note 59 above for relevant references. Millikan (1995) considers defusing the problem of accounting for misrepresentation by broadening the conception of the bearer of the relevant capacity to the "head-world" system. See McDowell (2004) for a response to Millikan that brings out the source of her (despairing? certainly peculiar) move as commitment to what I have been calling non-fallibilism about capacities.

the concept of a fallible capacity has not received a lot of attention in the existing literature, so it is difficult to know precisely how to proceed, and which metaphysical issues to address first.[125] But I think the preceding discussion is sufficient to justify the claim that there is room for a normativist metaphysics of mind, of a kind that can take the phenomenon of error in its stride, and that the legitimation project is worth trying, albeit on another occasion.

2.6 PUZZLES SOLVED AND UNSOLVED

The account of fallible capacities offered in §2.3.2 above solves the first of the puzzles left over from Chapter 1. It is hard to see how there could even be an essentially normative power of thought. But once one broadens one's horizons to include fallible as well as non-fallible capacities, one can see both how an essentially normative power of thought might be possible and what its essential normativity would consist in.

The account of fallible capacities also offers a solution to the second of the puzzles left over from Chapter 1, though perhaps not a very satisfying one. What mistaken beliefs and mistakes in performance have in common is that they are both normatively-imperfect exercises of fallible capacities. Where they differ is in the particular fallible capacity of which they are a normatively-imperfect exercise. But no account has been given of how to individuate capacities, beyond noting that not *every* difference in capacity-specification corresponds to a difference in capacity. In this respect, I think we must look to the phenomena for the relevant differences, and in Chapter 3, I offer an extended meditation on the nature of the fallible capacity to do what one has in mind to do, by way of starting in on this (large) project on the practical side.

---

[125] To my knowledge, Rödl (2007) and McDowell (2011) are the only ones to have done so.

What of the third puzzle left over from Chapter 1? The third puzzle was whether the standard of correctness for beliefs is truth, just as such, or something more robust, like knowledge, and similarly for telic events and success. In closing I would like to offer a brief argument for the claim that the standards of correctness set by fallible capacities do not allow for what I described in Chapter 1 as "external accord".

The standards of correctness set by fallible capacities do not allow external accord because capacities explain their perfect exercises as non-accidental. Consider the fallible capacity to know. There is no such thing as knowing *by accident*, relative to the fallible capacity to know. One may of course discover an unexpected truth, and in that sense one may come to know something by accident, but that is a quite different sense of 'accident'. If one has taken up a definite epistemic stance with regard to the truth of p, but the stance one has taken is merely accidentally united with the truth, then one precludes knowledge in oneself whilst believing the truth on the relevant (bad) epistemic grounds. The standard of correctness has not been met.

Knowledge is a high standard for cognitive activity, and might be thought to be a special case. What about apparently less demanding capacities, like the fallible capacity to believe the truth? Isn't it plainly possible to believe the truth by accident relative to this capacity, as when I believe that the lottery numbers will be 4, 8, 15, 16, 23, 42 and the numbers turn out to be that way (and no one cheats)? Hasn't the standard of correctness been met, albeit merely accidentally? The same point applies. There *is* such a thing as believing the truth by accident relative to the bare (fallible?) capacity to *believe*, because that capacity is indifferent to the truth of what is believed. But in its perfect exercises, the imagined fallible capacity to believe the truth unites elements in the non-accidental way discussed in §2.4.2.1. We might flag this with a convention of ugly hyphenation, saying that it is really the fallible capacity to believe-the-truth, and not the fallible capacity to believe (i.e. to believe something that might *happen* to be true). Or we could make up a new word for the relevant unity of

elements. Regardless, the basic point stands: there is no such thing as believing-the-truth by accident, relative to the fallible capacity to believe-the-truth. When one believes that the lottery numbers will be 4, 8, 15, 16, 23, 42 it might appear that one is perfectly exercising the fallible capacity to believe-the-truth. But the appearance is deceptive: one is actually precluding perfect exercise of that capacity in oneself on that occasion. Given that the standards of correctness set by fallible capacities do not allow external accord, there is a sense in which beliefs and telic events aim at (something like) knowledge, in the sense of non-accidental unity of what one has in mind with what is the case.

3.0 THE ANTINOMY OF BASIC ACTION

3.1 INTRODUCTION

It can seem mere common sense to suppose that there must be some things that agents simply *do* intentionally, without the mediation of doing anything else intentionally. Call these basic actions. An orthodox position in philosophy of action accepts that every intentional action decomposes into a set of basic actions: basic actions are the "practical atoms" out of which all the more interesting (more complex) intentional actions are made. Michael Thompson has constructed a powerful argument against the very idea of basic actions, based on little more than common sense considerations about the continuity of time. The disagreement between the orthodox view and Thompson's has the look and feel of an antinomy: what I call the *antinomy of basic action*.

I have sympathy for both poles of the antinomy, but they cannot be true together. I dispute the claims of necessity. It is neither the case that there *must* be basic actions in every case of intentional action, nor is it the case that there *must* not be. Neither necessity holds because (for the most part, and within certain constraints) it is up to individual agents themselves to settle how to carve up the instrumental joints of what they have in mind to do.

The chapter has the following structure. In §3.2 I summarize the orthodox position of basic action theorists, and Thompson's argument against the very idea of basic actions, and outline why neither position is, as it stands, an attractive one. In §3.3 I argue against Thompson's dialectically stronger position. Over the course of the argument a positive view of the nature of intentional

action emerges, where the individual agent's practical conception of what she is doing is of prime importance. In §3.4 I consider one key respect in which the positive view is best accommodated by the theory of fallible capacities outlined in Chapter 2.

3.2 THE ANTINOMY OF BASIC ACTION

3.2.1 Basic Action and Practical Atomism

At what point does the will get involved in changing the world? If one takes the wording of the question seriously, and believes that the will gets involved in changing the world *immediately*, at some *point*, then one believes in basic actions.[126]

Basic action theories take many forms, but they all agree that there are some actions (or for Davidsonians, descriptions of actions) that are *immediate* in some special way. Theories are individuated by their account of the kind of mediation in question. Some say the mediation is causal (basic actions *cause* non-basic actions, and are themselves not caused by any other actions); others say it is instrumental (basic actions are compositional or constitutive *means* to non-basic actions, and are themselves not performed by any other means); yet others say it is psychological (basic actions are e.g. *thought of* as means to non-basic actions, and are themselves not sustained by any thought of distinct means). Theories also differentiate themselves in non-essential ways by making further claims about how to recognize basic actions in the wild, independently of their definition: it is

[126] Throughout this paper the interest is in the *intentional* actions of thinking subjects, rather than the actions of pistons, slugs and acids. I will often omit the qualifier 'intentionally'. The reader may fill in the word 'intentionally' where required.

variously claimed that they are volitions, or neural events, or bodily movements, or (graceful skilful) things done 'all in one go'.[127]

Basic action theorists are united in thinking that the will *must* get involved in changing the world in or through something *immediate*: the mediated, non-basic actions cannot come to be except via the mediation of some basic (immediate) action(s). The reason for this necessity is often left quite opaque but is usually supposed to have something to do with a vicious regress suited to the kind of mediation in question. As Danto claims, it couldn't be that *every* action is caused by another action of the same agent. Or as Davidson claims, it couldn't be that *every* true description of an action under which it is intentional makes reference to a contingently achieved effect of the action. Or as Hornsby claims, it couldn't be that an agent has an *infinite* number of beliefs about her means to her ends, or has an *infinite* number of 'items of knowledge' about how to achieve her end. To the extent that the viciousness of the regresses is left implicit, as it usually is, the question of exactly *why* causes, or thought of contingent effects, or beliefs about means, or 'items of knowledge', can't go on forever is left to the reader's imagination. (Perhaps the reason is too obvious for words.)

The traditional basic action theorist is a kind of *practical atomist*. Practical atomists believe that all non-basic actions, which pre-theoretically encompass all the *interesting* actions in which humans engage, must be e.g. caused by, or constituted by, or composed of (or otherwise essentially mediated by) basic actions.[128] If true, this promises a certain simplification of a philosophical account of agency. For a practical atomist, practical *thought* of non-basic action may be supposed to mirror the metaphysics of the case, whereby non-basic "molecules" are grounded in relations with, or between, basic "atoms". When an agent does *A* by doing *B* (intentionally), and *B* is a basic action, the agent

---

[127] Examples of basic action theories include Danto (1965), Goldman (1970); Davidson (2001); Hornsby (1980); Ginet (1990).

[128] Or for Davidsonians: non-basic action *descriptions*, usually of contingently achieved effects of what is done, are essentially mediated by basic action descriptions

thinks that, for example, doing *B* will *cause A* (mirroring causation) or that doing *B is* doing *A* (mirroring constitution) or that doing *B is a part of* doing *A* (mirroring composition). The simplification is this: under the supposition that all basic actions are intentional, once we have an account of the metaphysically fundamental relationship between an agent and her basic actions, then the non-basic actions will fall into place with some minimal theory-jiggling, exploiting the kind of thoughts and relations described (and perhaps some extra conditions of knowledge, skill or reliability) to extend the status of intentional action from the basic to the non-basic case.

I do not think it is an overstatement to say that belief in basic action, and the kind of practical atomism that usually accompanies it, represents an orthodox position in contemporary philosophy of action. Given this, it would be remarkable (and of some philosophical interest) if a simple argument could show that belief in basic action is deeply mistaken, and so that the metaphysical foundations of much contemporary philosophy of action are rotten. I shall present an argument that purports to do that in just a moment.

In what follows, unless otherwise noted, I shall assume that an *instrumental* conception of basic action is the most promising (and will refer to *instrumentally* basic action by the term "basic action"). Basic actions are things an agent does intentionally without doing them *by* doing anything else intentionally as means to that end. The assumption is made partly for ease of exposition, but also because many philosophers treat instrumentally basic action as the most promising kind, because thought of means to ends promises to unite the various kinds of mediation appealed to (causation, composition etc.) under a common heading. I won't try to sketch how such a project of unification might go and I will make no further assumptions about how to recognize basic actions in the wild independently of their definition.

