

# Towards a convincing account of intention

by  
Niel Henk Conradie

*Thesis presented in fulfilment of the requirements for the degree of  
Master of Arts (Philosophy) in the Faculty of Philosophy at  
Stellenbosch University*



Supervisor: Dr. V. Roodt  
Department of Philosophy

April 2014

## **Declaration**

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

April 2014

Copyright © 2014 Stellenbosch University

All rights reserved

## **Abstract**

This thesis has two aims. The first is to assess the cogency of the three most influential theories of intention – namely those of Elizabeth Anscombe, Donald Davidson and Michael Bratman. I identify five requirements that a convincing account of intention must fulfil, and then assess each of these theories in light of these five requirements. In the course of this analysis, I demonstrate that, while each of these theories contributes to our understanding of intention, none of them meet all the specified requirements. This leads to the second aim of this thesis, which is to develop an account of intention capable of overcoming the problems inherent in the foregoing theories of intention and hence fulfilling the specified requirements. This account is built around the definition of intention as a complex mental entity, consisting of two components: a revisable pro-attitude and a belief that the agent will try to fulfil this pro-attitude. It must further be possible for the agent to reflexively reconstruct the belief component without external information.

I begin by setting out the five requirements for a convincing account of intention. In each case, I explain why it is necessary for a theory of intention to meet the relevant requirement, and elaborate on what is needed for an account of intention to fulfil this requirement. The five requirements for a convincing account of intention are: 1) It must explain the unity of the three seemingly irreconcilable uses of intention; 2) it must explain the epistemic requirements for intention; 3) it must clarify the relationship between intention and motivation, intention and causes, and intention and reasons; 4) it must explain the relationship between intention and practical reasoning, and 5) it must clarify the relationship between intention and moral responsibility. Together, these five requirements form the yardstick against which I evaluate the different theories of intention.

With this yardstick in mind, I am then able to assess each of the influential theories of intention developed by Anscombe, Davidson, and Bratman. In each case, I examine how the relevant theory of intention fares in meeting each of the five requirements. This analysis shows that, while each theory provides a number of important insights, none of them succeeds in meeting all five requirements. Such analysis further enables me to identify the specific difficulties that have stymied the attempts of all three thinkers to develop a convincing account of intention.

Having identified the strengths and weaknesses of the three preceding accounts of intention, I then try to work out an alternative account of intention that would not fall prey to the same complications. Following the same *modus operandi* as before, I evaluate my proposed account against the five requirements for a convincing theory of intention. In each case, I show that my account not only succeeds in meeting the specified criterion, but also, crucially, that it is able to overcome the difficulties that have plagued previous attempts to fulfil this criterion. I conclude that, while this account is not necessarily conclusive, it does meet the conditions for a convincing account of intention and thereby casts some light into the conceptual darkness surrounding intention that Anscombe identified more than half a century ago.

## Opsomming

Hierdie tesis het twee oogmerke. Die eerste is om die oortuigingskrag van die drie mees invloedryke teorieë van intensie te beoordeel – naamlik die van Elizabeth Anscombe, Donald Davidson en Michael Bratman. Ek identifiseer vyf vereistes waaraan 'n oortuigende verklaring van intensie moet voldoen en beoordeel dan elk van hierdie teorieë aan die hand van hierdie vyf vereistes. In die loop van hierdie analise wys ek dat, alhoewel elkeen van hierdie teorieë tot ons verstaan van intensie bydra, geeneen aan al die gespesifiseerde vereistes voldoen nie. Dit lei tot die tweede oogmerk van die tesis, wat die ontwikkeling van 'n teorie van intensie behels wat daartoe in staat is om die probleme wat inherent aan die voorgenoemde teorieë is, te oorkom en wat dus aan die gespesifiseerde vereistes voldoen. Hierdie teorie berus op die definisie van intensie as 'n komplekse mentale entiteit wat uit twee komponente bestaan: 'n wysigbare pro-houding en 'n oortuiging dat die agent hierdie pro-houding sal probeer vervul. Dit moet verder ook vir die agent moontlik wees om die oortuigingskomponent refleksief te rekonstrueer sonder eksterne inligting.

Ek begin deur die vyf vereistes vir 'n oortuigende verklaring van intensie uiteen te sit. In elke geval verduidelik ek hoekom dit nodig is vir 'n teorie van intensie om aan die relevante vereiste te voldoen en werk ek uit wat nodig is vir 'n verklaring van intensie om aan hierdie vereiste te voldoen. Die vyf vereistes vir 'n oortuigende verklaring van intensie is: 1) Dit moet die ooreenstemming tussen die drie skynbaar onversoenbare gebruike van intensie verduidelik; 2) dit moet die epistemiese vereistes vir intensie verduidelik; 3) dit moet die verhouding tussen intensie en motivering, intensie en oorsake, en intensie en redes verhelder; 4) dit moet die verhouding tussen intensie en praktiese redenering verhelder; en 5) dit moet die verhouding tussen intensie en morele verantwoordelikheid verhelder. Gesamentlik vorm hierdie vyf vereistes die maatstaf waarvolgens ek die verskillende teorieë van intensie evalueer.

Met hierdie maatstaf in gedagte is ek dan in staat daartoe om elkeen van die invloedryke teorieë van intensie, wat ontwikkel is deur Anscombe, Davidson en Bratman, te beoordeel. In elke geval ondersoek ek hoe die relevante teorie van intensie vaar in die voldoening aan elkeen van hierdie vyf vereistes. Hierdie analise wys dat, alhoewel elke teorie 'n aantal belangrike insigte bied, geen van hul daarin slaag om aan al vyf vereistes te voldoen nie. So

‘n analise stel my verder in staat om die spesifieke probleme te identifiseer waardeur die pogings van al drie denkers om ‘n oortuigende verklaring van intensie te ontwikkel, gestuit is.

Nadat ek die sterk en swakpunte van die drie voorafgaande verklarings van intensie geïdentifiseer het, probeer ek dan om ‘n alternatiewe teorie van intensie uit te werk wat nie aan hierdie selfde komplikasies onderhewig is nie. Deur dieselfde modus operandi as voorheen te volg, evalueer ek my voorgestelde verklaring aan die hand van die vyf vereistes vir ‘n oortuigende teorie van intensie. In elke geval wys ek dat my verklaring nie bloot daaraan slaag om aan die gespesifiseerde kriterium te voldoen nie, maar ook, van deurslaggewende belang, dat dit in staat daartoe is om die probleme te oorkom waardeur vorige pogings om die kriterium te vervul, geteister is. Ek kom tot die gevolgtrekking dat, alhoewel hierdie teorie nie noodwendig afdoende is nie, dit wel die voorwaardes vir ‘n oortuigende verklaring van intensie vervul en hierdeur lig werp op die konseptuele duisternis waarin intensie gehul is en wat meer as ‘n halfeeu gelede deur Anscombe geïdentifiseer is.

# Acknowledgements

I wish to thank my father and mother, without whose backing and support my interest in pursuing my philosophy may never have taken root. To my brothers, for being prepared to sit through my arguments, and for arguing back!

Most importantly, my thanks go to Dr. Roodt. It is no exaggeration to say that if not for her insightful comments and steadfast guidance, this thesis would never have been completed.

## TABLE OF CONTENTS

<b>INTRODUCTION.....</b>	<b>1</b>
<b>CHAPTER 1: FIVE REQUIREMENTS FOR A CONVINCING ACCOUNT OF INTENTION.....</b>	<b>10</b>
Introduction.....	10
1. Explaining the unity of the three seemingly irreconcilable uses of intention.....	10
2. Explaining intention’s epistemic requirements.....	14
3. Explaining the relationship between motivations, causes, reasons and intention.....	15
4. Explaining the relationship between intention and practical reasoning.....	17
5. Explaining the relationship between intention and (moral) responsibility .....	19
Conclusion .....	22
<b>CHAPTER 2: ANSCOMBE’S ACCOUNT OF INTENTION .....</b>	<b>23</b>
Introduction.....	23
1. The unity of the three seemingly irreconcilable uses of intention .....	25
1.1 The reduction of the three uses of intention to intentional action.....	25
1.2 The problem of pure intending .....	29
1.3 The problem of plural intentions.....	30
2. Epistemic requirements.....	33
2.1 Knowledge of intention as non-observational .....	33
2.2 Objections to knowledge of intention as non-observational knowledge .....	35
2.3 A refinement of Anscombe’s approach: Intention-as-belief.....	38
2.4 Acting under a description .....	39
3. The relationship between motivations, causes, reasons and intention.....	41
3.1 Distinguishing reasons and causes .....	41
3.2 The position that intention is not a motivation .....	45
4. The relationship between intention and practical reasoning.....	47
4.1 Intention’s role in practical reasoning.....	47
4.2 The problem of plural intentions.....	50
4.3 The charge of cognitivism .....	51
5. The relationship between intention and (moral) responsibility .....	53
5.1 Three types of consequences of intentional action: intended, foreseen and unforeseen .....	53
5.2 The “principle of side-effects” .....	54
Conclusion .....	57

<b>CHAPTER 3: DAVIDSON’S ACCOUNT OF INTENTION .....</b>	<b>59</b>
Introduction.....	59
1. Davidson’s early account of intention .....	61
1.1 Intention as a primary reason .....	61
1.2 Explaining the unity of the three uses of intention .....	62
1.3 Intention as a reason, motive and cause.....	63
1.4 The epistemic requirement for intention and practical reasoning.....	64
2. The unity of the three seemingly irreconcilable uses of intention .....	65
2.1 The turn to intention-for-the-future .....	66
2.2 Analysis of Knobe and Burra’s criticisms of Davidson’s account of intention.....	69
3. The relationship between intention and practical reasoning.....	76
3.1 Intention as a rational judgement .....	77
3.2 Apparent impossibility of plural intentions .....	78
3.3 The problem of bootstrapping and the charge of cognitivism .....	83
4. The relationship between motivations, causes, reasons and intention.....	84
4.1 Intention as a pro-attitude and a cause of action.....	85
4.2 Intention as an unconditional judgement .....	86
5. Epistemic requirements.....	87
6. The relationship between intention and (moral) responsibility .....	93
6.1 Reconstructing Davidson’s views on intention and moral responsibility.....	93
6.2 The problem of irrational intentions .....	95
Conclusion .....	96
<b>CHAPTER 4: BRATMAN’S ACCOUNT OF INTENTION.....</b>	<b>98</b>
Introduction.....	98
1. Explaining the unity of the three seemingly irreconcilable uses of intention.....	101
1.1 Reduction of the three uses of intention to intention-for-the-future .....	102
1.2 Two criticisms of Bratman’s reduction of the three uses of intention.....	105
1.3 The problem of plural intentions.....	107
2. Epistemic requirements.....	109
2.1 The intention-belief consistency requirement.....	109
2.2 Epistemic opacity and irrational intentions.....	110
3. The relationship between motivations, causes, reasons and intention.....	112
3.1 Intention as a conduct-controlling pro-attitude.....	113
3.2 Intention and causes .....	113
3.3 Intention as not counting as a reason for action.....	114

4. The relationship between intention and practical reasoning .....	117
4.1 The role of means-end coherence in Bratman’s account of intention .....	118
4.2 Critique of Bratman’s notion of self-governance .....	122
4.3 Analysis of Bratman’s critique of cognitivism .....	125
5. The relationship between intention and (moral) responsibility .....	129
5.1 Intended consequences and ethical rationalism .....	130
5.2 The problem of irrational intentions .....	131
Conclusion .....	132
<b>CHAPTER 5: AN ALTERNATIVE ACCOUNT OF INTENTION.....</b>	<b>135</b>
Introduction.....	135
1. The unity of the three seemingly irreconcilable uses of intention .....	138
1.1 Defending the reduction to intention-for-the-future .....	139
1.2 Critique of Searle’s argument for intention in action .....	145
1.3 Explaining the role of plans in cases of plural intentions .....	149
2. Epistemic requirements.....	152
2.1 Defending belief as a necessary component of intention.....	152
2.2 The role of epistemic opacity and understanding irrational intentions.....	154
3. The relationship between motivations, causes, reasons and intention.....	158
3.1 Intention as a revisable pro-attitude.....	159
3.2 Defending intention as a cause of intentional action .....	160
3.3 Intention-for-the-future and intentional action as sharing reasons .....	161
4. The relationship between intention and practical reasoning .....	163
4.1 Intention related to practical reason through deliberative procedures .....	164
4.2 Defending intention-as-belief without cognitivism .....	166
5. The relationship between intention and (moral) responsibility .....	169
5.1 Moral responsibility and the consequences of intentional action .....	169
5.2 Countering Knobe and Burra’s argument .....	172
5.3 Intentional pathological behaviour and moral responsibility.....	173
5.4 Are we morally responsible for anything we do unintentionally?.....	174
Concluding Remarks.....	177
<b>BIBLIOGRAPHY .....</b>	<b>179</b>

## INTRODUCTION

A man reverses out of his driveway and runs over a neighbourhood toddler. Bystanders rush to the scene and demand to know, “Why did you do this terrible thing?” He answers, “I didn’t mean to do it! It was an accident!” Based on this answer, we judge his actions in a certain way. Now consider a second example: Events progress exactly as before, except that in this case, when confronted by the “Why?” question, the man responds by saying, “The parents of this toddler killed my daughter; I did what I did because I intended to avenge her death.” This answer leads us to an entirely different judgement of the same set of actions. The two judgements involve different assessments of the causality at play, of the epistemic content of the mind of the driver, the driver’s motivations, the rational status of the action, and the grounds of moral responsibility. The main source of these differences is this: In the first example, there is no intention to kill the toddler, while in the second example, the action was undertaken with just such an intention. Adding intention to an action clearly has important ramifications for how we judge that action, not least for how we think about moral responsibility. And yet, despite the obvious importance of intention for making sense of action, we are often, in the words of Elizabeth Anscombe (1963: 1), “pretty much in the dark about the character of the concept which it represents.” The purpose of this thesis is to shed some light on this conceptual darkness. My specific concern here is to work out the requirements for a convincing account of intention. Such an account must explain the concept of intention itself, as well as its relationship to a number of other important concepts, such as practical reasoning, cause, motivation, knowledge, belief and responsibility.

The philosophical literature shows a number of attempts at developing a coherent account of intention. In this thesis, I examine three such accounts that have had the greatest impact on our philosophical understanding of intention. These are the theories of intention expounded by Elizabeth Anscombe, Donald Davidson and Michael Bratman.<sup>1</sup> As will become clear, each of these accounts offers some important insights, but each also suffers from certain crucial failings. Having set out these insights and failings in the course of a sustained analysis of the respective theories of intention, I then offer my own attempt at a convincing account of intention. This account will, I hope, retain the insights and avoid the failings of the foregoing

---

<sup>1</sup> These are not the only theorists of intention. Other accounts of intention are developed by Thompson (2008), Velleman (1989) and Harman (1986), amongst others. However, while I will refer to elements within these accounts where appropriate, Anscombe, Davidson and Bratman have had by far the greatest impact on the field and therefore deserve the most extensive analysis.

accounts, at least as far as my chosen requirements for a convincing account of intention are concerned.

Before I embark on the project outlined above, a remark on methodology: All three theories I discuss belong to a specific tradition of thinking about intention. This tradition begins with Anscombe's publication of her seminal text on the topic: the pointedly named, *Intention* (1963).<sup>2</sup> Davidson's work on intention is very much a response to the ideas and shortcomings of Anscombe's account. Similarly, Bratman's work is a direct attempt to improve upon the shortcomings in Davidson's account. Despite the substantial differences between these thinkers, they all follow a generally recognisable methodology: First, they specify a set of difficulties that an account of intention must overcome in order for it to be convincing, and then provide a definition of intention that is meant to overcome the stipulated difficulties. Although the definition of intention changes from one thinker to the next, what has crystallised across their different accounts is the set of difficulties that any account of intention must resolve in order for it to be convincing. For this reason, it is possible to judge these three accounts of intention in light of a set of generally shared requirements. In keeping with the tradition of Anscombe et al., I therefore first set out the requirements for a convincing account of intention, after which I explain and evaluate the accounts of intention developed by Anscombe, Davidson, Bratman and myself in light of these criteria. I devote a separate chapter to each of these topics.

Chapter 1 is divided into five sections, each of which deals with one of the criteria for a convincing account of intention. These criteria involve the ability of any such account to explain: (i) the three seemingly irreconcilable uses of intention; (ii) the epistemic requirements for intention; (iii) the relationship between intention on the one hand and motivation, causes, and reason on the other; (iv) the relationship between intention and practical reasoning and (v) the relationship between intention and the ascription of moral responsibility. The first four of these criteria are all present from Anscombe's *Intention* onwards. While Davidson and Bratman make a number of important contributions to our understanding of these criteria, neither of them adds to the list of criteria in their work on intention. The fifth criterion, on the other hand, is my own addition.

---

<sup>2</sup> Although Anscombe's 1963 monograph *Intention* is the most important contribution she made to the discussion about intention, it is in fact an expansion of her 1957 article of the same name. When I use the name *Intention* in this thesis I will always mean the 1963 work, unless I indicate otherwise.

The first criterion for a convincing account of intention relates to the difficulty that the concept intention is used in three seemingly irreconcilable ways: intention-for-the-future, intention-with-which and intentional action. A convincing account of intention must be able to determine whether or not these three uses are really just a matter of a single concept being applied in three different contexts, or whether they involve three separate concepts. If the latter is the case, then, instead of a single convincing account of intention, we would need three separate accounts of intention, one for each use. However, all three the thinkers I discuss argue for the former option, and each presents an argument for how to achieve a unity of the three uses. In my own account I follow suit, and argue for the view that intention-for-the-future, intention-with-which and intentional action are not three separate concepts, but different articulations of one overarching concept of intention.

The second criterion centres on the question: What is the nature and extent of the knowledge that an agent must have in order for it to count as having an intention? Any answer to this question will fall somewhere on the spectrum between certain knowledge and complete epistemic opacity. Answering this question will involve tackling specific problems, such as whether or not it is possible to have non-observational knowledge, the relationship between intention and belief, the possibility of unconscious intentions and how we understand cases of *akrasia*.

According to the third criterion, a convincing account of intention must be able to say whether or not intention is a special case of motivation, whether or not intention is a cause, and how intention fits into the domain of reasons. As we will see, any theory that sees intention as unrelated to motivation faces insurmountable difficulties, whereas understanding intention as a special kind of motivation yields a much more convincing theory of intention. Similarly, any theory that denies that intention is a cause ends up being incoherent on a number of fronts. On the other hand, treating intention as a cause raises the serious objection of causal deviance, and any convincing account of intention must be able to overcome this objection. As for the relationship between intention and reasons, this requires a further account of the conditions for the rational explanation of actions – something that, ever since Anscombe, has been considered an inviolable characteristic of intention.

The last of the criteria introduced by Anscombe and adopted to varying extents by Davidson and Bratman is that of the relationship between intention and practical reasoning. The nature

of this relationship is almost certainly the most contested issue in the debate about intention. In order to make sense of this criterion, I first define what I mean by practical reasoning, after which I consider the following set of questions: First, are intentions inputs or outputs of practical reasoning, or both? Second, if it is possible for an agent to have multiple intentions, what are the rules that govern their rational relationship? Finally, is Bratman correct in criticising so-called cognitivist accounts of intention, and if so, what implication does this have for intention's relationship to practical reasoning?

The final criterion introduced in Chapter 1 has the most profound implications for how we understand the course and purposes of our lives. It should be clear from the examples mentioned earlier that it is of cardinal importance that we are able to say how intention relates to moral responsibility. In many cases, the answer to the question: "Did agent X intend to Y?" can determine whether or not agent X faces moral as well legal sanction for actions related to Y.<sup>3</sup> It is my contention that an account of intention is incomplete as long as it doesn't explain how ascriptions of moral responsibility are related to the intentional status of the consequences of an action.

Armed with a proper understanding of the criteria for a convincing account of intention, as well as their attendant difficulties, as set out in Chapter 1, I then analyse the theories of intention developed by Anscombe, Davidson and Bratman. In each case, I focus on whether and how the theory fulfils each of the criteria specified in Chapter 1, and the extent to which it resolves the difficulties inherent in these criteria. Chapter 2 deals with the account developed by Anscombe, Chapter 3 with that of Davidson and Chapter 4 with that of Bratman. Chapter 2 is divided into five sections, one for each of the requirements for intention. Chapter 3 is structured slightly differently. Due to the turn from early Davidson to later Davidson, there are six sections, the first devoted to discussing early Davidson, while the next five will analyse later Davidson's account of intention in terms of the five requirements. Chapter 4 once again follows the structure of Chapter 3.

I begin my analysis of Anscombe's account in Chapter 2 with her attempt to unify the three uses of intention by reducing intention-for-the-future and intention-with-which to intentional action. I argue that this attempt fails, and that the reasons for its failure are indicative of the

---

<sup>3</sup> One need only consider examples such as those of Oscar Pistorius (Mail&Guardian, 2013), Sanele May (SABC, 2013) and Gilberto Valle (The New York Times 2013) to see this importance in action.

impossibility of any successful unity of the three uses through reduction to intentional action. It further becomes clear, in Section 2, that Anscombe argues for a very stringent epistemic requirement for intention, namely direct and non-observational knowledge. I claim that this position fails on account of giving insufficient credence to the epistemic opacity of intentional actions. The third section deals with Anscombe's crucial distinction between reasons and causes. Here I show that Anscombe thinks that an intention is *not* a cause, and, although it is necessarily related to reasons, it is not necessarily a reason itself. The distinction between intention and cause follows, in turn, from Anscombe's understanding of motivation – which she *does* consider a cause. She argues that, while motivation causes us to undertake a particular action, intention is an after-the-fact rationalisation – or at the very least, a rationalisation *in medias res* – of that action. I argue that this understanding of intention's relationship to motivation is untenable, and that, by extension, her view of the relationship between intention and causes is also untenable. However, I do consider her argument that intention is necessarily related to reasons to be convincing, and will later incorporate this view into my own account. In Section 4, I argue that Anscombe's understanding of the unity of the three uses of intention prevents her from giving a convincing explanation of the relationship between intention and practical reasoning, or of the relationship between multiple intentions. Lastly, in Section 5, I examine her claim that there are three types of consequences to intentional action (intentional, foreseen and unforeseen), and that moral responsibility is ascribed differently depending the type of consequence at stake. Although I disagree with elements of Anscombe's approach, most notably her use of the “what you did it for” question in cases of foreseen consequences, I do take her to be correct in thinking that any account of intention must be able to specify the relation between the different types of consequences and moral responsibility.

As I mentioned earlier, the analysis of Davidson's account in Chapter 3 has a slightly different structure. I first give a brief overview of the so-called “early Davidson,” particularly his attempt to unify the three uses of intention through a reduction to intention-with-which. However, this attempt fails – as Davidson himself admits – which moves him to develop a new approach to the unity of the three uses. I designate this new approach as “later Davidson.” Section 2 then examines how the later Davidson tries to unify the three uses of intention by reducing them to intention-for-the-future, understood as a mental entity and a rational judgement. Although this is a step in the right direction, I show that this definition of intention as a rational judgement is ultimately unsustainable. In light of this, Section 3

analyses Davidson's understanding of the relationship between intention and practical reasoning. Here I show that, given his commitment to intention as a rational judgement, he requires practical reasoning to play a very particular role. Unfortunately, this role is incapable of making sense of the relationship between intentions, or of irrational intentions. It also results in the problem of "bootstrapping" in practical reasoning. I expand on this point in Section 4, which deals with Davidson's treatment of intention as a reason. On a more positive note, I argue that Davidson is correct in asserting that intention is a motivation. The same goes for his position that intention is a cause, although he still provides no solution to the problem of causal deviance. In Section 5, I analyse Davidson's claim that intention is a rational judgement and show, once again, that this definition of intention leads to untenable outcomes. Lastly, since Davidson does not deal directly with the relation between intention and moral responsibility, I try to reconstruct an argument that Davidson might have made, and then measure it against the final criterion specified in Chapter 1. Overall, my analysis of Davidson's account shows that, while he makes a number of valuable contributions, his view of intention as a rational judgement is flawed, and intention cannot be considered a reason without raising the problem of "bootstrapping."

Bratman's account of intention, the focus of Chapter 4, is in many ways an attempt to overcome the difficulties afflicting Davidson's account. In Section 1 of this chapter, I make it clear that Bratman follows Davidson in trying to unify the uses through intention-for-the-future, still understood as a mental entity. However, Bratman has a different understanding of the content of this mental entity, and further argues that intention is a conduct-controlling pro-attitude rather than a rational judgement. Although Bratman is partially effective in overcoming the difficulties that plague Davidson's account, I argue that the former's theory of intention nevertheless does not unite the three uses of intention in a convincing way. In Section 2, I argue that, when it comes to our knowledge of intentions, Bratman gives too much scope to epistemic opacity, which prevents him from giving an adequate account of the epistemic requirements for intention. I further argue that this stems from his unwillingness to consider intention as belief. Section 3 shows that Bratman follows Davidson in considering intention to be both a motivation and a cause. However, like Davidson, he is unable to resolve the problem of causal deviance, leaving much of the important work on the relationship between intentions and causes still to be done. On the other hand, when it comes to the relationship between intentions and reasons, we see Bratman diverging dramatically from Davidson. Bratman argues that intention cannot be considered to be a reason, given that

this raises the possibility for “bootstrapping” in reasoning. This argument carries over into the analysis of Bratman’s intention’s relationship to practical reasoning in Section 4. In this section, I flesh out Bratman’s views on how rational pressures might apply to intention without intention being considered a reason. In the end, I conclude that, although Bratman’s criticisms of Davidson are correct, his solution is not convincing as it obscures the nature of the rational relationship between intention-for-the-future and intentional action, and requires too much work to be done by the notion of *self-governance*. Since Bratman doesn’t deal with intention and moral responsibility in a sustained way, Section 5 is an examination of his fragmented insights on this topic, namely: that intentional, and not only intended, consequences are subject to moral responsibility, that we must be cautious in allowing an account of intention to allow that having an immoral intention can rationally justify pursuing immoral means toward this intention, and lastly, that intention must strive to meet certain rational regularities. This final insight allows for a better understanding of moral responsibility in cases of *akrasia*.