3.2.2 Thompson's Argument against Basic Action

Michael Thompson has a deceptively simple argument against the very idea of basic action that relies on little more than an appeal to the continuity of time.[129] Take an intentional movement from *A* to *C* as a representative example of a supposedly basic action that takes some time to complete. (In order to address Thompson on his strongest ground, we may concede this restriction of our choice of basic action to those that take time to complete, even though some actions don't obviously take time to complete. Certainly actions that take time to complete are the *central* cases.) Because the action takes time, it must have parts that also take time. These parts are (or could intelligibly be) *rationalized* by the whole. That is to say, one could intelligibly ask of someone moving from *A* to *C* "Why are you moving from *A* to *B*?", where *B* is a point that is halfway on the way to *C*, provoking a rationalization such as "Because I am moving from *A* to *C*". If one can appropriately ask this question and receive an intelligible rationalizing reply, then one could ask the same kind of question of the movement halfway to *B*, provoking a similar rationalization, whether in terms of moving to *B*, or *C*, or some further point beyond *C*. Thompson notes that in the minimal cases such questions would be rather strange "conversationally speaking" but could be seen as less strange under the supposition that the tiny movement is all of the action that the questioner can see. Given that the questions and answers are intelligible, Thompson's conclusion is that we have no reason to deny that the parts discussed are intentional actions in their own right. After all, they are distinct from the action of which they are a part, yet they serve the whole as compositional means, and they are rationalized in just the same way that more obvious "macro-cases" of compositional means are

[129] Why continuity of time, rather than continuity of space? I think Thompson's argument can be read as following Aristotle's practice of using movement as the clearest example of determinate change. All that is required for Thompson's style of argument is an intentional action that is a determinate change that occurs over time, and the change need not be change in location for the parts to be rationalizable in terms of the whole. (This is perhaps what sapient chameleons might say to each other in the spirit of Thompson's argument: "I say old chap, why are you changing from blue to green?"; "Because I am changing from blue to yellow.")

rationalized, and they too can serve as rational ground for still smaller parts. So there are further rational grounds, all the way down, and the supposedly "basic" action is not so basic after all. Whenever one performs an intentional action one performs an *infinite* number of component intentional actions. Another way to state the same conclusion: there are no basic actions, only *more basic* ones.

3.2.3 The Antinomy of Basic Action

We seem to have wandered into the middle of an antinomy: what we might call the *antinomy of basic action*. On the one hand we have the basic action theorists with their (often inchoate)[130] horror of the infinite, and on the other we have Thompson with his (barely repressed)[131] disdain for the immediate. Yet if there is something peculiar in the idea that the will engages with the world "all of a sudden" in a bang, or at an immediate point, there is also something peculiar in the idea that *any* change that is formally subsumed by a change the agent has in mind to make is *thereby* a bona-fide intentional action in its own right. As will become apparent, I have qualified sympathy for both positions: I think there is something right, and something wrong, with both poles of the antinomy. Let me make some brief remarks about what is wrong with the poles of the antinomy, by way of motivating subsequent discussion.

---

[130] Danto (1965) is representative in this regard. He suggests that it is absurd to think of the agent as *always* having to do an infinite number of things first before they do the thing they have in mind to do, but that is all he has to say about the matter. One wonders how he would deal with Zeno's paradoxes.

[131] As Thompson (2008) puts it, he is tempted to adopt the manner of Quine and declare himself deep amongst the "don't cares" when considering the limit that supposedly marks off basic (immediate) action from non-basic (mediated) action.

Traditional basic action theory seems to depend on a thesis about the limitations of practical thought.[132] The practical atomist does not merely claim that many cases of intentional action do decompose into a set of basic actions, but that every case of intentional action *must* decompose into a set of basic actions. What could explain the necessity here? Suppose I hear of the practical atomists' claim, and desire to perform a metaphysical feat to prove them wrong: I will straighten my finger by means of an infinite number of compositional stages, and the stages will themselves be performed by means of an infinite number of further compositional stages, and so on *ad infinitum*. I will perform an infinite number of intentional actions, none of them basic, in under a second (or so the circus advertisement reads).[133] By virtue of what will the practical atomist deny that I have carried out my metaphysical feat? She does not deny that determinate changes, such as a change from having a bent finger to having a straight one, are continuously divisible into infinitely many parts. She also accepts that agents can turn anything they know about to their purposes, so long as what they know is (causally or constitutively or compositionally) relevant to those purposes.[134] It seems that the only plausible explanation for the limitation must come from some claim about limitations on powers of practical thought: agents (like us) just *can't* encompass that many means to an end in practical thought, even if we have (silly, philosophical) reason to.

Although there are many senses in which we are limited as thinkers, I do not think that this supposed limitation on practical thought is one of them. Take proof by mathematical induction as a (less silly) case in point. It is possible for me to prove something by mathematical induction "in my

---

[132] Hornsby (1980) says that an agent cannot have an infinite number of 'items of knowledge' about how she does something, for instance. She also claims that beliefs about means to ends just run out at some stage. But she doesn't tell us why either claim is supposed to be true, or why the associated regresses are vicious.

[133] Of course I will not pause between the stages, because physiology prevents me, but that is beside the point. The feat is not that impressive when you actually see it performed. There are no refunds at the philosophical circus.

[134] Even causal deviance can be legitimated once it is known about. Consider the traditional case of external causal deviance, where I aim at my enemy with a gun, miss, but make enough noise to stampede a herd of elephants that obligingly trample my enemy to death. Now that I know about them, I can enlist the elephants on my side when I aim to do in the next enemy. The same is true of Davidson's mountaineer, once he comes to know about the physiological reactions that can make him let go of a rope when he desires to let go of a rope.

head" without writing anything down. In doing so it seems that I have thought of an infinite number of objects in the infinite series that my proof exploits.[135] But not only that: the point of encompassing the members of the series is *instrumental*; each is a step in my proof. Here is a case in which there are an infinite number of means envisaged *and taken* to my end; precisely so, for if I missed one or more of them, my proof would be incomplete. (Another way of putting the same point; the ellipsis, as employed in written representations of proofs by mathematical induction, is not a representation of something *left out* of the proof but of something *present* in it.)

I do not pretend that this point is decisive: the nature of proof by mathematical induction, and of thoughts of infinite series, is a deep and difficult topic. But I think it is an unattractive feature of traditional basic action theory that it *requires* some (usually unmentioned, unmotivated) thesis about the limitations of practical thought. Until we are sure that there is such a limitation, and have a good idea of what kind of limitation it is, and have good reason to think that it applies in every case, we should avoid this presumption when theorizing about the nature of intentional agency.

Baier (1972: 282) provides another reason for resisting the claim that there *must*, in every case, be basic actions. She argues that we have good reason to think that there are cases where an agent does two things intentionally, each by doing the other. If possible, then neither action is basic because of their mutual instrumental dependence. Baier further argues that there are plausible cases of such mutual dependence at the purportedly basic level: her case is typing the letter 's' by making a finger movement, where one also, at the same time, makes the finger movement by typing the letter 's'. I will not pass judgment on the plausibility of Baier's case as a counterexample. Regardless of the plausibility of the particular example, it is very hard to see how to argue persuasively for the claim that no cases of mutual dependence could *ever* occur at a level that might cause trouble for the

---

[135] I haven't *imagined* or *paid special attention* to the elements of the proof, one by one, but only a Cartesian Theater picture of the mind would require one to trot elements of thought past a more or less imaginative and attentive audience.

traditional basic action theorist's claim. Although we have an intuitive grasp on means to ends, we do not have a generally accepted analysis of 'means to an end' that rules out cases of symmetric means-end dependence. Until we are sure that mutual dependence of means is *impossible* at some appropriately fundamental level of analysis of means to ends, and that every intentional action must be performed, one way or another, by means of some second action that is, so to speak, instrumentally independent of the first, we should avoid this presumption when theorizing about the nature of intentional agency.[136]

There is yet another reason to avoid the practical atomist's position: it risks making cases of basic *mistakes* rather mysterious. In Chapter 1, Anscombe provided an example of a basic mistake, where an agent simply missed pressing the button she meant to press, without the mediation of any false belief, and without it being a case of doing something else perfectly well, and without it being a case of interruption by (internal or external) interference. But the practical atomist doesn't seem to have the conceptual resources readily available to make sense of this case. Recall that the practical atomist requires basic actions to be *intentional*. It is because basic actions are intentional that they are a good candidate for being the foundation of practically-atomistic account of intentional agency: the intentionality of the basic parts can be transmitted to the larger wholes by means of thoughts about instrumental relations between the parts and the wholes, when those thoughts are reliable and true (or perhaps even: when those thoughts constitute knowledge). Intentional action carries with it the implication of *success* – doing something one has in mind to do. Accordingly, basic action is a *perfect* case of the will getting to grips with the world, even if only in an action that is quite small and limited in scope. If there is an *imperfect* case of the will getting to grips with the world, due to false beliefs, lack of grace or skill, or contingent interference, then either the thing that is imperfect is a

[136] The metaphysical feat mentioned earlier might qualify as a case in point. One might object that I can only perform an infinite number of non-basic actions *by* straightening my finger as a basic action. The rejoinder is, of course, that I also perform the purportedly basic action of straightening my finger by performing the infinite number of non-basic actions.

non-basic (unintentional) action, or else *there is no action at all* (that is, something may happen, but the will does not express itself, however perfectly or imperfectly, in action).

What could a basic mistake be under such a practically-atomistic conception of intentional agency? Either it isn't any kind of expression of the will, or else there is a kind of expression of the will that isn't in the form of (basic or non-basic) action. Now, as I will explain in §3.4, basic action theorists ought to acknowledge another form of expression of the will anyway: they ought to acknowledge that doing A intentionally is "processual" expression of the will that is not (yet) any kind of action. But I will also argue that the expression of the will characteristic of a basic mistake is not the same as that of doing A intentionally, because doing A intentionally is temporally-incomplete, and basic mistakes are a kind of temporally-complete expression of the will that I call a "deformed particular". These claims won't make much sense in advance of more detailed explanation: for the moment, consider the problem of accounting for basic mistakes a promissory note on why one should avoid the position of the traditional basic action theorist, just as it stands.

Let us turn to Thompson's pole of the antinomy. Here I have rather less to say. What is wrong with Thompson's view is that it seems to mistake *formal play* at rationalization for our *actual* practice of rationalization, where we have particular reasons for making the claims we make. For it is surely not merely conversationally speaking that Thompson's rationalizations concerning tiny geometrically-identified movements are odd: they seem to bridge the gap that exists between our actual practice of rationalizing actions and (somewhat nerdy) geometrical banter (Why is the chicken moving half a trillionth of the way across the road? To get a trillionth of the way across the road. Supposing the chicken sapient, this is a *paradigm* of rationalization for Thompson.)