Having examined the three most influential theories of intention in light of a standardised set of criteria, the final chapter of the thesis consists of my own attempt to develop an account of intention that would fulfil all of the specified criteria while avoiding the pitfalls that characterise each of the aforementioned theories. I begin by arguing that the unity of the three uses can be found in intention-for-the-future, understood as a mental entity. This mental entity consists of two components: a revisable pro-attitude and what I call an *intention-unique belief*, which can be reflexively reconstructed without external information. I then try to show that this definition of intention meets the five requirements for a convincing account of intention that I stipulated in Chapter 1 and successfully overcomes all the difficulties – thirteen in all – revealed by my analyses of Anscombe, Davidson and Bratman’s accounts. These difficulties are listed in the introduction of Chapter 5. After having defended the reduction of the three uses to intention-for-the-future in Section 1, the further steps in my argument are as follows: In Section 2, I argue that the epistemic requirements for intention can be characterised as “a belief to try,” while in Section 3 I argue that intention is to be considered both a motivation and a cause. However, it is not to be considered a reason, though it is conditioned by reasons. This leads into my argument in Section 4, that intention is not a necessary outcome of practical reasoning, but is always related to such outcomes. Finally, in Section 5, I argue that an agent is morally responsible for *all* the intentional consequences of her actions. By intentional consequences is meant all intended and foreseen

consequences, but none of the unforeseen consequences (which include consequences that *could* have been foreseen, but were not foreseen by the agent). Furthermore, I argue that, while in some cases an agent might be held responsible for unforeseen, though rationally foreseeable consequences, there is a distinction between the two types of responsibility here. In the case of one type of foreseeable consequences, the responsibility ascribed is *moral* responsibility, but in the case of the other type, the responsibility ascribed is what I term *corrigible* responsibility. I discuss what I mean by this distinction between two types of foreseeable consequences, and by corrigible responsibility in particular, in Chapter 5, Section 5.

Before turning to the criteria for a convincing account of intention, it remains for me to address two criticisms that might be levelled against either the scope or the methodology of this thesis. The first possible criticism is that I am making a mistake by focusing on *intention*, as opposed to *Intentionality*. Spearheaded by John Searle (1980),<sup>4</sup> some philosophers have argued that questions of intention are subsumed by questions of Intentionality, and that understanding the latter must take priority as the former dances to its tune, so to speak. Intentionality refers to “the ‘directedness’ or ‘aboutness’ of conscious experiences” (Moran, 2013: 317), whereas intention refers to that quality of an action that distinguishes the actions of the driver in the second example mentioned at the beginning of this introduction from the actions of the first. The difference is one of scope: where Intentionality can be applied to beliefs, or motivations, or perceptions as well as intentions (Thompson, 1986: 85), intention applies only in cases where an agent does something with an intention, or has an intention to do a thing (Bratman, 1984: 375). Although Searle may be correct in claiming that intention falls within his broader theory of Intentionality (see Searle 1980, 1983), this is not within the scope of this discussion. Rather, my focus falls on the tradition of inquiry beginning with, or at least given contemporary impetus by, the work of Anscombe. And this tradition is concerned with intention in the narrower sense.

The second possible criticism is that trying to unify the three uses of intention is an inherently fruitless venture, as no such unity exists. Since my project aims to provide a single convincing account of intention, as opposed to three disparate accounts each focussed on one

---

<sup>4</sup> The project of explaining Intentionality did not start with Searle, he himself builds on the work of thinkers such as Franz Brentano (1874; 2009) and Edmund Husserl (1900/1901; 2001). However, it is Searle who has been the strongest proponent of making understanding *intention* subordinate to understanding Intentionality.

of the uses of intention, if it is proven that the three uses are irreconcilable then my account must be error. This criticism is made by Knobe and Burra (2006), based on the results of an experimental study by Knobe of how respondents ascribed intention-for-the-future and intentional action to agents (Knobe, 2004). The argument is that, since the respondents ascribe intention-for-the-future to agents differently from how they ascribe intentional actions, they cannot be understood as uses of the same concept. Furthermore, they argue that the rules that govern the relationship between intention-for-the-future and intentional action indicate that these two concepts are not related to each other except for sounding similar. I discuss this criticism in some depth in my thesis (see Chapter 3, Section 2). At this point, it is sufficient to mention my primary arguments against this criticism: In the first place, Knobe and Burra's argument concerning the relationship between intention-for-the-future and intentional action is mistaken, as this relationship fails to account for the authority of *self*-ascriptions of intention. In the second place, their argument is built on a mistaken understanding of the distinction between intended and intentional consequences of intentional action. When this distinction is properly understood, the results of their experiments no longer prove the point they think it does.

Having outlined the methodology that I employ in this thesis and introduced its subject matter, as well as addressing two possible criticisms of my approach, I now turn to the first step in my argument: defining the five requirements of intention.

# CHAPTER 1: FIVE REQUIREMENTS FOR A CONVINCING ACCOUNT OF INTENTION

## Introduction

What is it that one looks for from an account of intention? What should such an account seek to explain? My aim in this chapter is to set out what I consider to be the five most important criteria by which a convincing account of intention should be judged. The first four of these criteria were introduced by Anscombe in *Intention* (1963), and all of the main thinkers on the topic have followed her lead in trying to develop a theory of intention that meets these criteria. The four criteria are: (i) Explaining the three seemingly irreconcilable uses of intention; (ii) explaining the relationship between intention and its epistemic requirements; (iii) explaining the relationship between motivations, causes, reasons and intention, and (iv) explaining the relationship between intention and practical reasoning. The fifth criterion I will be using is my own addition. I have settled upon it because not only does it receive insufficient attention in the literature, but also because it has the most profound consequences for how we, as creatures who act intentionally, live our lives, namely (v) explaining the relationship between intention and responsibility. Each section in this chapter is devoted to one of these criteria. In each case, I explain the nature of the criterion involved, as well as some of the notable difficulties involved in fulfilling this criterion. My contention is that whatever account of intention is best able to meet these criteria (and overcome the difficulties that stem from attempting to do so) can be considered to be the most convincing account of intention so far.<sup>5</sup>

### 1. Explaining the unity of the three seemingly irreconcilable uses of intention

Ever since the publication of Anscombe's *Intention* (1963), there has been near unanimous agreement amongst philosophers that any convincing theory of intention must meet at least one key criterion: it must explain the unity between the three seemingly irreconcilable uses of intention, namely as "intention-for-the-future, intentional action and as the intention with which someone acts" (Anscombe, 1963: 1).<sup>6</sup> In this thesis, I will likewise take the need to

---

<sup>5</sup> This does not exclude the possibility that there may be additional requirements that a convincing account of intention should meet. However, even if this is the case, the five requirements I have chosen here would remain necessary, even if they do not prove to be sufficient, for any convincing account of intention.

<sup>6</sup> See Davidson (2001), Velleman (1989), Bratman (1985) and Setiya (2011a).

explain the unity of the three uses as a criterion for adjudicating the value of any accounts of intention.

Right at the beginning of *Intention*, Anscombe observes that there are three common ways in which we employ the concept intention:

1. The agent intends to G (verb)
2. The agent G'd intentionally (noun)
3. The agent F'd with the intention of Ging (adverb)

The first use of the concept (1), which Anscombe defines as intention-for-the-future or prospective intention, refers to the fact that we commonly say “I intend to do X at time Y,” where Y is still to come. I can have such an intention for a very long time before I take any actions to bring about X; for example: as a sixteen year old I may intend to become president one day, and even though I take no immediate action towards that goal until I am considerably older, it remains the case that I intend to do so (become president). Conversely, prospective intentions could be only very slightly anterior to the performance of action, such as forming the intention to scratch my back and doing so (almost) immediately. As will become clear, there is much contestation about exactly *when* such an intention can be said to be present, and whether the presence of such intention requires that some action takes place at some point. For example, if I never actually take steps to become president, can I ever be said to have had the intention of doing so? It seems obvious that the intention entails a sense of commitment, but must that commitment manifest in action for it to qualify as such? Additionally, there is the question of whether intention-for-the-future should be understood as a mental entity as proposed by (Davidson 2001) and Bratman (1987), or as something else as is the case with Anscombe (1963).

The second use (2) refers to intentional action, meaning that we identify a given action as taking place due to *human agency of a certain sort*, as opposed to by accident or due to forces of nature external to the agent. For example: I lift my hand intentionally as opposed to it being lifted by somebody else, or it being lifted due to an involuntary reflex. I take the question of what exactly it is that is added to an action in order for it to count as intentional – that is, what is meant by *human agency of a certain sort* – to be the most important question that an account of intention must answer if it is to explain intentional action.

The third use (3) is intention-with-which, i.e. the intention I have when performing an action. This use of intention is often used in a teleological sense; that is, it is used to describe an action as directed toward the achievement of a certain goal. For example: I am reading an article by Davidson because I want to better structure my analysis of his arguments. Bayne (2010: 2) makes the additional point that this third use of intention is the only one that makes *necessary* demands on the forces of nature. In his words, “Sentences of this sort pertain to occasions where what we do intentionally is done *with the intention* of accomplishing something with the help of nature.” By “the help of nature” Bayne means that the success or failure of my enterprise in performing an action aimed at a goal depends not only on my own actions, but on the external world reacting in a way amenable to my success. If nature does not play along, so to speak, then my action will not succeed in achieving its goal.

The problem is that these three uses seem to be “not equivocal,” and our inability to explain the rules underlying the use of the term “intention” in these three different ways reflects that “we are pretty much in the dark about the character of the concept which it represents” (Anscombe, 1963: 1). Until these underlying rules are understood, a theory of intention that can convincingly account for all three uses is unachievable. For this reason, addressing the disparities in our use of the term remains one of the fundamental requirements of any convincing theory of intention. Perhaps unsurprisingly, the common strategy for resolving this difficulty is to explain the three uses of the concept in terms of one. Anscombe, for example, argues for an account that reduces uses (1) and (3) to use (2). Early Davidson sees the solution in bringing the uses together under use (3), while later Davidson placed (1) in the primary position. Bratman, though the least concerned of these three thinkers with the unity of the uses, gives an account that follows later Davidson in explaining intention almost solely in terms of use (1), although his approach is distinct from Davidson’s own.

To complicate matters even further, Wilson and Shpall (2012) contend that there is an additional use of intention for which Anscombe had failed to account. They state this use to be the following: (1) “in *Fing* (by *Fing*), the agent intended to *G*”, which they argue is related to but distinct from intention-with-which. Given that the most basic expression of intention-with-which is: “the agent *Fed* with the intention of *Ging*,” they employ the following example to illustrate their point:

[A]lthough it may be true that

(8) Veronica mopped the kitchen then with the intention of feeding her flamingo afterwards,

it normally won't be true that

(8') In (by) mopping the kitchen, Veronica intended to feed her flamingo afterwards.

The irreconcilable nature of these two statements is meant to show that the latter use of intention must be considered distinct from the first, and so should be added to Anscombe's list of uses. However, this would be a mistake. In both examples the two actions – mopping the floor and feeding the flamingo – are separate intentions, and so what is at stake here is not some new use that seems close to a case of intention-with-which, but rather a statement that includes an intention-for-the-future and an intentional action. The action of mopping the floor does not bring about or aid the progress of the action of feeding the flamingo, or if it does then the statement (8') would not be problematic in the sense Wilson and Shpall take it to be. The problem centres around the word “with” as it is used in the first statement, which seems to indicate a relation to intention-with-which, but this misses the fact that there is a difference between doing an action *with* an intention in the sense of “I am thinking of another intention I have while I am separately and intentionally performing my current action” and doing an action *with* an intention in the sense that “I am performing my current action with the intention of performing another.” The first type does not imply a connection between the two actions, the latter does. Assume a case where the flamingos cannot be fed until the floor is clean. In that case, it would be perfectly correct to say that, “in mopping the floor, Veronica intended to feed her flamingos afterwards,” precisely because here the two actions are linked, and the intention to feed the flamingos is the intention *with which* Veronica mops the floor. Let us call this case (8''). Assuming, as Wilson and Shpall do, that no such necessary connection as in (8'') exists between Veronica's actions, we are left with not one new use of intention, but a statement that expresses two different uses of intention, each use relating to a given intention. The mopping of the floor is a case of intentional action: “I am (intentionally) doing X,” and the feeding of the flamingo is a case of intention-for-the-future: “I intend to do X.” In the case of (8'), intention-for-the-future has simply been given in a reported form: “she intends to do X.”