If it is possible to have a view that avoids the unattractive elements of traditional basic action theory and Thompson's view then I think we should try for it. Where to start? Thompson's position is dialectically stronger. His regress serves to make his point explicitly: it is (it is supposed) a

completely ordinary feature of the continuity of time that supports his conclusion that there are no basic actions. By contrast, the support that the basic action theorist's regress offers for their conclusion is only as strong as the case made for the viciousness of the regress, and this viciousness is usually left implicit, so to that extent it is unclear whether the relevant regress constitutes a *reductio* or is simply a real feature of the case that we ought to accept. If our aim is to steer for a position somewhere in the middle, we should start with Thompson's dialectically stronger position, and then turn to the mysteries of basic action theory. The bulk of the chapter will consist in responses to Thompson's regress. In the course of responding to the regress a positive view will emerge that will also help us understand what is right, and wrong, about Thompson's view, and what is right, and wrong, about traditional basic action theory.

3.3 RESISTING THOMPSON'S REGRESS

The obvious point at which to resist Thompson's regress is the claim that we have no reason to deny that the relevant parts are intentional actions in just the same way that the encompassing supposedly "basic" action is supposed to be. In what follows I shall assume that the appropriateness of Anscombe's special question "Why?", understood as a demand for reasons for action, delimits the domain of intentional action. If the question is not properly applicable to something the agent is doing then (whatever else it is) that thing is not an intentional action.[137] With reference to Thompson's regress in particular, Anscombe's special question "Why are you moving from *A* to *B*?", understood as a demand for reasons for action, must be shown to have application to the

[137] The question may be applicable when the agent has no particular reason for acting, beyond the null-reason provided by the answer "No reason, I just thought I would."

movement from *A* to *B*, no matter how small the part (and movement) described happens to be, if Thompson's argument is to be successful.

3.3.1 Thompson's Argument is Inconclusive

Anscombe says that her special question "Why?" is denied application when the agent is completely unaware of whatever is asked about. When we ask for reasons for action we ask after something the agent has in mind to do: if she is *completely* unaware of the aspect in question, then obviously she doesn't have that in mind at all. But it is hard to deny that Thompson's agent is aware of going halfway (quarter-way etc.) in some sense of 'aware'. Everyone who has a minimal grasp of the nature of movement knows that one must go halfway if one goes all the way. Even those who are not *au fait* with the concept of division might be able to *see* or *imagine* that one goes halfway (*this* far) when one goes all the way (*that* far).

That said, it makes a difference *how* the agent is aware of moving to the halfway point. There are many cases where one may truly observe or infer (or predict or remember) that one is (or will be or has been) centrally involved in some change, such that one is aware of the change in some sense, whilst also truly claiming that one does not intend to do *that*, and does not do *that* intentionally, and does not have any reasons for doing *that*. Anscombe takes it that the kind of *practical* awareness of what one is doing that is relevant to demands for reasons for action is so far divorced from other ways of knowing what one is doing that her special question "Why?" is denied application when the agent must have recourse to observation or inference in order to answer it.[138] Anscombe's claim here is bound up with her conception of *practical knowledge* as distinct from other kinds. There is some puzzle about whether such non-observational, non-inferential practical knowledge is knowledge of

[138] Anscombe (1957: 13-14; 49-51).

what one is *actually doing* or only knowledge of what one has in mind to do (whether or not one is actually doing *that*).[139] We can dodge the puzzle, for we only need the weaker claim. We need not delve into what Anscombe means by "practical knowledge" in order to appreciate the plausibility of the claim that it is not by observing or inferring that an agent knows what she has in mind to do in the way that is relevant to demands for reasons for action. Even if I need a moment to become alert and find the words to express what I have in mind to do as I drive (quite habitually, with a minimum of care and attention) to work, I do not *look* at my current movements, nor *infer* from some premise ("Well it's 8.50am on a Monday morning…") in order to work out that what I have in mind to do (here, now and thus) is to drive to work. The same is true of all those actions undertaken (habitually, perhaps with a minimum of care and attention) as compositional means to the end of driving to work: taking the shortcut, speeding through the stop sign, cutting that guy off at the lights etc.

To capture our subject matter by stipulation, let us say that if the agent has it in mind to do something, such that demands for reasons for action apply to what she has in mind to do in the way that interests philosophers of action, then the agent *intends* to do that, and that various expressions of her intent express her *practical conception* of what she is doing and how. The agent intentionally driving to work intends to drive to work, intends to take the shortcut etc. etc. According to the stipulation, it is only of those things that the agent intends to do that a demand for reasons for action is appropriate: the agent's intentions capture all the things that the agent has in mind to do in the relevant sense. So we may ask a driver many questions about her driving, but it is inappropriate (or anyway, missing the agent's point) to demand a reason why she is dribbling snot out of one nostril (unless she intends to be gross, or intends not to be). In order to not beg questions against Thompson's (unorthodox) approach to the subject of intentional action, I hereby stipulate that I

---

[139] Davidson's carbon copier is the standard case for beating up some puzzlement about this issue – see Davidson (2001: 92).

make no further assumptions about the nature of intention and its "psychological reality".[140] Presence of intention is, for present purposes, just a device explicitly designed to coordinate with the appropriateness of Anscombe's special question "Why?" and the presence of practical awareness of what one has in mind to do. (In particular, there is no implicit suggestion that intentions are occurrent psychological states, whatever 'occurrent' means in this context.)

The stipulation is sufficient to show that Thompson's argument is inconclusive. For although we know that the agent in question intends to move from *A* to *C* and does so intentionally, and although it is reasonable to suppose that the agent knows that there must be some halfway-point *B* on the way to *C*, we do *not* know *how* the agent knows this, or whether the agent *intends* to move from *A* to *B*. If the agent knows about the halfway point *only* by means of observation or inference, then (following Anscombe) she is not *practically* aware of the halfway point. Thompson's argument is inconclusive because he has not shown that the agent knows about the halfway point *otherwise* than by observation or inference.

Now it might be objected that the agent who intends to move from *A* to *C* *must* intend to move from *A* to *B*. For, it might be said, agents (at least those who are halfway rational and knowledgeable about such matters) must intend the *necessary means* to their ends, and going halfway is a necessary (compositional) means to going all the way. So it would be unintelligible (or anyway,

---

[140] What do I mean by "psychological reality" here? For example: nothing is claimed about whether such intentions are beliefs, or desires, or combinations of belief and desire, or more nebulous states (or habits, or frames, or aspects) of mind: the stipulation is supposed to be as ecumenical as possible. Nothing is claimed about whether the adoption of such intentions requires a specially-attentive act of consciousness or a decisive furrowing of the brow: probably many things the agent intends to do are adopted as means or as ends quite automatically and habitually (or perhaps as a function of having learned a skill or having been raised a certain way), without any obvious feeling or sign that marks their adoption. Nothing is claimed about whether one must adopt an intention *prior* to executing it: perhaps some intentional actions have the intention *in* the action, such that no relevant, prior, psychologically-distinct state can be pointed out apart from the fact of doing the thing (for this or that reason). Nothing is (yet) claimed about whether the adoption of one intention must be accompanied by the adoption of some others: perhaps if I intend to move from *A* to *C* I must intend to go halfway, but then again, perhaps not. It *is* claimed that if one intends to *Φ* one knows that one intends to *Φ* without recourse to observation or inference.

incredible) to suppose that an agent who intends to move from *A* to *C* might not intend to move from *A* to *B*, at least according to our permissive (stipulative) use of 'intention'.

The objection fails because it needs to be shown that the movement from *A* to *B* is a necessary *means* rather than a mere necessary *aspect* (part, presupposition, consequence) of what is done. Even granting that agents must intend the necessary means to their ends, agents need not intend all the necessary aspects of something they intend to do.

Suppose I am pounding a nail into a board to fix it to a post, and suppose further that there's no way in the circumstances to hammer without making noise, and that I know this, and that a would-be questioner cannot see *any* of the movement I am making, but can hear the godawful racket. Just as Thompson's question "Why are you moving from *A* to *B*?" makes some sense when the questioner cannot see all of what the agent is doing, it would be perfectly natural "conversationally-speaking" for the questioner in the hammering case to ask "Why are you making that godawful racket?" as if that were something that I really meant to be doing. But I don't conceive of making a godawful racket as promoting any of my ends, nor do I conceive of it as desirable in any particular way (nor do I think it manifests justice, or some other intrinsic value). To that extent I don't have a *reason* for making a godawful racket (not even the null reason: "I just thought I would"). Making a godawful racket is accidental to my intent, although it is something I am doing, and it is something I *must* do if I am to hammer. The fact that it is a necessary aspect of what I am doing (perhaps even one I *must* know about) does not show that it is something I do *intentionally*, or that it is a necessary means to some end of mine, or that it guides or delimits my activity in any interesting way.[141]

---

[141] The antecedent aptness of a question about what one is doing ought to be distinguished from an important feature of discourse whereby aspects of what one is doing may be brought to one's attention by a question and thus *incorporated* into one's practical conception of what one is doing (whether one likes it or not). Thus Cavell (2001: 232) notes that when one is ignorant of an aspect, that aspect is not part of what one is doing, but once someone has asked a question like

The hammering case exhibits systematic similarities with the tiny movements from *A* to *B* that we are considering. When I respond to the question "Why are you making a godawful racket?" by saying "Because I am hammering in a nail" I do not rationalize what was asked about. Such a response does not use the 'because' of rationalization, but rather the 'because' of (something like) mere efficient causal explanation. (We might say the hammerer's response provides a *starting point* for a conversation about reasons for action, rather than providing a reason for action, because the questioner has latched onto some aspect of the case that is accidental to the agent's intent and the *particular* instrumental order the agent has in mind to pursue.) Similarly, we might say of the agent who responds to the question "Why are you moving from *A* to *B*?" with the answer "Because I'm moving from *A* to *C*" that she does not (or at least: need not) rationalize what was asked about. Such a response does not use the 'because' of rationalization, but rather the 'because' of (something like) mere *formal* causal explanation, exploiting the ratio 2:1 to explain why she goes halfway to *C*.