What this analysis shows us is twofold. First, it reinforces Anscombe's position concerning the three – and only three – uses of intention. Second, and more importantly, it raises the fact that we can, and commonly do, hold multiple intentions. Statements such as (8') are complexes in the sense that they contain multiple intentions. The statement as a whole is not an example of any *one* use of the concept of intention, but each intention in the statement is itself an example of one use of the concept. This is a concern that Anscombe herself does not address directly. It is likely that there are rules governing the relationships between multiple intentions (such as are present in (8) and (8')) that we ought to try to explicate. This question is best addressed not under the criterion of the unity of the three uses, but as part of a consideration of the relationship between intention and practical reason. At this point, it is enough merely to note the existence of these complexes, as this supports the view that there are only three uses of intention.

## **2. Explaining intention's epistemic requirements**

All accounts of intention make at least some epistemic assumptions. It seems largely undisputed that to hold an intention requires some level of knowledge on the part of the agent. Although the nature, quantity and quality of this requisite knowledge is a matter of contestation, what is not contested (at least not by any of the three thinkers examined in this thesis) is that an agent who performs an action X without any knowledge of having done so cannot be said to have had an intention to do X. Consequently, the question that a convincing account of intention must answer is: How much (and what) does the agent have to know about herself (and the external world) for it to count as having an intention or acting intentionally?

It should be immediately obvious that there are two kinds of knowledge at issue here, or more correctly, knowledge about two different things: "self" and "the external world." In order to answer the question mentioned above, it is first necessary to gauge whether the difference between knowledge of "self" and knowledge of "the external world" refers to the gap between my physical person and the external world, or between the mental and the physical (where the physical includes my body). Anscombe takes the former approach, whereas Davidson and Bratman take the latter. This is important as it pertains to the issue of certainty. The problem at stake is how certain I must be in order to claim an intention, as well as about what I must be certain about. Any attempt at defining the epistemic requirements for

intention falls on a spectrum, with complete certainty at one end, and complete epistemic opacity at the other. An account of intention that occupies a position closer to the former end of the spectrum requires an agent to have access to some kind of special or privileged knowledge. However, as we will see, this view has serious difficulties in accounting for the nature and possibility of such knowledge. Accounts of intention that lie more toward the middle of the spectrum allow for some epistemic opacity; that is, they allow that an agent may not be aware of all the epistemically relevant aspects of the action. The question then becomes how much, and what type, of epistemic opacity a given account of intention allows. For instance some theorists – notably Davidson – claim that intentions are rational judgements to which agent may not necessarily have complete access. Besides having to explain how such judgements are possible, such accounts face the further difficulty of having to explain cases of irrational intentions (such as in some cases of *akrasia*<sup>7</sup> and pathological behaviour). Another approach – which is closer to the epistemic opacity end of the spectrum – is to tie intention to belief rather than knowledge. Such accounts claim that the only epistemic requirements for intention are that it must be consistent with the agent’s pre-existing beliefs. Such accounts, however, face the difficulty of having to explain cases where an agent has an intention X, but no belief to X, or where an agent has a mistaken belief regarding her intention.

### **3. Explaining the relationship between motivations, causes, reasons and intention**

As regards the relationship between motivations, causes, reasons and intention, there are three issues at stake. First, and tying back to what was said about intention-for-the-future, there is the question about the motivational status of an intention. Now it is a point of general consensus that intending to do X amounts to wanting to do X, and that even in cases where I do what I intend to do reluctantly, intention must be the dominant “want,” in the sense that it motivates me to act and no counter-force overrides it. In other words, there is nothing I can do, to my knowledge, that I want more (that *motivates* me more), or else I would be doing that instead. This raises the problem of whether or not it is possible to hold mutually exclusive intentions, what it is that differentiates an intention’s volitional quality from any

---

<sup>7</sup> The definition of *akrasia* that I use throughout this thesis is that provided by Davidson. He identifies *akrasia* as present when an agent “acts, and acts intentionally, counter to his own best judgement” (Davidson, 2001: 23).

other kind of motivation, and how to account for failed intentions. These are all questions that a coherent account of intention must attempt to answer or diffuse.

Second, are intentions to be taken as causes of the actions that follow them, or are those actions necessary components of the intentions themselves?<sup>8</sup> Davidson (2001: 3) argues that intentions are indeed causal explanations. Bratman agrees with this. Anscombe, on the other hand, while she does speak of “mental causes,” strongly denies that these are equivalent to intentions (Bayne, 2010: 53-54), or that the relationship between action and intention is a causal one (Anscombe, 1957: 322-325).<sup>9</sup> As I will show, the divergence between Anscombe and Davidson/Bratman on this point is intimately related to the different approaches they employ in trying to unite the three uses of intention.

Also in need of clarification in cases where intentions are understood as causal explanations is the issue of causal deviance (Davidson, 2001: 79). Causal deviance refers to the apparent possibility that an agent could have an intention to X, and this intention to X could cause the agent to perform X, yet the action performed would not be intentional. That is to say, it is not sufficient (if we embrace intentions as the causal explanations of actions) to simply have the intention and then perform the action; the action must be performed in the “appropriate” way, where the meaning of “appropriate” is what is at stake. An illustrative example taken from Setiya: “if I intend to be shaking to signal my confederate, and this intention makes me nervous, so I shake, I am shaking because I intend to do so – though not intentionally” (2011a). This is a problem as it means that any account that takes an intention to be the cause of an action must explicate the rules for when this causation does or does not bestow intentional status on the action. This clearly demonstrates that there is a need to clarify the “correct” causal path from intention to action.

Finally, are intentions to be understood as some species of reason or rational judgement? The question here is not whether or not intentions are linked to a process of reasoning, but whether intentions themselves should be identified as being a kind of reason. If intention is a

---

<sup>8</sup> Or could it be that intention follows action, as the famous experiment by Libet, Gleason, Wright and Pearl (1983) purports to show. I take the answer to be no, for two reasons: First, the method that Libet et al. employs for determining the moment at which conscious intention is present is flawed (see Haggard and Libet 2006, and Gomes 2002). Second, if intention is understood as a brain state, then the experiment can be interpreted as in fact providing evidence that intention *does* play a causal role in action. I will not be addressing these arguments in this thesis.

<sup>9</sup> Mental causes should not be confused with mental entities, the existence of which she rejects as nothing more than grammatical notions (Bayne, 2010: 51).

reason then there are issues to resolve concerning the possibility of irrational intentions (as are present in certain examples of *akrasia* and pathological behaviour). The reason this difficulty arise is because if intention is taken to be a reason, or a rational judgement, then it seems that it is impossible to have an *irrational* intention. However, this conflicts with our everyday experience of irrational intentions. There is also the danger of permitting illegitimate bootstrapping in reasoning, where an individual may have a reason X (about which I could be mistaken – that is, it could be irrational) which, through a process of practical reasoning, leads to the formation of the intention Y, and then that intention Y (even if it turns out that it is actually founded on an irrationality) will count as a reason Z in future processes of practical reasoning. This would then mean that having an intention is a reason in favour of performing that intention, which is clearly a matter of circularity, or pulling ourselves up by our bootstraps.

#### **4. Explaining the relationship between intention and practical reasoning**

It is not my aim in this thesis to present a complete or novel account of practical reasoning. For the purposes of my own argument, I will simply avail myself of the following general definition of practical reasoning from the Stanford Encyclopedia of Philosophy:

Practical reason is the general human capacity for resolving, through reflection, the question of what one is to do. Deliberation of this kind is practical in at least two senses. First, it is practical in its subject matter, insofar as it is concerned with action. But it is also practical in its consequences or its issue, insofar as reflection about action itself directly moves people to act. (Jay, 2009)

Given this definition, I think that it is quite obvious that intention is related to practical reasoning. After all, one can scarcely claim to have undertaken an action as the conclusion of a process of practical reasoning and then claim that action to be unintentional. So what is not in doubt is that there is a relation. What is in doubt is the nature of this relation. We can ask three important questions in this regard:

First, and building on the discussion of the relationship between intention and reasons, are intentions to be understood as reasons for the purposes of practical reasoning (as later Davidson seems to indicate), or should they be understood as something else (as argued by

Bratman)? If they are the former, then the problem of bootstrapping arises. Either some account of intention as reasons must be presented that avoids this circularity, or intentions must be treated as discrete from reasons for the purpose of practical reasoning.

Second, a convincing account of intention must be able to clarify the rules governing the relationships *between* intentions. It is my contention that the rules that govern such relationships are an aspect of the practical reasoning process, and as such a discussion of intention's relationship with practical reasoning requires an account of the relationships between intentions.

Third, and finally, any convincing account of intention must avoid the charge of "cognitivism" (Bratman, 2009a: 15), which is to say that reasons for belief and reasons for action (and intentions) are collapsed into one category. This collapse can be described as the collapsing of two forms of rationality, namely practical (reasons for action) and theoretical (reasons for belief), into one. The argument is that if to have an intention is to have a belief about oneself as committing an act (or going to commit an act), then the reasons that must result in the formation of an intention are the same as those at stake in the process of belief-formation. This would mean that an agent's reasons for intentional action would be the same as their reasons for belief. And since reasons for belief is the domain of theoretical reason, practical reason (asking after the reasons for acting) is then "skipped," or subsumed by theoretical reasoning, since one would have moved straight from belief to intention.

Now it may not seem apparent at first why this would be a problem. So what if practical reasoning is nothing but a subset of theoretical reasoning? The answer to this question will depend on the extent of "cognitivism" that one allows in a given understanding of intention. There are different degrees of cognitivism that a theory concerning intention and/or practical reasoning can have. To put this more clearly: the more cognitivist the theory of intention the smaller the unique role played by practical reason in the description of reasoning about action. Thus a wholly cognitivist account of intention would describe the formation of the reasons for action and the intention to act purely in terms of theoretical reasoning. Such an account therefore effectively rejects the idea that there is any qualitative difference in reasoning about beliefs or reasoning about actions. Such a cognitivist approach cannot properly account for intention and reasoning about action for two reasons: First, it requires an understanding of the relationship between intention and belief that some theorists – most

notably, Bratman (2009a; 2009b) – have argued is impossible, and second, that it is missing a crucial element that only practical reasoning can provide, namely: the actual commitment to act.

For this reason, any account of intention, to be convincing, must either answer these two difficulties, or must avoid endorsing cognitivism. This is particularly relevant for accounts of intention that seek to argue that intention should be understood as a belief, or that intention necessarily entails belief, as this position implies that reasoning about intentions is equivalent to reasoning about beliefs. Such intention-as-belief accounts – of which those of Velleman (1989), Harman (1976) and Setiya (2010) are the most notable – must either both endorse cognitivism, and argue against both of two difficulties that afflict it, or must argue that understanding intention as belief does not lead to cognitivism.

## **5. Explaining the relationship between intention and (moral) responsibility**

The final criterion I propose for testing the validity and coherence of theories of intention is whether they are able to explicate the relationship between intention and questions of responsibility (which is linked to the question of consequences). This entire criterion can be encompassed by two questions: Am I morally responsible for all the consequences of my intentional actions? And, am I responsible for anything I do unintentionally? The first clue I would put forward in resolving these questions is that the everyday usage of the term, “unintentional” carries with it a strong connection to the term “accidental,” a fact that does not appear to me to be at all an accident! It seems to indicate that intention is more than contingently related to responsibility.

To find some clarity on this, it is first necessary to determine what kind of responsibility – causal responsibility, moral responsibility or yet some other kind – is at stake. *Causal responsibility* refers to whether or not a given entity X can be described as (even partly) the cause of event Y. So the water cycle is causally responsible for rain, and when I kick a stone on the road I am causally responsible for its movements following that kick. Of course, few (if any) events have only single entities as causes. In both of the above cases there are obviously various entities playing roles of various extents in bringing about the event in question, and each one of these can be ascribed causal responsibility. This also means that chains of causal responsibility are long, very long. My kicking a stone and then causing an

avalanche which brings about a regime change in the country in which I did so are all things for which I can be called causally responsible, although the more intervening steps the smaller my relative contribution, and hence the smaller my causal responsibility (i.e. the more likely it is that the event would have taken place without my contribution at all, which is the limit of causal responsibility).