What seems to go missing in Thompson's argument is the connection between intentional action and having *particular reasons* for doing whatever one does intentionally. The hammerer is *practically indifferent* to making a godawful racket – he neither intends to make one, nor intends not to – because he has no particular reason to make a godawful racket. (Should he realize he is pissing off the neighbors thereby, he might then *intend* to do so, for the reason that they deserve it). To the extent that it is up to the hammerer whether or not he intends to make a godawful racket, so too might an agent moving from here to there think that it is up to *her* whether or not moving to any particular point short of her final destination is an intentional movement or not. And in some cases

---

"Why are you making that godawful racket?" one's action cannot continue to have just the same character that it did before one was aware of the fact of disturbance. To *continue* to hammer the nail in after the question would no longer be to merely fix a board to a post, but to fix a board to a post *despite the irritation it causes one's neighbor*. Our topic is rather sparser than the question of how one's will bears on the will of others. We are interested in the agent's relationship to what they do *as such*. Later we can consider the agent's relationship to changing their mind about what to do by taking into consideration what they come to know (of the will of others, and of the circumstances of action) in the course of their deed. For the moment, when considering whether a question "Why are you doing X?" is apt or not, we suppose that the case is morally neutral and hold the agent's end (and knowledge) *fixed* at the time the question is asked.

(particularly for short or arbitrarily chosen trajectories) she may not see any particular reason to think that a movement from *A* to *B* is an *intentional* movement of hers. She can say (truthfully) that she is practically indifferent to the movement, because it is open to her to treat the movement as one of the many aspects of the case – such as neurons firing, muscle contractions, the precise placement of her foot, accompanying noises, and movement of air that is in the way – that in some sense take care of themselves.

3.3.2 Thompson's Conclusion is False

In response to our stipulative use of 'intention', Thompson needs to secure the conclusion that whenever an agent intends to move to *C*, she also intends to move to *B* (which is halfway on the way from *A* to *C*) and intends to move to *AA* (which is halfway on the way from *A* to *B*) and so on. The preceding discussion should be sufficient, I think, to raise doubts about the claim that particular practical thoughts, such as the intention to do *A* (rather than *X*, or any of the other particular things one might do), just do reach down to further particular practical thoughts, such as the intention to do *B*, where doing *B* is a proper part of doing *A*. Can we do any better than this in resisting the thought that there are no basic actions?

We can if we consider that intentional actions need not be *perfectly* articulated in their parts. We allow that someone who intentionally crossed the Sahara may have unintentionally fallen and rolled down a sand-dune during some part of her movement, without thereby ruining the intentionality of the movement across the Sahara as a whole. It does not seem to matter when or where the accidents occur so long as they aren't overwhelming: people get off on the wrong foot, stumble, and fall exhausted across finish lines yet still manage to win races. An argument against Thompson's conclusion then takes the following form. We imagine a case in which such an accident

occurs at the beginning of a movement, such that the agent moves from *A* to *B* but does not move from *A* to *B* intentionally, and yet *does* move from *A* to *C* intentionally (she gets off on the wrong foot and then rescues the movement overall from utter disaster). For what is an overwhelming accident – the kind that destroys intentionality – with regard to one movement need not be overwhelming with regard to another.

It may be objected that this argument turns on a bad analogy. Although we *can* make sense of the movement across the Sahara as an intentional movement that encompasses some embarrassing unintentional mistakes, pratfalls, and other failures, we cannot make sense of the *tiny* movements relevant to Thompson's argument as mistakes or failures in their own right. Who cares if you move one micron *forward* (say) rather than one micron backward at the beginning of your movement from *A* to *C*? But the objection helps our case more than it hinders it. For if we cannot make sense of the tiny movements as failures or mistakes, then neither can we make sense of them as *successes* – as the agent having done something *in particular* that the agent intended to do (according to our stipulated use of 'intention').

It may still be objected that we only credit the agent who moved across the Sahara with an intentional movement because of the things she did *right* in the course of that deed. Had her progress been an endless series of pratfalls, we could not credit her with an intentional movement overall, but only a lucky (probably hilarious) accidental one. And all Thompson needs for his style of argument to go through is the claim that for any completed intentional action, there is some *proper part* of that action that is an intentional action in its own right. The proper part need not encompass any particular geometrical point (like halfway, or a micron at the beginning, or a micron at the end), but there has to be *something* (in particular) that the agent did right.

As far as it goes, this latter objection is a good one. We cannot credit an agent with an intentional action if she doesn't do *anything* right. But the objection does not show that in cases

where no pratfalls, mistakes or failures occur *what* the agent must do right is anything short of (anything that is a proper part of) the completed deed.

When the agent encounters certain kinds of obstacle, we have good reason require *more* of her in the way of practical thought if we are to credit her with an intentional deed overall. Her original thought, in its relative simplicity, does not suffice – she must *divide*, and thereby multiply, her practical cognition of the case, to account for the problem that is now a proper part of her endeavor. The agent prior to falling had no thought about getting up: after she has fallen we require such a thought of her, dividing her project into parts and addressing intentions to *particular* means that can make amends for the particular setback. But where there are no unexpected obstacles or difficulties, we have no good reason to posit further particular practical thoughts, and so no reason to think that the agent *must* intend to do something that is a proper part of a larger project. To think otherwise would be to treat life as if it were one *continual* overcoming of error, misfortune and particular (infinitely small) practical problems. (A tempting, but perhaps overly neurotic, thought.)

By way of making the role of what is necessary in the circumstances clear, we might question the setup of Thompson's central (iterative) case. Thompson says that the halfway point *B* is a *particular* place along some particular path to *C* that the agent is following, where the path is given to the agent in sense or imagination, such that it would be as much true to say of the agent that she is heading to *B* as to say that she is heading to *C* when she sets out from *A*. If so, then moving to *B* is probably not a *necessary* means to moving to *C*. Most of the time one's movement from one place to another is not constrained by tunnel walls such that there is only *one* way to get to one's goal. Indeed, in most cases, for any *particular B* within sight, one could miss that *B* and not rule out the possibility of moving from *A* to *C* successfully.

A further argument against Thompson's conclusion then takes the following form. We imagine a case in which an agent moves from *A* to *C* intentionally but does so *without* an intention to

follow some determinate particular path from *A* to *C* that is marked out in obvious ways from other possible paths to the goal (such as a road, or some other narrow, complete, determinate trajectory) because there is no *need* to settle on such a path in order to go to *C* intentionally. Such an agent intends to move to *C*, but leaves her practical thought at that level of relative determinacy. The explanation of why she takes the *particular* path she does take will then appeal to her skills and habitual ways of moving rather than to her particular intentions. (We might note in passing that the explanation of why the agent starts *when* she does could similarly appeal to her skills and habitual ways of moving. If there's no need to posit a particular determinate specification of a path in the content of her intention, there's no need to posit a particular determinate specification of a time at which to get going either.) For such an agent, at each moment of the movement there will be some segment of an ultimately successful trajectory from *A* to *C* that she has completed, but given the absence of an intention to move to those places in particular, or to follow a particular path that included them, we may say that prior to reaching them she was *practically indifferent* to them: for any of these places, whether specified as parts of some particular path or described in their own terms, she neither intended to go *there*, nor intended not to, before she got there. If she was practically indifferent before she reached them, we may say that she *is* practically indifferent now – by the time a particular intention might be relevant to them, the places are already taken care of (just as the starting point was). So although it is true of such an agent that during the whole time of the deed she was moving to *C* intentionally, at no time was there some particular *B* halfway along a particular path to *C* such that she was moving *there* intentionally (and similarly for any other geometrically-identified place one might fix upon short of *C* itself). The agent lacks an intention to do something that is a proper part of moving from *A* to *C* intentionally, so Thompson's argument fails.

It might be objected that an agent who intends to go to *C* and does so intentionally must at least have intended to go *towards C* (or *forwards*, or *roughly that way*) during the movement, even if she

stops short of settling on some particular determinate path to the goal. If the agent's employment of her skills and habitual ways of moving is to count as an exercise of *practical thought*, rather than a blind response to some psychological occurrence, then she must determine her own activity *in* thought so as to make it instrumentally relevant to her end: she must choose a rough *direction* to move in, even if she does not thereby choose a determinate *path* from point of origin to goal, by way of addressing the requirement to settle on a sufficient means to her end.

If this is right, then at each moment of the movement the agent will have completed a stretch of activity of moving *towards C* (or *forwards*, or in *that* direction). When completed, the stretch of activity as a whole has the look of a successfully completed intentional action of moving from *A* to *C*. Dividing up this mass of activity, we can isolate an infinite number of component stretches of activity in thought, by isolating each stretch in time. Supposing such isolation and division is legitimate in the realm of *practical* thought, each component stretch of activity looks to be a successfully completed intentional action, and a variation on Thompson's conclusion – a *strong* variation, that grants intentionality to *all* the parts of this kind of action, rather than just some – seems to go through.

The isolation in question is not legitimate in the realm of practical thought. Consider the content of the proposed infinite series of intentions in Thompson's original example. The contents are all different: "I intend to go to *B*"; "I intend to go to *AA*"; "I intend to go to *AAA*" and so on. To the extent that one is swayed by Thompson's argument, one has good reason to think that *distinct* intentions address these various distinct endeavors, because the agent has chosen a path to follow, the places on the path are distinct, and the content of each intention varies with the place on the path identified, and intentions are individuated by their content. Now consider the content of the proposed infinite series of intentions in the new example proposed, of an agent who intends to move towards *C* but does not settle on some determinate path. These intentions all have the *same*

content: "I intend to move towards *C* (or forwards etc.)". We cannot distinguish these practical thoughts from each other on the basis of their content, so we have no reason to posit a multitude of intentions. It will not help to say that they are differentiated by distinct implicit temporal indexes: we have no reason to suppose that temporal indexes must be a part of their content. We would do better to say that the agent has *one* practical thought here that unites her continuous activity during this period of time into *one* continuous motion to *C* (and perhaps beyond). Were there an unexpected obstacle, we would require more of the subject in the way of practical thought, for merely moving forwards (say) will not carry her all the way to her goal. But in the ordinary case there are no obstacles, and her practical thought, in its relative simplicity, is sufficient to sustain the relevant intentional action all the way to completion.

So: given that it is possible to move intentionally to *C* without settling on a determinate path, we have reason to deny Thompson's conclusion that there are no minimal units of intentional action. We have not established that there must be basic actions in every expression of the will, but only that there can be. So there is a sense in which Thompson in quite right to ask the question "Why *must* there be basic actions in every case?" In fact, for all that has been said, it seems relatively easy for Thompson to prove that it is possible to act intentionally without acting through a series of basic actions. The discussion above has revealed intentional actions as individuated by their criteria of success or failure, much as beliefs are individuated by their truth conditions. We *insist* on criteria of success or failure when we have *particular reason* to do so. The agent has a particular reason to move to *C*, which is why it counts as a success, but she needn't have a particular reason to move to any particular *B*, which is why moving to the particular *B* she does move to need not count as a success, and so need not count as an intentional action in its own right. (It could have, if she had *planned* on moving to that particular *B*.) Similarly, given that the particular (philosophical) reason to

do so is intelligible, Thompson can prove that it is possible to act without acting through a series of basic actions simply by intending to do so, and doing so.