*Moral responsibility* refers to the fact that we hold people accountable for certain actions in a very particular way, where the agent “should not” have done whatever it was she did (this is also the context in which we use the terms “wrong” and “evil”). Causal and moral responsibility are distinct, though interrelated, concepts: moral responsibility always requires causal responsibility (I cannot be called morally responsible for what I had no contribution in bringing about or maintaining), whereas in the vast majority of cases causal responsibility is amoral (of no moral value). The consequences of my involuntary breathing may be vast, and I would be causally responsible for all of them (to various degrees), but few would hold an individual morally responsible for such consequences, and we would think that those who do so are mistaken. In short, we are not held morally responsible for what we do *by accident*. On the other hand, it would be truly unusual to treat an intentional action as not open to moral responsibility.

In light of this, and in so far as unintentional and accidental seem to be necessarily related, I will, in this thesis, be assuming that intentional action can be taken as necessarily open to moral responsibility.<sup>10</sup> This means that moral responsibility must apply either to some of, or all the, consequences of intentional action. Of the three thinkers I will be examining, only Anscombe deals directly with this problem, which she does by way of a particular understanding of the doctrine of double effect.<sup>11</sup> This doctrine states that “it is sometimes less morally objectionable knowingly to bring about (or allow) some bad effect in the course of achieving some good end, than it would have been to bring about (or allow) that bad effect with the *intention* to do so in order that a good end is achieved” (Chan, 2000: 405). Her solution, as I understand it, is built on a distinction between *intentional* as opposed to merely

---

<sup>10</sup> My assumption of the necessary relationship between moral responsibility and intentional action is, admittedly provisional. There is undoubtedly much more to say on this topic. However, as none of the three thinkers I analyse in this thesis oppose this assumption, and as it seems to be in line with general intuition, I do not consider the assumption to be without basis.

<sup>11</sup> Anscombe also discusses causality more specifically, primarily as part of a critique of Hume’s approach on the subject. I will not deal with this facet of her work; however, my argument requires only that we recognise that there is a difference between those consequences of our actions that carry moral value and those that do not.

*foreseen* consequences to actions, where only the former is certainly a locus of moral responsibility, but intention itself is expanded beyond the limits assumed by many forms of the doctrine of double effect (Anscombe, 1961). I discuss this argument in more detail in my treatment of Anscombe's account of intention in Chapter 2. In the cases of Davidson and Bratman it will be necessary to deduce an account of moral responsibility from the implications of their accounts of intention.

One of the most fascinating things about the concept of intention is that it plays a role in ascriptions of moral responsibility in virtually any ethical theory, with only one general exception: most forms of consequentialism. Although this might not be true for all and any consequentialist theories (it is far too wide a field to make such a pronouncement), it is true at the very least for utilitarianism. In practice, as my earlier discussion of causal responsibility as opposed to moral responsibility was intended to highlight, there are ways to be held responsible that do not involve a moral dimension. However, be it within a deontological framework, a form of virtue ethics or a contractarian theory, moral culpability and the basis of punishment are intrinsically linked to the intentional status of the action and the agent. My hope then is that my account of intention, by helping to understand the nature of this relationship between intention and moral responsibility, can provide insights helpful to discussions on ethics regardless of the ethical theories at stake (provided the ethical theory is not utilitarianism!).

I should emphasise that I am not interested in attacking or defending any particular moral theory, or in relating moral responsibility to any particular moral theory. My aim rather, is to understand what enables us to say "X performed an intentional action Y," and then to understand how that statement relates to the statement, "X *can be taken to be* morally responsible for Y," where X and Y are not specified. Stated differently, various ethical theories will hold various types of actions worthy of moral responsibility, but regardless of the exact actions in question there seems to be a further requirement involving the intention of the agent when it comes to the ascription of moral responsibility. It is this second requirement I seek to address.

## **Conclusion**

My goal in this chapter has been to identify and describe the requirements for a convincing account of intention. I have identified the following five requirements in light of which I will appraise the accounts of intention proffered by Anscombe, Davidson and Bratman: (i) the unity of the three uses of intention; (ii) explaining the relationship between knowledge and intention; (iii) explaining the relationship between intention, reasons and motivations; (iv) explaining the relationship between intention and practical reasoning and, finally, (v) explaining the relationship between intention and responsibility. What will become clear as the thesis progresses is that, although each of the accounts of intention I examine provides important insights with regard to one or more of these requirements, none of these accounts satisfies all five of the specified requirements. In each case, I will identify a number of difficulties that remain to be resolved. Having done so, I then attempt to develop an account of intention that successfully meets all the specified requirements and, crucially, overcomes the various difficulties that plague the foregoing accounts.

## CHAPTER 2: ANSCOMBE'S ACCOUNT OF INTENTION

### Introduction

It is not much of a stretch to say that the modern conversation about intention started with Anscombe's arguments in her article, *Intention* (1957), and her expansion on this work in a monograph of the same name, *Intention* (1963). She herself, of course, draws upon an earlier tradition, notably the work of William James and Wittgenstein (Bayne (2010: xvii-xxiv). However, the way in which she uses certain critical concepts, and her arguments, are largely original. These remain the definitive starting points for subsequent attempts to engage with the question of intention. Her recognition of the problem regarding the three guises of intention is still one of, if not the, most important cornerstones of the contemporary conversation about intention. Her claims about the relationship between self-knowledge and intention, and the nature of such self-knowledge (Anscombe, 1957: 322), inaugurated the discussion about the epistemic requirements for intention, while her distinction between reasons and causes would become one of the most contested conceptual battlegrounds in the field. Many of her concepts and distinctions would be adopted in an altered form by Davidson in his own account of intention, which would go on to become the dominant position on the topic of intention.

Before discussing Anscombe's account of intention further, there is an important point to make regarding her definition of intention, as it influences much of her account, including the unity of the three uses. For Anscombe, intention can be understood as behaviour about which it is meaningful to "raise the question 'Why?' in the sense of inquiring into reasons for acting" (Bayne, 2010: 15). In other words, "what we do for reasons, we do intentionally" (Setiya, 2010: 171). A good example, provided by Julia Driver, to clarify Anscombe's meaning runs as follows:

[W]hen someone knocks a glass off of a table he may give an explanation that he saw a face in the window and that made him jump. This provides a causal explanation for why he knocked the glass off the table, but it doesn't give a reason. The knocking of the glass off the table was not intentional, though it was caused by his being startled. (Driver, 2011)

This “Why?” question will play a significant role throughout Anscombe’s account, as well as all the accounts I examine in this thesis.

In this chapter, I will examine Anscombe’s account of intention in relation to the five criteria specified in Chapter 1. In Section 1 of this chapter, I show that, in order to achieve the unity of the three uses, Anscombe comes to the conclusion that the three uses can be reduced to intentional action. This collapses intention-for-the-future and intention-with-which into intentional action. I further show that this is an unacceptable solution to the problem of the three seemingly irreconcilable uses of intention, as it fails to account for “pure intending.”

In Section 2, I examine Anscombe’s attempt at clarifying intention’s epistemic requirements. This is the matter that Anscombe seeks to tackle when she makes the assertion that “[i]ntentional actions are a sub-class of the events in a man’s history which are known to him *not* just because he observes them” (Anscombe, 1963: 24). She argues that that which we do intentionally (and hence knowingly), we know we do without observation. This is a very stringent epistemic requirement, and one that I will show cannot hold.

In Section 3, I examine how Anscombe’s account deals with the relationship between intention and motivation, causes, and reasons respectively. I discuss the last of these relationships – intention and reasons – first, as it plays an important role in Anscombe’s account. Her argument is that a case of intentional action is characterised by being a case where the agent responds to a certain meaning of the question “Why?” with a rational explanation. This is a position that is widely accepted, and I will employ it in my own account of intention in Chapter 5. On the question of the relationship between intention and causes, Anscombe first differentiates between *reasons* and *causes*, and then argues that intention is certainly not a cause. As I will show, the latter assertion is mistaken. Anscombe similarly denies any necessary relationship between intention and motivation. Once again I will argue that she is wrong to deny such a relationship.

In Section 4, I examine Anscombe’s argument that, since intention is not a motive, it is not necessarily related to practical reasoning. I will argue that this view rests on a mistaken understanding of intention’s relationship to motivation, and a misguided commitment to the primacy of intentional action over the other two uses of intention. I argue further that she does not provide an adequate (or indeed, any) explanation of the relationship between

multiple intentions. In the final part of this section, I acknowledge that Anscombe's account avoids the charge of cognitivism, but only because it contains elements that result in fatal flaws with regards to other aspects of this account.

The fifth and final section of this chapter examines Anscombe's treatment of the relationship between the moral relevance of an action or consequence – whether or not the agent is morally responsible for the action – and its intentional status. She argues that there are three types of consequences: intentional, foreseen and unforeseen, and that intentional and foreseen consequences are morally relevant, while unforeseen consequences are not. I discuss her distinction between the three types of consequences and the two measures of moral relevance she advances for intentional and foreseen consequences respectively.

### **1. The unity of the three seemingly irreconcilable uses of intention**

In analysing Anscombe's argument for the unity of the three uses of intention, I focus, first, on outlining the approach that she adopts in order to achieve this unity, namely: the reduction of the three uses to intentional action. Following this, I look at the difficulties that this approach faces. I hope to make clear that Anscombe's approach is not able to overcome these difficulties, and that the flaws that prove fatal for her account applies to *any* attempt to unite the three uses of intention through a reduction to intentional action.

#### *1.1 The reduction of the three uses of intention to intentional action*

Anscombe begins her account of intention by describing the three seemingly distinct ways in which the term intention is employed. To repeat, these are: the use of intention as a verb when referring to an intention toward a future action (when I say, "I intend to complete my thesis"), the use of intention as a noun when referring to a teleological explanation of action (when I say, "my intention in reading that book was/is to learn the plot") and finally the use of intention as an adverb to describe an action (when I say, "I read the book intentionally" or "I intentionally read the book"). The problem that Anscombe identifies is the following:

Realising this [the existence of the three uses] might lead us to say that there are various senses of 'intention', and perhaps that it is thoroughly misleading that the word 'intentional' should be connected with the word 'intention', for an action can be

intentional without having any intention in it. Or alternatively we may be inclined to say that ‘intention’ has a different sense when we speak of a man’s intentions *simpliciter* – i.e. what he intends to do – and of his intention *in* doing or proposing something – what he aims at in it. But in fact it is implausible to say that the word is equivocal as it occurs in these different cases. Where we are tempted to speak of ‘different senses’ of a word which is clearly not equivocal, we may infer that we are in fact pretty much in the dark about the character of the concept which it represents. (Anscombe, 1963: 1)

The problem can be succinctly summarised as follows: If we assume that intention is not an equivocal concept, then how do we explain the three seemingly irreconcilable uses of this concept? It is this problem that Anscombe takes as the departure point for her account, and as I have stated in Chapter 1, it is a problem I think a convincing account of intention must answer. Anscombe’s attempt at a solution, which is the focus of this section, is to argue for the reduction of three uses to intentional action. The core of her argument is that both intention-for-the-future and intention-with-which are only examples of two different types of intentional action, both possessing elements that differentiate them from each other, but both containing the full gamut of elements required for intentional action. Said differently, it is the presence of intentional action that permits the self-ascription of intention by the agent. In the case of intention-for-the-future it is an obvious matter of reduction: to have an intention-for-the-future there must intentional action or no intention at all. In the case of intention-with-which, it is a matter of arguing that there is no “sharp distinction between ‘I am doing A’ and ‘I am going to do A’ offered as answers to the question ‘Why are you doing B?’” (Anscombe, 1963: 40). The end result of this approach is to treat “having an intention” as being in progress towards the intentional completion of an act.

There are a few forms that such an argument can take. The first would try to stress the open nature of the idea of “doing” – that is to say, if I reach the mental resolution to bake a cake I am already in progress towards doing so. On this understanding, my intentional action starts the moment I have formed the resolution to act. From this point onwards, any failure to achieve or fulfil the resolution is not fair grounds for the dismissal of the *self-ascription* of intention, but rather an example of a *failure to fulfil* my intention, nothing more. A second possible form of the argument for collapsing intention-for-the-future and intention-with-which into intentional action – keeping in mind Anscombe’s ordinary language

methodology<sup>12</sup> – could start as follows: The ordinary use of the term intention seems to be intimately linked with the notion of activity. The notion of an intention seems to require *something* to be intended; an intention is therefore always *directed* in this sense. Stated differently, it seems like a contradiction to speak of an aimless intention, or an intention to nothing. This last point raises the very interesting matter of omissions and how they are to be understood in relation to “actions.” After all, it is certainly not a contradiction to have an intention to *do nothing*, but this is different from having an intention to nothing. The first evokes the idea of an omission, whereas the second seems to be an admission of not having any intention at all.