Put in simpler terms: Thompson is right that there is an infinite *potential* for rationalization in any case of intentional action. His mistake is to treat the potentiality as an *actuality*. The problem is not that we *can't* encompass the required infinite series in thought, but just that we usually *don't*.

3.3.3 Practical Thought as an Immanent Order of Reason

Earlier I said that I would not presuppose anything about the "psychological reality" of intentions, beyond their tie to practical awareness and Anscombe's special question "Why?", so as not to beg questions against Thompson. One might object that the distinction between intentions and habits or skills – where habits or skills, rather than the content of an intention, explain the determinate path that the agent takes to *C* – sneaks in some objectionable presuppositions about the "instrumental reality" of exercises of skills. In particular, one might think that it presupposes that an instrumental order cannot be (anyway) present in the exercise of a skill, even when it is unaccompanied by explicit awareness of that order.[142]

Consider the following passage from Thompson (2008):

> …as Aristotle (for example) teaches, skill or craft or *technē* often drives out deliberation. What is done in accordance with skill in doing B, or in exercise of a practical capacity to do B, is not, as such, determined by deliberation or reflection – unless by a peculiarity of the skill itself (which might involve measurement and calculation, say, as laying carpeting does). But the absence of reflection does not make the action thus skillfully performed, making a pot of coffee, as it might be, or raising a hand, into a sort of unanalyzable whole; egg-breaking certainly does not lose its character as an intentional action after the agent's thirty-fourth omelet. Why should we suppose that acquisition of the type of skill that interests us, skill in moving a limb or object along this or that type of path, must deprive movement along sub-paths of their status an intentional?[143]

---

[142] I am grateful to participants of the Time and Agency conference 18th-19th November 2011 at George Washington University for pressing me on this point.
[143] Thompson (2008: 108)

In this passage Thompson is responding to the idea that when someone acts intentionally, the concept expressed by the description under which what the agent does is intentional must figure in some occurrent thought of the individual agent. If by an occurrent thought one means that the object of thought is given explicit attention, or the thought is said to oneself (*sotto voce*, or *in foro interno*) Thompson is surely correct that this is not always required for intentional action. Perhaps when we are learning omelet-making we must attend to the instrumental stages of omelet-making as the steps in the recipe that they are, but once we have internalized the relevant instrumental order, attention fades away, having done its work. After internalizing how to make an omelet through repetition, it doesn't matter for the intentionality of what is done whether one attends to the things one does whilst making an omelet. What matters is that there *is* a determinate and discernible instrumental order inscribed in what one does. Perhaps bodily movement is like that too, where exercises of our skill in bodily movement follow a "roadmap of the body" that was inscribed in our activity long ago and has now fallen *well* below the level of consciousness.

I am not opposed to the idea that practical thought could just *be* an immanent order of instrumental reasoning inscribed in the exercise of acquired skills, in a way that often falls below the level of consciousness (where 'consciousness' has the sense of 'explicit awareness' or 'attention'). But it is crucial that the elements of this order are available to practical awareness on given occasions. I do not claim that practical awareness of the kind coordinated with our stipulated sense of 'intention' requires reflection or attention at the time of action. The well-practiced early-morning omelet-maker is practically aware that she is breaking eggs in the service of omelet-making, even if she is doing it almost "in her sleep" (as we say), and responds to queries about what she is doing with nothing more than a distracted, half-somnolent grunt. Her practical awareness is shown merely by the fact that when she *is* fully awake and understands the point of our questioning, she can truthfully answer in the affirmative: "Yes, I was breaking eggs in order to make an omelet". If she

couldn't answer this way, and yet clearly understood the point of our questioning, we would have a real puzzle as to whether any intentional action took place or not: the kind of puzzle that leads one to weigh forgetfulness, repression and sleepwalking against each other.

I suspect that practical awareness does not even require much facility with analysis and articulate verbal expression. There are cases where an agent might need some *education* about the instrumental order that is really present in the exercise of her skills, including education about how to talk about that order and how to attend to its joints, in order to respond appropriately to inquiries about the instrumental order she is pursuing. Consider a ballet dancer who performs a subtle special dance move as part of a very complicated dance, but has never been taught the move's name, or discussed how it fits in the instrumental order of that particular dance. Perhaps she has never attended to the move in particular before, except in the sense that during practice she repeats that section of the dance if she happens to muck up the special unnamed move. If you ask her about what she is doing right at the moment she is making the move, she may be puzzled as to what to say. But after a few minutes (or hours) of education and discussion, the dancer might say truthfully: "Yes that *is* a step in the dance, and it has just the point you describe, instrumentally speaking." This need not be a *revelation* to her, except in the sense that she now has a way to talk about something she was (all along) practically aware of, although she didn't attend to it explicitly before, for she had no particular reason to attend to it.[144]

In this last case, the element of confirmation by the individual agent is all important. For if the dancer were to (truthfully) deny that the special move were part of the instrumental order she is pursuing, we could not say it was a means she was taking to her end, regardless of the fact that she

---

[144] By contrast, the conclusion of Thompson's argument, were the argument sound, really *would* be a revelation: it would show us something extraordinary about the depth of practical thought (apparently) at work in our everyday activities. Compare Anscombe (1957: 87): "What is necessarily the rare exception is for a man's performance in its more immediate descriptions not to be what he supposes. Further it is the agent's knowledge of what he is doing that gives the descriptions under which what is going on is the execution of an intention."

made the relevant movements. Of course, the individual agent's denial that the move is a step in the relevant instrumental order pursued by her would only be intelligible if she did *not* implicitly regard mucking up that move as a failure. But given the individual agent's avowed practical conception of the case that excludes the move, and her counterfactual lack of concern for cases where she "mucks up" (or omits) that part of the dance in particular, we have no good grounds to suppose that the special move forms a proper instrumental part of what she has in mind to do. It is just this element of confirmation by the individual agent that can go missing in the cases of the tiny movements relevant to Thompson's argument.

Supposing this is right, *why* does the element of confirmation by an individual agent go missing? Here it may help to consider Anscombe's description of an Aristotelian practical syllogism:

> … an Aristotelian doctor wants to reduce a swelling; this he says will be done by producing a certain condition of the blood; this can be produced by applying a certain kind of remedy; such-and-such a medicine is that kind of remedy; here is some of that medicine– give it.
>
> It has an absurd appearance when practical reasonings … are set out in full. In several places Aristotle discusses them only to point out what a man may be ignorant of, when he acts faultily though well-equipped with the relevant general knowledge.[145]

There are several points of interest in this passage, but my present concern is with the elements of the instrumental order represented by the syllogism and their conceptual articulation.[146] Anscombe's setting out of the practical syllogism comes to an end with a simple general term for a kind of action: *giving*. My supposition as to what explains the lack of confirmation for the tiny movements relevant to Thompson's argument is this: practical awareness of the elements of an instrumental order is articulated by means of concepts, but that articulated structure is what we might call a *lazy* one. (My apologies for the imprecision of the term. If I could be more precise about what is at issue here I would. I do *not* mean 'vague' by 'lazy'.) The concepts through which practical awareness is

[145] Anscombe (1957: 79).

[146] One of the points of interest is the claim that Aristotle often only sets out a practical syllogism in full in order to point out what an agent may be ignorant of in cases of faulty action. This may seem to apply to the case of a failed 'giving': if the doctors drips the medicine into the patient's eye, instead of onto their lips as intended, one might think that an expansion of the syllogism is in order to account for what went wrong. But it would be a mistake to think that beliefs about the levels of care and attention required explain the faulty action: clumsiness is not false belief or ignorance.

articulated are as intricately specific as is required by the case in hand given what one has in mind to do. Now the case at hand might require the employment of various skills that have techniques, and those techniques might display "their own" instrumental order that is articulated by means of concepts, and elements of that order often fall below the level of conscious attention. But once again, the conceptually-articulated structure characteristic of techniques is a *lazy* one: the concepts that articulate it are as intricately specific as they have been made to be (usually, as intricately specific as they need to be, and no more) for the purposes at hand. In particular, they are not as intricately specific as the sections of Thompson's sub-paths require, and that is why the element of practical awareness goes missing at a suitably deep level of analysis of a movement along a path.

I think that Anscombe's description of the practical syllogism marks a surprisingly deep difference from Thompson, given other similarities of their views about the nature of intentional action. Anscombe agrees with Thompson that the point of practical syllogisms is not to record the contents of what one says to oneself (*sotto voce*, or *in foro interno*) but to display the order in what is done, and why what is done makes sense in the light of the given end.[147] But Anscombe, in her description of the practical syllogism, thinks that the syllogism has *already* been set out in full, in all its absurdity. There is, for instance, no further *instrumental* analysis of 'give it' into 'carry the medicine a quarter-way to the patient's lips along the appointed path'; 'carry the medicine half-way to the patient's lips along the appointed path' etc. to be made.[148] The analysis of the immanent instrumental order present in what is done ends with a *general* term; one that subsumes a multitude of ways of giving the patient medicine under it.

---

[147] Anscombe (1957: 80)

[148] As she says, the "… mark of practical reasoning is that the thing wanted is *at a distance* from the immediate action". She is quite happy to countenances an immediate action as the terminating point of the relevant instrumental order. See Anscombe (1957: 79).