The exact understanding of omissions is a grey area all of its own and falls outside the scope of my argument. The one important point to understand in this regard is that an omission can be differentiated from other sorts of *not doing something*. When I do not jump into the air with every third step I take, that is a case of me not doing something, whereas if I am told that if I want to have my thesis evaluated I must actually write it and then do not, it is a matter of *not doing something* AND omission. In fact, I would argue that all examples of omission are examples of *not doing something*, but there are cases of *not doing something* that are not omissions. The difference is that cases of omission entail intention, whereas cases of *not doing something* that are not omissions are identifiable exactly in that they do not entail intention. If this is accepted, we can then speak of a difference between having an intention to do nothing as tied to omission, whereas having no intention to do something *and* not doing it is an example of having no intention.

This ordinary language method would seem to indicate that intention cannot be ascribed where there is no action involved, so that any instance of intention-for-the-future must be more than a mere recognition of a desire: it must indicate *progress toward* that desire. There must be some action linked to the intention, not in the mere sense that the intention is associated with an action, but in the sense that it implies an *active* response from the agent, otherwise it would seem incorrect to call it an intention. In the case of intention-with-which, the argument would be that the action embarked upon constitutes a part of the progress

---

<sup>12</sup> Following Wittgenstein, Anscombe purposefully avoids ascribing ontological status to ideas wherever possible, treating them instead as purely grammatical. This approach lends itself to “ordinary language methodology” (Driver, 2011). This means that one examines how we use terms such as “intention” in everyday life, and then draws conclusions regarding the rules for the use of such concept’s connotation without making any claims about the independent existence of that to which they refer.

toward the final intention, so that the action identified as the outcome of the intention is the actual intentional object. An example will help: “I am pumping the water with the intention of poisoning the well.” In this case, the action of “poisoning the well” would then be the intentional object, and the subordinate action of pumping the water is part of the progress towards the intentional object.

Setiya (2011a) also claims that Anscombe’s approach has the advantage of explaining the “unity of what Thompson (2008: 97-99, 118-119, 132-134) calls ‘naïve’ and ‘sophisticated’ rationalization.”<sup>13</sup> By “naïve” rationalisation is meant that we sometimes explain actions by action, as in “I am boiling the water because I am making tea,” whereas by “sophisticated” rationalisation is meant that sometimes we explain actions by intentions (or wants), as in “I boil the water because I intend (want) to make tea.”<sup>14</sup> The problem that Thompson identifies is that we mesh these two methods of rationalisation all the time, explaining actions by intentions and intentions by actions. What is the relationship between explanation by actions and explanation by intentions that facilitates this interchangeability?

Anscombe’s reduction of the three uses to intentional action could provide an answer to this question. If to have an intention always means to be in progress toward the completion of a given action, then the interchangeability of action and intention seems obvious: when we refer to the action the related intention is always implied, and vice versa. When I describe myself as intentionally boiling water, it goes without saying that this intention is tied to the action of boiling water. I *cannot*, on this view, have an intention without simultaneously making progress in the relevant action. However, this does not mean that the two terms are synonyms. Although all cases of intention imply action, not all cases of action imply intention. Action may be a necessary requirement for intention, but intention is not necessary for action. This is evinced by the very possibility of unintentional action.

Anscombe’s account could also give us clarity concerning the exact relation between intentional action and intention-with-which. In Setiya’s (2011a) words: “An intention with which one is doing A is an intentional-action-in-progress that explains one’s doing it.” With

---

<sup>13</sup> Thompson’s explanation for the use of the terms “naïve” and “sophisticated” can be found in his comment that “naïve” rationalisation is more commonly employed in everyday life, whereas the “sophisticated” variety is “a form which, if it is less common in life, is all the same much more common in the pages of philosophy” (Thompson, 2008: 97).

<sup>14</sup> Or more clumsily, but still possible: “I am intentionally boiling this water with the intention of making tea.”

this addition, we could claim to have resolved the difficulties caused by the identification of the three seemingly irreconcilable guises by reducing both intention-for-the-future and intention-with-which to forms of intentional action.

### *1.2 The problem of pure intending*

There is something troubling, however, about this demand for action as a necessary component of intention. What about cases of what Davidson (2001: 83) calls “pure intending?” An example: “Someone may intend to build a squirrel house without having decided to do it, deliberated about it, formed an intention to do it, or reasoned about it. And despite his intention, he may never build a squirrel house, try to build a squirrel house, or do anything whatever with the intention of getting a squirrel house built.”<sup>15</sup> The issue here is whether this is an instance of (a) fully-fledged intention, (b) failed intention or (c) no intention at all.

A proponent of Anscombe’s approach would have to choose either (b) or (c), since choosing (a) would undermine the position that all instances of intention must reduce to intentional action. The third option would be the easiest to choose, but that seems to require us to abjure our ordinary use of the term “intention,” which Anscombe would be very hesitant to do. The difficulty is that we *do*, at least sometimes, think of a case such as the one presented by Davidson as being case of intention.

This leaves the second option, that the case presented is that of a failed intention. Recall that the reduction to intentional action was made possible, at least in part, by adopting a very wide understanding of the notion of “in progress.” In order to incorporate all cases of intention-for-the-future, it is necessary to place the beginning of being “in the progress of an action” at as early a point as possible. So the moment that the intention is “in the head” is also the moment you first count as being in the progress of an action. If you deviate from the course of that progress, and hence do not accomplish your intention, even if that deviation is that you never do anything other than have the intention in your head, it is still a case of failed intention. However, this seems a serious stretch! In the Davidson example it would be a very unusual individual who would claim that the potential squirrel house builder was *ever* in the progress of the action of building a squirrel house. Does the notion of action not imply that it must be

---

<sup>15</sup> A more extensive discussion of Davidson’s argument can be found in Chapter 3, Section 2.

more than what is in the head? This is not to dispute that Davidson's example is one of a failed intention (Davidson himself would doubtlessly agree); the issue has now become whether the term "action" has any applicability. Might it not be the case, in other words, that the successful *fulfilment* of an intention might necessitate action, but the mere *presence* of an intention (and remember, even if it is a failed intention, it remains an example of intention) does not inherently require it? This is a battleground that will be discussed in greater depth in the section on Davidson's approach to the problem of the three uses in Chapter 3.

### *1.3 The problem of plural intentions*

A second difficulty with Anscombe's approach concerns the complex ways in which we use the concept intention, which often makes it difficult, if not impossible, to identify which action is actually taking place. If we cannot ascertain this, then – in terms of Anscombe's argument – we must forsake the claim that intention is present. Setiya (2011a) provides an example of the kind of complex statement that causes problems for Anscombe's account: "I intend not to be hit by a car as I walk home". In this case, there is a clear intention to walk home, and an intention not to be hit by a car. These intentions are not related in the sense that pumping the water and poisoning the well are related, nor are they as unrelated as the case of mopping the floor and feeding the flamingos (assuming the situation is not like the one I sketched in (8'')) (see Chapter 1, pp. 7-8). The problem here seems to be that there is no clear rule that governs the relationship between these two intentions.

Anscombe never answered this objection in her work. In the rest of this section, I will attempt to construct a plausible defence of her position in order to see how it holds up. The part of the sentence that poses the greatest difficulty for identifying the relevant action is the phrase: "I intend not to be hit by a car." The additional part of the sentence, "as I walk home," expresses both a second intention "to walk home," as well as contextualising the first intention. With regards the objection, that it is difficult to know what action is taking place – at least as far as the first part of the sentence is concerned – one answer that Anscombe could give would be to say that this objection results from misunderstanding (a) the nature of actions of avoidance and (b) the demarcation of action. By "actions of avoidance" I mean "action(s) Y undertakes

to avoid some outcome X,” such as “...not being hit by a car as I walk home.” By the demarcation of action is meant the way in which we differentiate one action from another.<sup>16</sup>

Anscombe could then say that actions taken to avoid an outcome are no different from any other actions as far as the presence of intention is concerned. There does, however, seem to be an inherent vagueness regarding the actual actions to be taken to accomplish such intention. However this is only apparently a problem, brought about by the proportionately larger role played by external forces in deciding the success or failure of the intention. When I mow the grass, what I require from nature (understood as “requirements external to the mind and to bodily movement”) in order for my intention to succeed relative to my own contributions stands at a certain proportion, let us say X to Y, where X is nature’s contribution and Y mine. In the case of avoiding the car it could be said that this relationship is now 2X to Y, with nature’s contribution being the more important contribution. This proportionately greater contribution from nature can lead to vagueness or uncertainty regarding which actions I should in fact take to accomplish my objective since, while I know my goal, I might be unsure how to achieve it.<sup>17</sup> It is merely a contingent fact of our universe that acts of avoidance *tend* to entail greater uncertainty. Anscombe could therefore argue that all actions are characterised by such uncertainty, as all action requires the help of nature. As such, there is always some uncertainty about the actions I am or should be undertaking in pursuit of my goal. The difference between the two actions in our example is therefore one of the quantity, not the quality, of certainty. In the case of avoiding the car, the actions to be taken would be contextually dependent on the situation: walk in the right direction, do not drive into oncoming traffic, etc. The fact that we cannot tell from a theoretical perspective exactly which actions the agent should take is not a concern, since in a sense we could never be certain in any given case. We can merely have various degrees of confidence in our guesswork.

---

<sup>16</sup> Of course this becomes tremendously complicated, as most (perhaps all) actions are actually complexes of numerous subordinate actions. This is obvious in cases like “building a house” where in order to complete the action I must perform a myriad other actions such as “mixing the cement” and “laying the bricks.” However even the simplest of actions has this structure if we attribute the term action to muscular movements. Whether we do so or not, or to what extent we do, is an intriguing question, but not one that I will pursue here. The point I want to make is that we *do* differentiate between the individual actions that make up a more complex action, and that these actions are themselves composed of identifiable subordinate actions.

<sup>17</sup> Note that I do not mean to imply that intention as my goal can be talked about separately from the action I am to take to accomplish it. Intention cannot be identified purely as a goal; otherwise it would amount to nothing more than motivation or desire. Intention does seem to require, at the very least, a “plan of action” even if the action never eventuates.

The second element of Anscombe's hypothetical defence against objections based on the complexity of our use of the term action – a defence I have characterised as the demarcation of action – is straightforward: We have no problem identifying what action is at stake when a person says “I intend to run for President,” despite the fact that this intention is far more vague in terms of the actions it demands than in the case of the car avoider. The difference seems to be that, in the case of former, we accept that a broad series of subordinate actions will be necessary to achieve the goal, and that we will not know exactly what these will be, although we may hazard a few guesses. In the case of the car avoider, Setiya seems to assume that we must have some concrete idea of what actions should be the subordinate actions taken to achieve the goal. But there is no reason why the car avoider case should be considered differently from the case of the presidential hopeful in this regard. Just as with the presidential hopeful, I could specify a series of subordinate actions that comprise the actions that the car avoider performs. These would be open to the same uncertainty as the subordinate actions required by the presidential hopeful. I take this to mean that Setiya's assumption that we are required to have a concrete idea of what the subordinate actions should be in the case of the car avoider, but not in the case of the presidential hopeful, is unfounded.<sup>18</sup>

In conclusion, Anscombe's account of the unity of the three uses still has two unresolved difficulties: the matter of “pure intending” and the case of certain complex expressions of intention. The former seems a truly insurmountable hurdle, while the latter results from the fact that Anscombe did not address complex expressions of intention in her work at all, which leaves her account incomplete. Difficulties of the first sort – those concerning “pure intending” – would apply to any attempt to unite the three uses of intention through a reduction to intentional action. I therefore take the failure of Anscombe's account to indicate that any account that follows this approach would be unworkable. This being said, it remains to be seen if any other method of reduction can resolve the issue of the three seemingly irreconcilable uses, or whether the problem Anscombe identified in her watershed paper remains unsolved and, perhaps, unsolvable.

---

<sup>18</sup> More challenging in this regard is another example provided by Setiya (2011a): “I intend to drink wine or beer with dinner.” In this case the only remedy seems to rely on a discussion of the relationship between multiple intentions, which I will only tackle in Section 4 below.

## 2. Epistemic requirements

In this section I investigate Anscombe's understanding of the epistemic requirements for intention. This requirement is that intentional action must be known to the agent performing it, *without observation*. I then raise two objections to this approach, and argue that Anscombe's account is incapable of overcoming them. I then briefly discuss the position of intention-as-belief, a position that attempts to salvage useful elements of Anscombe's epistemic requirements, while abandoning those that are untenable. Finally, I examine Anscombe's conception of "acting under a description," where I conclude that this conception is useful in discussions concerning the ascription of intention, but not the definition of intention.

### 2.1 Knowledge of intention as non-observational

Perhaps no aspect of Anscombe's work on intention has received as much scrutiny as her position concerning the epistemic requirements for intention. The core of this position can be summed as follows:

*Anscombe's Principle:* If A has the capacity to act for reasons, she has the capacity to know what she is doing without observation or inference – in that her knowledge does not rest on sufficient prior evidence. (Setiya, 2010: 174)<sup>19</sup>

For the sake of a more extensive understanding of Anscombe's position in this regard, let us start by considering the most intriguing set of cases she uses to build her argument, namely cases of bodily movement. It is a relatively uncontroversial claim that any and all human action must entail bodily movements; in fact, it seems a matter of definition that bodily movement is to be involved whenever human action is taking place. This is important, as it implies that anything we say that applies to cases of bodily movement will necessarily apply to any and all human action, or, at the very least, to certain subordinate actions undertaken to accomplish the main action.

For Anscombe, knowledge of such bodily movement is a case of non-observational knowledge (Anscombe, 1963: 14).<sup>20</sup> She contends that in such cases there is no separation

---

<sup>19</sup> Note that this is Setiya's own formulation of Anscombe's view, which she herself never stated as succinctly.

between our sensation of an event and the facts of the event itself. Stated differently, in such cases the sensation is not mediated by some intermediary sensual apparatus that can be described separately. When I am asked, “Did you move your arm?” I do not need to check with my sense of taste, sight, smell, hearing or touch; I can answer with confidence in my knowledge that I did or did not move my arm. This, at least, is Anscombe’s view. This view follows from her assertion, as summed up by Bayne (2010: 43), that “I can know something by observation only when the sensation which goes with the observation is ‘separable.’ A sensation is said to be ‘separable’ when its ‘internal’ description can be distinguished from the description of the very fact, purportedly known by observation.” That is to say, I cannot provide a description of the sensation of the movement of my limbs without using a description of the actual position of my limbs. We must then conclude that our knowledge of bodily movement is non-observational but direct in some meaningful sense.

This does not imply that such knowledge disregards empirical reality. When I intend to clench my fist behind my back, but a temporary paralysis prevents me from doing so, I cannot claim to have knowledge of the movement of my hand. The point is simply that in those cases where I *do* have knowledge regarding bodily movement, this knowledge is never observational. To clarify: in the case where I have knowledge of my bodily movement I must first know that I am not suffering from paralysis, but once I know this, the knowledge of my bodily movements is not perceptual or inferential. In other words, I can still be wrong about what my body is doing, but in cases where I know what my body is doing, this knowledge is non-observational. In Anscombe’s (1963: 50-51) words:

[The topic] of an intention may be matter on which there is knowledge or opinion based on observation, inference, hearsay, superstition or anything that knowledge or opinion ever are based on; or again matter on which an opinion is held without any foundation at all. When knowledge or opinion are present concerning what is the case, and what can happen – say Z – if one does certain things, say ABC, then it is possible to have the intention of doing Z in doing ABC; and if the case is one of knowledge or if the opinion

---

<sup>20</sup> It is worth noting that the requirement of knowing, non-observational or not, should not be confused with the more narrow idea of “consciously knowing.” I know, for example, the names of my parents (or at least hope I do), and at times such knowledge is consciously before my mind, at other times it is not. More dramatically, when I sleep it is not usual to assert that I am conscious at all, yet it would seem strange to say that while I sleep I do not know who my parents are. The point of this is to show that though some knowledge I hold consciously at any given moment in time, other knowledge I do not, but this does not mean that the latter is not something I know. This point is worth keeping in mind for the entirety of this enterprise, it is not limited in application only to Anscombe’s account.

is correct, the doing or causing Z is an intentional action, and it is not by observation that one knows one is doing Z; or in so far as one is observing, inferring etc. that Z is actually taking place, one's knowledge is not the knowledge that a man has of his intentional actions.

Having presented her case for the non-observational status of our knowledge of bodily movement, she then extends this status to our knowledge of all our intentions. To be precise, she states that “the class of things known without observation is of general interest to our enquiry because the class of intentional actions is a sub-class of it” (Anscombe, 1963: 14). In the words of Bayne (2010: 45), Anscombe “appears to feel no discomfort in moving from what she says about knowledge of a state of our body to knowledge of actions.” This is a jump, to be sure, and one for which Anscombe provides preciously little justification. The main thrust of her argument is that, in all cases where our answer to the relevant question “Why?” takes the form of reasons (that is to say, in cases of intentional action), our knowledge of these (intentional) actions is non-observational. In fact, this becomes a method by which to identify cases of intentional action. To explain this it is useful to consider a vignette from *Intention*: In order for me to know what I am doing when I am pumping water into a house, I first have to know that the equipment is working properly. But once I know this and begin pumping, my knowledge of what I am doing is not known through observation.

## 2.2 *Objections to knowledge of intention as non-observational knowledge*

Needless to say, Anscombe's position on the epistemic requirements for intention has drawn considerable criticism from other thinkers, perhaps best expressed by Grice (1971) when he describes it as “licensed wishful thinking.”<sup>21</sup> There are few thinkers who attempt to defend Anscombe's argument as it stands. There is, however, a minority of action theorists who have appropriated various elements of her argument, with greater and lesser modification, in order to develop convincing accounts of intention. For the purposes of my own argument, I focus on one attempt at such an account, namely that of Keiran Setiya.<sup>22</sup> Before outlining Setiya's attempt to rescue Anscombe's views on the epistemic requirements for intention, however, I

---

<sup>21</sup> See Michael Bratman (1987), Keith Donnellan and Sidney Morgenbesser (1963), and Paul Grice (1971) among others.

<sup>22</sup> I have selected Setiya's approach both because I think it to be the most convincing and because it is a paradigmatic example of the general trends in defences of Anscombe's position.

will first set out the most important objections to Anscombe's argument so that we may ascertain whether Setiya's theories resolve all, some or none of these difficulties. These objections can be divided into two groups: (i) those that argue that we do in fact acquire bodily knowledge through observation (which builds up to a bigger argument concerning the possibility and scope of non-observational knowledge), and (ii) those that argue that even if such knowledge is direct, it cannot be linked to intentional action in the fundamental way Anscombe contends it does.

The most poignant argument presented by those raising objection (i), as I see it, is the straightforward point that "[s]urely, one wants to say, proprioception and kinaesthetic sensation play some role in informing the agent of the positions and movements of his body, and it is uncertain why these informational roles should fail to count as modes of inner 'observation' of the agent's own overt physical behaviour" (Wilson and Shpall, 2012). This objection takes the discussion into very difficult waters, as it seems to call for empirical testing to ascertain the actual directness of sensation.<sup>23</sup>

This is not at all Anscombe's own position, but a possible counterargument against critics who raise objection (i) against her claim that intentional action be known without observation, might run as follows: Let us accept that proprioception and kinaesthetic sensation are senses like any other, and that bodily movements are known through these senses. However, can I really describe my experience of the sensation of a bird alighting upon my car bonnet separate from the event itself? When describing the sensation I must, surely, employ descriptions of what I actually saw, and heard and smelt, all of which are (assuming I am not hallucinating) facts of the event. By this measure, all knowledge would be "non-observational" in Anscombe's sense of the term, a conclusion that conflicts with all normal usage of the term "observational knowledge." Thus the same point can either be used to argue that bodily knowledge is not non-observational or that at least some non-bodily knowledge is non-observational. This is just one of several loops that this fascinating subject can lead to, and though a thorough analysis of the matter might be engrossing, it falls outside the scope of my thesis. As I hope to make clear, it is possible to develop a convincing account of intention without first resolving this matter. Given this aim, I will not devote further

---

<sup>23</sup> As even the smallest of gaps between a movement and my knowledge of it would indicate that the sensation and the event are indeed separable, if testing for such a delay reveals that I could describe the sensation of movement with reference to the neural firing that serves as an intermediary between the event and my sensation, of it, then this would show that Anscombe is incorrect.

discussion to the relationship between bodily knowledge and observation, although I will note whether a particular thinker agrees or disagrees with Anscombe on this point when it is relevant.

Those who, like Davidson (2001) and Bratman (2009a), raise objections of type (ii) focus on the issue that our knowledge of what we do intentionally can never be anything but contingent, which means that Anscombe's claim that what we do intentionally we do knowingly (regardless of the observational or non-observational nature of this knowledge) cannot be upheld in all cases. The obvious proof of their argument would be any case where a person did not know what he/she was intentionally doing. Broadly speaking, such cases are not difficult to come by. For example; when I intentionally start my car I have knowledge of only part of what I "do," as I do not understand the inner workings of an internal combustion engine. But it could be argued that this is too broad a conception of "knowing," that I do in fact have full knowledge of my actual actions (the turning of the car key), and that only knowledge of my actual actions, as opposed to the reactions of nature to my actions, is necessary.

What about the case of tying one's shoes, however? I intentionally tie my shoes, yet when I do so I do not *know* the exact movements of my fingers. I could not, for example, duplicate the finger motions in the air, or describe them in detail without first analysing the movements myself (Bayne, 2010: 35-36). This seems to imply that I lack knowledge concerning even some cases of bodily movement. I know that I intend to tie my shoes, but I do not have full knowledge of what I am actually doing intentionally, even by the more narrow measure of "knowing." Another example – from Davidson – goes: I wish to make ten copies of a document using carbon paper, and so I press down with the imprint of what I write on the top page. I do not know whether this will imprint through to the tenth page, though I can believe it will or will not. What I do know is what I intend to do, what I do not know is what I intentionally do (Davidson, 2001: 92). Davidson's point here is that, while my knowledge of what I intend to do may be direct, my knowledge of my bodily movements is not necessarily direct. What seems to be direct is my knowledge of what I intend to do, not my actual intentional actions.

If Bayne and/or Davidson are correct, this would problematise Anscombe's unification of the three uses under intentional action. If what I know is what I intend to do (intention-for-the-

future) and what I do not know for sure is what I am actually doing (intentional action), then it seems that intention-for-the-future has priority, as this is what is known to me. It also means that the attempt to define intention-for-the-future as a case of intentional action where the action is in a very nascent stage comes under attack, as we now have grounds for distinguishing between the two uses, with one being associated with *knowledge* and the other with *belief*. This means that any proponent of the unity of the three uses of intention will have to provide some alternative theory for such unity if their account of intention is to be taken as convincing. This is of course exactly what Davidson tries to do, and the point of his example is to build up an argument for uniting the three uses of intention under intention-for-the-future (see Chapter 3, Section 2).

### *2.3 A refinement of Anscombe's approach: Intention-as-belief*

Setiya, on the other hand, tries to refine Anscombe's argument in a different way by offering up a version of what I will call intention-as-belief. In his essay, "Knowledge of Intention," Setiya presents us with his formulation of Anscombe's Principle, which I have already mentioned earlier. To remind the reader:

*Anscombe's Principle:* If A has the capacity to act for reasons, she has the capacity to know what she is doing without observation or inference – in that her knowledge does not rest on sufficient prior evidence. (Setiya, 2010: 174)

Setiya then proceeds to "argue directly that the principle is true." He does not, however, "aim to give reasons that Anscombe would herself endorse. The plan is rather to rely on grounds acceptable to those who deny Anscombe's Principle and thus to give a novel argument on its behalf" (ibid.).

Before I examine these grounds, it is worth noting that the principle as laid out by Setiya differs in one crucial respect from the arguments Anscombe actually made. By framing the matter in terms of the "capacity" of an individual to know without observation when acting for reasons (acting intentionally), Setiya has done away with the necessary relationship between intention and knowing that Anscombe espoused. On Anscombe's view, if I knowingly acted for a reason then I acted intentionally and vice versa (Anscombe, 1963: 11). In terms of Setiya's wording, I could act for a reason and have the *capacity* to have non-

observational knowledge of this reason, without *necessarily* possessing such knowledge. Although this move enables Setiya to evade some of the criticisms levelled against Anscombe's position, it is important to keep in mind that at this early stage his argument has already shifted away from the original.

Setiya provides indirect justification for this move when he explains that the critical difference between his own approach and that of Anscombe is that his approach accounts for cases of partial knowledge, and more importantly, of belief. He concludes that the epistemic requirements for intentional action are not necessarily knowledge, but rather "justification of confidence, which comes by degrees" (Setiya, 2010: 174). An action can therefore be termed intentional if the agent had reasons for action and had the capacity to know what she was doing, even if this capacity was only realised in the form of a belief regarding what she was doing. This is, of course, a response to objections of type (ii), and it is obvious to see how the movement from knowledge to belief serves to mitigate that objection. To recall the shoe-tying example: I no longer require *knowledge* of what I am doing for the action to count as intentional, but rather a *belief* about what I am doing. Moreover, while I must hold the belief in order for the action to be intentional, I need not be correct in my belief. I might hold a mistaken belief about what I am doing, without that counting against my action being intentional.<sup>24</sup>

Setiya's argument is a superb exemplar of the intention-as-belief approach to the problem of intention. This approach is of particular relevance to the argument I develop in Chapter 5. Since a significant part of that chapter will deal with the criticisms levelled against the intention-as-belief approach and possible ways of overcoming them, I will not rehearse these criticisms here.