I think that Anscombe's judgment that the relevant instrumental order has already been set out in full is more in keeping with how we learn concepts of movement and concepts of paths, sub-paths, landmarks, directions etc., and how we apply such concepts when refining skills and techniques of self-movement.[149] Human babies learn self-movement that is (relatively?) un-conceptualized by doing things like stuffing toys in their mouths, kicking and grasping. It is only later that these skills are drawn into the service of conceptually-articulated structures with sharp bounds, like those enshrined in the parental imperatives "Come to me!"; "Simon says: arms in the air!"; and "Get away from that power outlet!".[150] Those sharp bounds encompass any of a range of possible fulfillments, and they do not discriminate between the parts of the path taken that Thompson's argument depends upon. Similarly, specification of the *technique* of navigation along a route comes to an end with general terms for kinds of activity limited by terms for orienting landmarks: one *walks* down to the *corner*, *turns* left (relative to the direction one is *facing*), *keeps going* past the *schoolyard* and so on. One can of course *refine* one's technique by working out further tiny details as one goes along. As one is walking down to the corner, one might stick to the sunny side of the street, step around the pile of refuse, put a swagger in one's walk, and so on. But again the terms by which one refines a route along a sub-path are general terms, so there is much about what falls under them that they do not specify. (We do not need to specify them: one of the great beauties of the generality of thought is the way in which it allows us to disregard irrelevant details and get on with living well.) The doctor who fails to give the patient medicine, dripping it all over the floor instead, may invent a new technique for giving on the second attempt. When she does, she pursues a new instrumental order, that could be (absurdly) expressed in narrative form: "Now I'll go slowly

---

[149] See Hornsby (2007a, 2007b, 2011a) for reflections on the logical form of abilities. I think the present proposal, with its sharp emphasis on practical awareness of the individual agent, is distinct from Hornsby's position, where the element of practical awareness present in the exercise of skills is not brought squarely into view.

[150] Or as Anscombe (1981: 137) has it, such causative verbs as "*scrape*, *push*, *wet*, *carry*, *eat*, *knock over*, *keep off*, squash, *make* (e.g., noises, paper, boats), *hurt*".

towards the chair, now from the chair to the bed, now tilt back my hand to make it level, now pause until my hand stops shaking…OK, *finally* I did it *right*." There is no need to settle on line-thin trajectories here; what one settles on is just an ordered structure of continuous stretches of activity of a unified kind (e.g. 'going'; 'tilting'; 'pausing') bound by limits (e.g. 'the chair', 'until it's level', 'until it stops shaking').[151]

Compare Anscombe's description of the practical syllogism as already being set out in full to Thompson's thought that an order of reason is inscribed *infinitely-deep* within every intentional action:

> … it is not so much by being caught up in a rationalizing order, or in a "space of reasons", that behavior becomes intentional action; rather, the rationalizing order, that peculiar etiological structure, is inscribed *within* every intentional action proper… Any intentional action (proper) figures in a space of reasons as a region, not as a point; or, equivalently, each of them, whether hand-raising or house-building, is itself such a space.[152]

As a general description of the *possibilities* for intentional action, I think Thompson's description is an excellent one. Our capacity to reason, deliberate, calculate, analyze, symbolize, imagine, hypothesize and construct novel concepts give us what seem to be infinite degrees of freedom in dividing up our activity into novel instrumental orders particular to the individual on given occasions. Just as we can decide what to do, we can decide how to do it. Obsessive-compulsives (arguably) give an example of what can go wrong when the conceptual articulation of an instrumental order is much more intricately specific than is required for the agent's ends. But that is just one extreme: we are quite familiar with the gradations of such instrumental depth available to individual agents (who don't have psychiatric disorders), as with the disaffected teen who *insists* on having his own special walk, or the stickler for the rules who crosses the road at the pedestrian crossing when it is clear that there is

---

[151] Mourelatos (1978) has suggested that kinds of activity bound by limits can figure in the specification of open-ended (unlimited) kind of activity, as when one thinks of running-a-mile as a qualitatively different kind of activity from running-a-hundred-meter-dash. What Mourelatos contrasts is better described as a contrast between short-distance running (or sprinting) and long-distance running: the kind of activity involved is *indifferent* to the contingent limits we place upon it at the Olympics, or the name we give it drawn from our most common employment of the activity in the service of a finite end (in the service of an *action*). Someone could, for instance, run a hundred meter dash by exercising her skill at long distance running until she has run a hundred meters (she probably wouldn't win), or she could "run-a-hundred-meters" (sprint) for two hundred meters without a break in this *single* stretch of activity.

[152] Thompson (2008: 112).

no one for miles around. What unites these various projects is the criteria for success and failure articulated by means of general concepts applied in practical awareness of what one is doing. The individual depth and intricacy of that conceptual structure on given occasions is for the most part up to the individual agents themselves, just as it is up to them to determine their ends.

I will leave the proper (more precise) expression of what a 'lazy conceptual articulation' is to another (more appropriate) occasion. I only mean to sketch an explanation of what I take to be *plainly* true: namely that agents (like, I hope, myself) who understand the point of the relevant questioning deny that they are *practically* aware of the relevant tiny, purportedly intentional, movements, even if they are aware of them in some other sense.

3.4 THE FALLIBLE CAPACITY TO DO WHAT ONE HAS IN MIND TO DO

The arguments against Thompson's conclusion reveal a conception of action individuation as up to individual agents in a distinctive way. For the most part we go along with the concepts and conceptual structures we were brought up with in our practical projects. But once we realize the power we have to determine our ends and our means to our ends, we can mess with our practical conception of what to do, coming up with novel ends to pursue, and novel ways and means to those ends. There are standards of intelligibility for doing so: not just anything counts as deploying the concept *moving towards C*. There are also standards imposed by the techniques we have learned: not just anything counts as deploying the skill of omelet-making. There are also limits imposed by the structure of our bodies and the materials available to us. But within the bounds of those standards and limits we have a certain freedom to move.

I think that this broad conception of action individuation is best captured by treating intentional action as an exercise of the fallible capacity to do what one has in mind to do, where 'what one has in mind to do' refers to the kind of practical awareness that is so central to the discussion above. A proper defense and articulation of a fallible capacity theory of intentional agency would require a whole book (or at least: another dissertation). In this closing section I would like to point out just one respect in which I think a fallible capacity theory of intentional agency is an improvement on both traditional basic action theory and Thompson's position. A fallible capacity theory of intentional agency can explain why mistakes in performance are *particulars*.

3.4.1 Processual Expression of the Will

Traditional basic action theory seems to stand in need of supplementation.[153] We can see this by considering an interrupted basic action. Suppose I am straightening my finger as a basic action and some petty and violent rival chops it off before I manage to straighten the finger. According to the traditional basic action theorist, no basic or non-basic action has taken place: the will has not expressed itself in action. Yet there is a clear sense in which the will was *expressing* itself in what (otherwise) *would* have been a basic action. I may not have straightened by finger intentionally, but I was *straightening* my finger intentionally before the violent interruption. So there is such a thing as doing *A* intentionally that is different from having done *A* intentionally. Call doing A intentionally *processual* expression of the will.

---

[153] By traditional basic action theories I mean theories such as those of Davidson (2001) and Goldman (1970), who do not pay much attention to processual expression of the will. Hornsby was in this camp as well, although she seemed dimly aware of processuality even then – see Hornsby (1980: 80). These days Hornsby is all for processual expression of the will – see Hornsby (2011b).

Processual expression of the will has only received widespread attention in philosophy of action fairly recently.[154] Processual expression of the will exhibits what has been called the "openness of the progressive".[155] The fact that I am doing A intentionally does not imply that I will be doing A or will have done A at some time in the future. Generally speaking, if it takes time for me to do something, there is time for something to interrupt me (or for me to drop dead).[156] The fact that I am doing A intentionally does not imply that I have *done* anything intentionally in the service of my end either, unless the substitution for A and circumstances of action are quite special. If I wake up and decide to remain in bed, then I have already remained in bed intentionally, even if only for a little while. But most actions are not like remaining in bed, and remaining in bed, just as such, is not like remaining in bed *for ten minutes*, which represents the more usual case of doing something that takes time to complete.[157]

A supplementation of traditional basic action theory would include some account of processual expression of the will and would explain how it relates to basic and non-basic action. The proposed supplementation is more than a modest revision. One of the main claims of the traditional basic action theorist is that whenever one can speak of expression of the will one can speak of *an* expression of the will: expression of the will always falls into the logical category of *particulars*. But there is reason to think that a supplemented basic action theory would have to give up that claim.

For the supplemented basic action theory, we might think of processual expression of the will as the "stuff" or "activity-material" out of which particular (basic and non-basic) actions are

---

[154] For example, Wilson (1989), Thompson (2008), Moran and Stone (2009), Hornsby (2011b).
[155] See Comrie (1976) and Galton (1984) for accounts of the "openness" of the progressive.
[156] Of course one could define a sense of doing A intentionally that had this future-oriented implication. But my present concern is not with possibilities for regimentation of senses, so I will not go through the options here.
[157] Here I am ignoring progressive locutions as applied in analysis of intention for the future. There is a relatively clear sense in which one might say "I am remaining in bed tomorrow morning" whilst one is up and about today without contradicting oneself.

formed.[158] On this theory, basic actions are still the smallest units in which this "stuff" takes on determinate form, and until the stuff has reached completion in a basic action, the proponent of the supplemented basic action theory must say that there is no determinate particular action to which one might refer. This latter thought is well expressed in Thompson (2008):

> …just as "I baked a loaf of bread" entails "There is or was a loaf of bread, x, such that I baked x," so also "I performed an act of baking a loaf of bread" must entail something on the order of "There is or was an act of baking a loaf of bread, a, such that I performed a"; and similarly, just as "I was baking a loaf of bread" does *not* entail anything on the order of "There is or was a loaf of bread, x, such that I was baking x," so also "I was performing an act of baking a loaf of bread" should not be supposed to entail anything on the order of "There is or was an act of baking a loaf of bread, a, such that I was performing a"[159]

If Thompson is right, then we cannot attach a name to the developing deed as it develops, for "it" is not complete *as a deed* yet. And there are reasons to think that there is something right in this thought. Consider the very first instant at which it is true to say that I am doing A intentionally. Suppose I am hit by an asteroid at that instant and vaporized. Intuitively at least, I need not even have braced myself to take a step (or to straighten my finger). There need be nothing *in particular* in which my doing A intentionally consists. Even if the proponent of the supplemented basic action theory does not grant this point, and supposes that there is some other kind of particular present, she will have to concede that there is no particular *action* (yet) at the first instant I am doing A intentionally. Perhaps the mind does squeeze out discrete particular chunks of more or less shapeless "practical-thought-dough" whilst there is processual expression of the will, but for the proponent of the supplemented basic action theory, basic actions are still the smallest "loaves of the mind" that individual agents can bake.

---

[158] Given my ignorance of physics, I'm not sure how appropriate the analogy is, but we might say that the proposed supplementation of practical atomism is a "quantum field theory" of intentional action, which does its work shoring up the foundations. Both Thompson and the proponent of the supplemented basic action theory will agree that that the *original* or *fundamental* expression of the will is imperfective in form.
[159] Thompson (2008: 136).

3.4.2 Mistakes in Performance are Particulars

Mistakes in performance seem to belong to the logical category of particulars. When one makes a mistake in performance one does *something* determinate, complete and irrevocable that precludes perfection on that occasion and in that regard. Recall the example from Chapter 2: the baseball player's fumble consists in an inappropriate ordering of preparatory activity, such that she couldn't bring her hand up in time to catch the ball. Mistakes in performance need to be in the logical category of particulars so we can refer to their features to explain *why* they preclude perfection on that occasion and in that regard.

The sense of irrevocable completion that attaches to mistakes in performance is not the same as that which attaches to perfect expressions of the will (i.e. intentional actions). But it is intuitively plausible that it *is* a kind of completion: the *bad* kind. The bad kind of completion exhibits systematic similarities with the temporality of intentional actions. An ordinary intentional action isn't complete until the agent has done the thing in question. We might express this by saying that in the normal case of successful action the agent is doing A intentionally until it is impossible for her to fail, at which point she has done A intentionally. (The sense of 'impossible' here is temporal impossibility.) She might not know that she has done A intentionally, but that is beside the point. Similarly, whilst it is still possible to succeed, no irrevocable mistake has yet been made either. One makes a mistake in performance when one closes off the possibilities for success on that occasion and in that regard. For example, the shopper who has margarine in her cart instead of butter has made a local mistake, relative to a local intention: that is now an *event* that is part of history. But it need not be damning, relative to her overall goal. She can check the contents of her cart before going through the checkout register, thereby rescuing the performance overall from disaster, and turning what otherwise would have been a mistake (relative to the larger project, not the smaller one,

which is forever damned) into a perfect expression of the will. To express the various intuitions at work here, let us say that mistakes in performance are "deformed particulars". An adequate theory of intentional action would explain why the will's *normatively* imperfect expressions are particulars, while its *grammatically* imperfect expression in doing A intentionally doesn't exhibit particularity.

Can we explain why mistakes in performance are particulars by treating their particularity as parasitic on the particularity of intentional action? Davidson argues that all mistakes are actions that are intentional under one description but unintentional under another. His example is a naval officer who sinks the *Bismarck*, thinking it is the *Tirpitz*.[160] Davidson's treatment of mistakes will only work for mistakes that exploit constitutive means to ends and are partly explicable by false belief about those constitutive relations. There is some plausibility to the analysis of the given case: the agent did *one thing* perfectly well (sinking that ship over there), even if she thereby precluded doing what she intended to do (we may presume it was her last torpedo). But if there are basic mistakes in performance, then there are non-parasitic deformed particular expressions of the will, and their particularity will need separate treatment. And given that there are basic *actions*, we should expect there to be such basic *mistakes*.

3.4.3 Basic Mistakes in Performance

When Anscombe stubs her thumb on the phone machine, rather than pressing the button that would return her coins, she fails to do something she had in mind to do, through no external or internal interference, and without having any relevant false beliefs about where her thumb was,

[160] Davidson (2001: 46).

where the button is, and so on.[161] Let us say that this is a basic mistake in performance: a normatively-imperfect expression of the will that is not explicable in terms of the unintended results of some perfect expression of the will (such as a basic action). I take it that the category of a basic mistake is an intuitively plausible one. Stubbing one's thumb is not at all like causal deviance: it may be perfectly ordinary in its causal history and articulation. Stubbing one's thumb is not at all like cases where paralysis or convulsion interferes with one's normal exercise of one's powers, or when one's rival chops off one's finger. To countenance basic mistakes as imperfect expressions of the will is just to make the ordinary point that there is such a thing as clumsiness or lack of grace, and clumsiness is not a case of convulsion, ignorance, interruption, theoretical error, bad reasoning, or bizarre (causally deviant) coincidence. It might be embarrassing to admit, but the will can express itself in *something* that is both immediate and graceless. Shit happens.[162]

Basic mistakes in performance, like mistakes in performance generally, seem to be particulars. Now that Anscombe has stubbed her thumb, she must *try again* to press Button A. Presumably she will be more careful on the next attempt, and will adjust her technique accordingly. The sense in which Anscombe has tried once (and failed) to press Button A isn't the only sense of 'trying' that is in play in the example. Presuming that she hasn't changed her mind about *how* to get her money back (as she would, were she to pull out her crowbar and smash the thing) there is a different sense in which she is *still trying* to press Button A, and may do so on the second (presumably more careful) attempt. Similarly, whether or not she changes means midstream, she is *still trying* to get her money back, and so long as she does not do something that precludes the

---

[161] Anscombe's own case was pressing Button B, not stubbing her thumb. This is an irrevocable mistake that is more obviously irrevocable than stubbing one's thumb. On the old phone machines in the United Kingdom at the time, pressing Button B instead of Button A would ensure that you didn't get your money back. Thanks to John McDowell for educating me as to how the phone machines used to work.

[162] Or, perhaps more precisely, *turds* happen.

possibility of success, she may well succeed in getting her money back, all in one and the same performance, regardless of the fact that the overarching performance started so inauspiciously.[163]

I don't think that the supplemented basic action theory can account for the sense in which Anscombe, having stubbed her thumb, has to try again to do something she had in mind to do. For according to the supplemented basic action theory, no basic action has yet taken place, and basic actions are the only relevant countable that the basic actions theorist countenances. The supplemented basic action theorist can of course account for the second kind of trying, according to which Anscombe is still trying to press Button A, by saying that it is processual expression of the will. But that is not what we are after: we are after an account of the deformed particularity characteristic of basic mistakes, which is exhibited in the first kind of trying.

On this point, the fallible capacity theory of intentional agency is on stronger ground than the supplemented basic action theory. The concepts of an occasion and regard are built into the definition of normatively-imperfect exercises of fallible capacities. When it comes to intentional agency, what determines the occasion and regard is what the agent *has in mind to do*. And what the agent has in mind to do may come in many forms and many individually different articulations. Sometimes, it comes in the kind of form that delimits a small action like pressing Button A, such that if Anscombe stubs her thumb, she has precluded doing what she had in mind to do and must try *again*, if she still wants to pursue that kind of means to the end.[164]

---

[163] There is another kind of case. Consider the braggart trying to impress a girl by hitting the button to get his money back without looking at the button and instead looking the girl in the eye the whole time. (To make the example plausible, we must suppose the braggart is also a bit of a fool.) When the braggart misses, he need not be *still trying* to get his money back. Now (embarrassed) he might try to, scrabbling with the machine as required, but there is a clear sense in which this could be a *new* project that he did not have in mind before.

[164] I suspect that the concept of *continuity* is important here. Thompson (2008: 141) says that the progressive has a use 'in hiatus' where one can be playing poker intentionally even whilst one is, at that moment, not handling or thinking about cards, but getting a cup of coffee whilst waiting for the next deal. If Thompson means to imply that nothing is going on in the service of the end of poker playing during this time then he is mistaken: waiting is going on. There are some instrumental orders that cannot include waiting for the appropriate cues as proper parts, given what one has in mind to do. Consider tiptoeing to the end of the catwalk in one graceful *continuous* motion, for example. Any interruption here (e.g. to get a cup of coffee from somewhere off the catwalk) marks a break in that kind of activity and failure to do what

I should make clear that the sense of 'trying' in which Anscombe has already tried once (and failed) to press Button A can be slippery and difficult to sort out in individual cases. It is easy to tell what counts as an attempt at the Olympics; matters have been regimented for us. But when an agent is groping around for a light switch in the dark, there doesn't seem to be much in the nature of the case, beyond the agent's judgment of what she is doing and has done, to help us decide e.g. whether she has failed three times to hit the light switch and is now on her fourth flailing attempt, or whether she is *still groping* for the switch such that she hasn't (yet) failed to do anything she has in mind to do. But the fallible capacity theory of agency can make a virtue of this slipperiness in individual cases: the differences follow from the individual agent's practical conception of the case (and perhaps the instrumental order inscribed in her skills, where the supposed elements of that order must be available to practical awareness if they are *actually* present in the deed).

3.4.4 A Final Objection: How Could Basic Mistakes Be Mistakes?

In Chapter 2 I said that all mistakes consist in an inadequate or inappropriate ordering of activity, in a suitably broad sense of activity. This may prompt the objection that basic mistakes in performance cannot be mistakes, in the sense I described in Chapter 2, for they do not consist in *ordered* activity; they are rather things one just does, and does wrong.

I do not think that this objection, even if a good one, is harmful to the overall theory of fallible capacities outlined in Chapter 2. If necessary, I could give up the claim that all mistakes consist in an inadequate or inappropriate ordering of activity, in a suitably broad sense of activity,

---

one had in mind to do. In the case of Anscombe stubbing her thumb, it is tempting to suppose that if it is a basic mistake (and it need not be), then it is because some similar concept is deployed that doesn't tolerate the kind of interruption or waiting that poker playing does tolerate. I won't develop the thought here, beyond pointing out how it is parallel to the conceptual articulation of kinds of *continuous* activity bound by limits described towards the end of §3.3.3.

and say instead that most of them do. But I think there is reason to retain the claim, and a simple way to do so.

The solution to the objection is prompted by simple-minded reflection on what Anscombe did wrong when she stubbed her thumb. The natural criticism of what she did wrong is expressed in sentences along the following lines: "You missed it, pay attention!"; "Be more careful next time!"; "Slow down!" and so on. And when Anscombe tries again, she adjusts her technique by e.g. paying more attention, being more careful, and slowing down. With her adjusted technique, there is no mistake and no problem. It would be amazing if, focusing *all* her powers on the case at hand, including her power to take her sweet time, she nevertheless missed Button A.

Attending to *what's important*, as required by the case at hand given the circumstances in which one acts, and adjusting the ordering of one's activity as appropriate, is a learned skill or technique in the sense outlined in §3.3.3. This kind of skilful adjustment of one's activity to what's required in the circumstances is *continual*: whilst alive, we don't take breaks from it, except perhaps when incapacitated (as the word suggests). Often little attention or care is required, given the nature of the case and the other complicated skills one has internalized. (There are some things one can do intentionally whilst *literally* asleep, and only one of them is sleeping.) But sometimes more attention or care or time is required than is given, in such a way that a mistake is made, and perfection precluded. This is particularly obvious when it comes to *omissions*, like forgetting to put the salt in the soup. One precludes perfection in such a case because one doesn't figure routines for checking one's memory and progress *into* one's routine and progress: one did not reflect, in a quiet moment, on what steps would be required, nor did one go slowly, checking as one went along, nor did one consult the cookbook. The temporal bounds of the agent's ordered activities that are relevant to putting the salt in start well before the stage in the recipe when salt is called for.

Usually one doesn't care very much about these careless mistakes, because there are more important things in life than being careful and bothering oneself with every little detail, or regimenting every aspect of one's life, especially if one's skills are such that most of the time the soup is appropriately salty. But sometimes – particularly when internalizing new skills – these mistakes are of great importance.

BIBLIOGRAPHY

Anscombe, G. E. M. (1957). *Intention*, Harvard: Harvard University Press.

——(1981). "Causality and Determination" in *Metaphysics and the Philosophy of Mind*, Oxford: Basil Blackwell.

Aristotle (1995). *Metaphysics Theta* in Barnes J. (ed.), *The Complete Works of Aristotle*, Princeton: Princeton University Press.

Armstrong, Martin & Place (1996). *Dispositions: A Debate*, London: Routledge.

Baker, L. R. (1991). "Has content been naturalized?" in Barry M. Loewer & Georges Rey (eds.), *Meaning in Mind: Fodor and His Critics*, Oxford: Blackwell.

Baier A. (1970). "Act and Intent", *Journal of Philosophy*, Vol 67 No 19: 648-658.

——(1977). "The Intentionality of Intentions", *Review of Metaphysics*, Vol 30 No 3: 389-414.

Beere J. (2010). *Doing and Being: an Interpretation of Aristotle's Metaphysics Theta*, Oxford: Oxford University Press.

Bird A. (2012). "Dispositional Expressions" in Russell G. & Fara D. G. (eds.), *Routledge Companion to the Philosophy of Language,* London: Routledge.

Brandom R. (2001). *Making it Explicit*, Harvard: Harvard University Press.

Cartwright N. (1994). *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Cavell S. (2001). "A Matter of Meaning It" reprinted in *Must We Mean What We Say*, Cambridge: Cambridge University Press.

Clifford W. K. (1999), *The Ethics of Belief and Other Essays*, Amherst: Prometheus Books.

Comrie B. (1976). *Aspect: an Introduction to the Study of Verbal Aspect and Related Problems*, Cambridge: Cambridge University Press.

Danto A. (1965). "Basic Actions", *American Philosophical Quarterly*, Vol II, 1965: 141-148.

Davidson (2001). *Essays on Actions and Events*, Oxford: Oxford University Press.

Dretske F. (1986). "Misrepresentation", in R. Bogdan (ed.), *Belief: Form, Content, and Function*. Oxfrod: Oxford University Press.

——(1989). "The Need to Know," in Marjorie Clay and Keith Lehrer, eds., *Knowledge and Skepticism*. Boulder: Westview Press.

——(1991). *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass: MIT Press.

—— (1995). *Naturalizing the Mind*, Cambridge, Mass: MIT Press.

Fara M. (2008). "Masked Abilities and Compatibilism", *Mind*, Vol 117: 843-865.

Fine K. (1994a). "Essence and Modality" in Tomberlin J. (ed.), *Philosophical Perspectives* 8: 1-16

——(1994b). "Senses of Essence" in Sinnott-Armstrong W. (ed.), *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus*, Cambridge: Cambridge University Press.

——(1995). "The Logic of Essence", *Journal of Philosophical Logic* 24: 241-273

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass: MIT Press.

——(1997). "Special Sciences: Still Autonomous After All These Years." In Tomberlin J. (ed.) *Philosophical Perspectives* 11: 149-164

——(2008). *LOT 2: The Language of Thought Revisited*, Oxford: Oxford University Press.

Galton A. (1984). *The Logic of Aspect*. Oxford: Oxford University Press.

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

Goldman A. (1967). "A Causal Theory of Knowing", *Journal of Philosophy*, Vol 64: 357–372.

——(1970). *A Theory of Human Action*, Prentice-Hall.

Hornsby J. (1980). *Actions*, London: Routledge and Kegan Paul.

——(2007a). "Knowledge, Belief and Action.", in Beaney M., Penco C. and Vignolo M., *Representing and Inferring*, Cambridge: Cambridge Scholars Press.

——(2007b). "Knowledge and Abilities in Action.", in *Cultures: Conflict, Analysis, Dialogue (Proceedings of the 29th International Wittgenstein Symposium)*, Ontos Verlag

——(2011a). "Ryle's Knowing How, and Knowing How to Act" in *Knowing How: Essays on Knowledge, Mind, and Action*, Bengson J. and Moffett. M. (eds.), Oxford: Oxford University Press.

——(2011b). "Actions in their Circumstances" in Ford A., Hornsby J., Stoutland F. (eds.), *Essays on Anscombe's* Intention*,* Harvard*:* Harvard University Press.

Horwich P. (1998). *Meaning*, Oxford: Clarendon.

Humberstone I. L. (1992). "Direction of Fit", *Mind*, Vol 101 No 401: 59-83.

Jackson F. (1998). *Mind, Method and Conditionals*. Londond: Routledge.

Jacobson-Horowitz H. (2006). "Motivational Cognitivism and the Argument from Direction of Fit", *Philosophical Studies*, Vol 127 No 3: 561-580.

Kim J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kimhi I. (forthcoming). *Thinking and Being*. Harvard: Harvard University Press.

Lewis D. (1997). "Finkish Dispositions", *The Philosophical Quarterly* 47: 143–158.

Little M. (1997). "Virtue as Knowledge: Objections from the Philosophy of Mind", *Nous*, Vol 31 No 1: 59-79.

Machamer P., Darden L, and Craven C. (2000). "Thinking About Mechanisms", *Philosophy of Science*, Vol 67 No 1: 1-25.

Makin S. (2006). *Metaphysics, Book Θ*, Oxford: Oxford University Press.

Manley & Wasserman (2008). "On Linking Dispositions and Conditionals", *Mind* 117: 59–84.

McDowell, J. (2004). "Naturalism in the Philosophy of Mind" in de Caro M. and Macarthus D. (eds.), *Naturalism in Question*, Harvard: Harvard University Press.

——(2011). *Perception as a Capacity for Knowledge*. Milwaukee: Marquette University Press.

Millikan R. (1984). *Language, Thought and Other Biological Categories*, Cambridge, Mass.: MIT Press.

——(1995). *White Queen Psychology and Other Essays for Alice*, Cambridge, Mass.: MIT Press.

——(2004). *Varieties of Meaning*. Cambridge, Mass.: MIT Press.

Milliken J. (2008). "In a Fitter Direction: Moving Beyond the Direction of Fit Picture of Belief and Desire", *Ethics, Theory and Moral Practice*, Vol 11: 563-571.

Molnar G. (2003). *Powers: A Study in Metaphysics*. Oxford: Oxford University Press.

Moran, R. and Stone, M., 2009, "Anscombe on Expression of Intention" in Sandis C. (ed.) *New Essays on the Explanation of Action*, Basingstoke: Palgrave Macmillan, 132–168.

Mourelatos A. (1978). "Events, Processes and States", *Linguistics and Philosophy*, Vol 2: 415-434

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

Neander K. (1995). "Misrepresenting and Malfunctioning", *Philosophical Studies*, Vol 79: 109-141.

Nozick, R. (1981). *Philosophical Explanations*. Harvard: Harvard University Press.

Papineau D. (1993). *Philosophical Naturalism*. Oxford: Blackwell.

Platts M. (1997) *Ways of Meaning*, Cambridge, Mass.: MIT Press.

Price H. (1989). "Defending Desire-As-Belief", *Mind*, Vol 98 No 389: 119-127.

Pritchard, D. H. (2007). "Recent Work on Epistemic Value', *American Philosophical Quarterly*, Vol 44: 85–110.

Rey G. (1997). *Contemporary Philosophy of Mind: a Contentiously Classical Approach*, Oxford: Blackwell.

——(2007). "Resisting Normativism in Psychology" in McLaughlin B. & Cohen J. (eds.) *Contemporary Debates in Philosophy of Mind*, Oxford: Blackwell.

Rödl S. (2007). *Self-Consciousness*, Harvard: Harvard University Press.

Rosen G. (2001). "Brandom on Modality, Intentionality and Normativity", *Philosophy and Phenomenological Research*, Vol 63 No 3: 611-623.

Ryle G. (1963). *The Concept of Mind,* London: Penguin.

Schroeder T. (2003). "Donald Davidson's Theory of Mind is Non-Normative", *Philosopher's Imprint* Vol 3 No 1: 1-14.

Schueler G. F. (1991). "Pro Attitudes and Direction of Fit", *Mind*, Vol 100 No 2: 277-281.

Searle J. (1979). *Expression and Meaning*, Cambridge: Cambridge University Press.

——(1983). *Intentionality*, Cambridge: Cambridge University Press.

Shoemaker S. (2003). *Identity, Cause and Mind: Expanded Edition*. Oxford: Oxford University Press.

Smith M. (1994) *The Moral Problem*. Oxford: Blackwell.

Sobel D. (2009). "Subjectivism and Idealization", *Ethics*, 119: 336-352

Sobel D. and Copp D. (2001). "Against Direction of Fit Accounts of Belief and Desire", *Analysis*, Vol 61 No 1: 44-53.

Thompson M. (2008). *Life and Action*, Harvard: Harvard University Press.

van Gulick, R. (2002). "Non-reductive Materialism: Still the Best Buy at the Mind Body Bazaar" in Michael Pauen (ed.), *Phaenomenales Bewustsein: Entstehung und Erklarg. Berlin:* Mentis Verlag, 297-327.

——(2006), "Functionalism, Information and Content" in Bermudez J. L. (ed.) *Philosophy of Psychology: Contemporary Readings*, London: Routledge.

Velleman J. D. (2000). *The Possibility of Practical Reason*, Oxford: Oxford University Press.

Wedgwood R. (2007a). *The Nature of Normativity*, Oxford: Oxford University Press.

——(2007b). "Normativism Defended" in McLaughlin B. & Cohen J. (eds.) *Contemporary Debates in Philosophy of Mind*, Oxford: Blackwell.

Williamson T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Wilson G. (1989). *The Intentionality of Human Action*, Stanford, CA: Stanford University Press.

Zangwill N. (1998). "Direction of Fit and Normative Functionalism", *Philosophical Studies*, Vol 91 No 2: 173-203.

——(2005). "The Normativity of the Mental", *Philosophical Explorations* 8: 1-20

——(2010). "Normativity and the Metaphysics of Mind", *Australasian Journal of Philosophy* 88: 21-39.Another bibliography entry.