#### 2.4 Acting under a description

I now want to conclude this section with a brief remark on Anscombe's conception of "acting under a description" (Anscombe, 1963: 37). What Anscombe means by this conception is that a particular action performed by a particular agent is only ever intentional under a certain

---

<sup>24</sup> As for the objections of type (i), Setiya argues that our epistemic relationship to our intentional actions – i.e. either knowledge or belief – is non-observational. He also hopes to maintain the unity of the three uses based on belief rather than on intentional action (Setiya, 2011a). As I indicated earlier on, I will not pursue the issue of type (i) objections further here. I will, however, return to the role of belief in uniting the three uses in Chapter 5.

description of that action. For example: say I am chopping wood. I believe that the wood I am chopping is oak, when in fact it is cedar. In this case my actions can be legitimately called intentional under the description “I am chopping wood,” but not under the description “I am chopping cedar wood.” This follows from the requirement that I must know what I am doing in order for it to count as intentional. It is also not correct to take my actions as intentional under the description “I am chopping oak wood,” since this is not in fact what is taking place. The result of this point is that in any given situation the question regarding whether or not actions are intentional depends on the description under which the ascription is made. My suggestion, very simply stated, is that the notion of “acting under a description” is useful purely as a heuristic device in the ascription of intention to others, but nothing more. That is to say, it tells us nothing meaningful about the *definition* of intention, but only about our *ascription* of mental states and reasons to ourselves and others.

Anscombe argues that intention is present in a given act if, and only if, it is the case that the given agent knowingly undertook that action for a reason (Anscombe, 1963: 9-11). Considering what has already been said regarding her account of intention, this statement can be reformulated as: intention is present in a given act if, and only if, it is the case that the given agent acted intentionally. This is because “knowingly undertook that action for a reason” is synonymous with “acted intentionally” in Anscombe’s account. This reformulation shows that the statement is at least partly tautological. This is not an indictment of Anscombe’s point, as I see it, but rather an indication that the *definition* of intention is a separate matter from *ascribing* intention, and such ascription is correct when it is applied to cases that meet the definition of intention. Since the notion of “acting under a description” doesn’t add to our understanding of the epistemic grounds for *having* an intention, I will not devote further attention to it here.

In conclusion: Anscombe’s stringent epistemic requirement that intentional action must be known without observation has been shown to call up a number of insurmountable objections. Indeed, I have argued that in order to keep the epistemic requirement of non-observational knowledge it is necessary to forsake a unity of the three uses of intention based on intentional action. If such knowledge is possible of an intention, it can only be possible of intention-for-the-future. However, I do think that there is value in the arguments of thinkers attempting to retain useful elements of Anscombe’s account, such as that of Setiya. In

particular, I am convinced that belief plays a crucial and fundamental role in any convincing account of intention, and will argue this point in Chapter 5.

### **3. The relationship between motivations, causes, reasons and intention**

In this section I discuss Anscombe's treatment of the relationships between intention and reasons, intention and causes, and intention and motivation. As we will see, Anscombe argues that intentions are intimately related to reasons, and that reasons should not be understood as causes. This position allows her to avoid the pitfalls associated with the view that intention is a cause, but forces her into a position regarding intention, reasons and causes that places her distinctly in the minority. As for the question about the motivational status of intention, her answer to this follows directly from her solution to the unity of the three uses: collapsing intention-for-the-future and intention-with-which into intentional action.

#### *3.1 Distinguishing reasons and causes*

I have already mentioned in Section 1 that Anscombe considers intention to be essentially linked to reasons. In to her account, an action is only intentional if the agent would respond to the "Why?" question by providing a rational explanation (that is, an explanation by reasons) of the action. This idea that intentions are necessarily related to rational explanation is one of Anscombe's most important and enduring contributions to the discussion on intention. It is important to note that she is not asserting that an intention *is* a reason, but that intentions are necessarily related to reasons in some way. However, Anscombe is categorical in her rejection of identifying reasons with causes, where causes refer to effective empirical causes (Anscombe, 1963: 10, 24). This is a move that Bayne (2010: 10-11) links to her philosophical inheritance from Wittgenstein, as exemplified by her recalling "some notes on a lecture of Wittgenstein in which he imagined some leaves blown by the wind and saying 'Now I'll go this way...now I'll go that way' as the wind blew them" (Anscombe, 1963: 10). Her point, as Bayne sees it,<sup>25</sup> is to show that the causal power sometimes attributed to reasons (recall that for Anscombe to do a thing for a reason is fundamental to having an intention to do that thing, which is assumed to be the case with the talking leaves) is an illusion resulting from the confluence of a reason (as manifested as an intention) and a cause (the wind). The leaf has

---

<sup>25</sup> Although Bayne himself considers her to have missed the point of the example. See Bayne (2010: 10).

reasons, insofar as its statements are taken to reflect intentions, and it believes that these are the causes of its movement, when in fact this is not the case.

The basis of this distinction between reasons and causes begins with her consideration of the difference between *expressions of intention* and *predictions*. As pointed out by Driver (2011), there is some similarity between expressions of intention and predictions, insofar as both of them “are future-directed. Both seem to require a *belief* that a future state of affairs will occur.” The difference Anscombe identifies is that, whereas we justify predictions through the provision of *causal evidence*, we justify expressions of intentions through the provision of *reasons*, a fact that appears to say a great deal about what makes intention, intention (Anscombe, 1963: 1-5). Hence, for Anscombe, reasons and causes are distinct and not interchangeable, at least when employed as means of justification. Still, the link between intention and prediction is a deep one. Indeed it is arguable – and Bayne (2010: 5) does so argue – that intention requires prediction on a conceptual level. If the formation of intention requires a belief regarding the future (as Anscombe (1963: 5) certainly assumes it does), and prediction is the means through which we form beliefs about the future – or a belief we have regarding the future – then prediction is necessary for intention to have any content. I can scarcely form an intention if I am incapable of forming a belief concerning the future, since, as we have already said, intention is future-directed. Bayne (2010: 5) states this point more strongly when he says “[b]ut one thing seems to be certain: without, at least, a belief that we can make successful predictions, no intention will be formed.” This does not mean that intentions are derived in an unmediated fashion from our predictions (two people having the same predictions could very well form two disparate intentions), but that the ability to form intentions is at least partly founded on our belief in our ability to predict the future to a sufficient extent that our actions could change it in a meaningful way.

To return to the distinction between reasons and causes, it should be noted that Anscombe does introduce into her work the notion of a “mental cause” (as has been mentioned in Chapter 1, Section 3) but she is very clear that these should be distinguished from both the terms “intention” and “motive.” In her own words, part of her motivation in having “isolated this notion of a mental cause” was “because I want to distinguish it from the ordinary sense of ‘motive’ and ‘intention’” (Anscombe, 1963: 17-18). For Anscombe, a mental cause refers to a cause that is non-observable, whereas other types of causes (what I will simply call

“causes” as opposed to “mental causes”) can be observed (Anscombe, 1963: 16; Bayne, 2010: 50-51).

To explain this point Bayne (2010: 50) makes use of the following comparison: Case 1: I provide my reasons for the belief, “There will be an eclipse.” Case 2: I knock over a glass because I am startled by the appearance of a scary face at the window and then I give reasons for the action of knocking over the glass. The case of the eclipse is certainly not a case of intention, according to Anscombe. This is a case of providing reasons to justify a belief in a prediction or fact, which, for her, is distinct from providing reasons to justify an action – a distinction that I discuss in greater detail in the next section. Case 2 is more ambiguous, as it involves an action, and there is certainly a sense of the “Why?” question that is relevant here. Bayne (2010: 50) asserts that for Anscombe, this “Why?” should be answered by causal explanation (the identification of causes) and not by reasons, and it is therefore not an example of intention either. She views it as self-evident that justifications given for belief in a prediction or fact (as in Case 1) are not examples of intention as they are not related to action<sup>26</sup> and the “reasons” to be given are causal rather than rational explanations.<sup>27</sup> While Case 2 seems to have potential as a case of reasons for actions, what is actually at stake here is not intention, but a mental cause, which is a species of causal explanation, not rational explanation (Anscombe, 1963: 17-18).

Case 1, then, is an example of a prediction if it is directed toward the future, or of causal observation (observation of a fact in the world) if taking place in the present. Given what we have already discussed regarding prediction, this means that the justifications to be provided must be causal explanations. Case 2 is an example of involuntary action, precisely because the action’s explanation depends upon causes rather than reasons; even though it involves an action, it is not a case of intention, even though it might seem that way. The reason for the possible confusion is the existence of mental causes, and the potential mistake of conflating such causes and intentions. The reason we can easily identify Case 1 as not being an example of an intention is because it is obvious that causal explanation is the correct answer to the

---

<sup>26</sup> Or rather, the belief must be related to an action for the reasons provided to be related to intention.

<sup>27</sup> It can be easy to become confused with the employment of the terms “reason” and “cause” in Anscombe. In the simplest terms she declares that when a “Why?” question is pertinent, then an individual must respond with reasons for their intentional actions, beliefs, predictions or involuntary actions. Different senses of this “Why?” question are pertinent to each of the four listed cases: while the first requires the provision of reasons as rational explanations (which she often simply calls “reasons”), the third and fourth require causal explanations (or simply “causes”). Answering the “Why?” question with regards to belief means giving reasons to accept it as true.

“Why?” question. In Case 2, the causal chain is partly un-observable: the movements and actions of the person involved do not have an observable relationship to the foregoing cause (the face in the window).

It is this un-observable link in the causal chain that Anscombe refers to as a mental cause (Anscombe, 1963: 16). So what does she mean here by observable? She clearly does not mean that we can see causality, in the sense of seeing a thing in the world (though she opposes Hume’s conception of causality, she does share his conviction that causality cannot be considered a “thing in the world”). Rather, she seems to mean that we can observe the change that takes place and identify the participants: I can see the moon and I can see the shadow that progressively moves in front of the moon until the eclipse is complete. In Bayne’s words, “[w]hen the eclipse takes place I can see what is causing it; I observe the moon moving between the earth and the sun. The causing that is going on is visible to the naked eye, and I can see *that* the intervening moon is doing the causing” (Bayne, 2010: 50). By contrast, in the case of the face at the window, although the face caused me to jump, I did not observe that the face was causing me to jump. There is an un-observed gap between the appearance of the face and my jumping. Now, it is quite apparent that this gap is filled by the response of my mind to the face, which functions in a sense like a neuronal connection between the appearance of the face and my bodily response to it. In some cases where intention is present a mental cause may also be present, for example, if I hear a doorbell ring (a case of a mental cause) and this leads me to intentionally open the door. But an intention is not a species of mental cause or vice versa. Part of Anscombe’s goal is to clarify that cases such as Case 2 are not instances of intention, since the correct answer to the relevant “Why?” question is a causal explanation. The only reason for confusion in such cases is that mental causes and intentions are commonly (and mistakenly) conflated.

But what, then, is the status of intentions? Is the link between my intention to type the next word on this page and the bodily motions that bring it about only “a favour granted by fate” as Wittgenstein (1961: 6.374) claimed? Anscombe rejects this view of intentions as wholly separate from what happens in the world. As I have mentioned before, she also strongly disagrees with the Humean model of causality, where event-event causality is only a construction of habit (Anscombe, 1963: 16). Her reasons for rejecting Hume’s account and an exposition of her own account of general causality are not at issue here. What is relevant, however, is that she does not see intention as any sort of cause at all – a position that renders

her as much of an outsider to the mainstream accounts of intention as she is on the matter of the non-existence of mental entities.

### *3.2 The position that intention is not a motivation*

Anscombe's arguments for why an intentional (rational) explanation of action can never be identified with a causal explanation are inextricably tied to her rejection of intention as being understood as a motive. This approach, which Wilson and Shpall (2012) call "neo-Wittgensteinian," has lost most of its influence in the discussion around intention as a result of the work of Davidson and other so-called "causalists." For causalists, intentions are essentially related to desires, pro-attitudes or means-end beliefs (Davidson 2001; Bratman 1984, 1987, 2009a, 2009b), and these should be understood as causes of the behaviour and actions that stem from them. This is precisely the view that Anscombe rejects. For her desires, pro-attitudes and means-end beliefs are motives, and intentions are not motives, since: (i) it is possible to have an intention without a dominant desire and (ii) intentions do not form prior to action, but are in a sense retrospective, whereas motives must be anterior to action.

The first claim requires an example. Before I turn to it, it is important to keep in mind the following: Anscombe is not claiming that motivations cannot be present when an intention is present, but that it is possible to have the one without the other, and that this constitutes proof that the two concepts are not inextricable. For Anscombe (1957: 325) intention is closely related to the idea of choice. This gives us perhaps the best way of describing her position: it is possible that the action I chose (intended) was not the action I most wanted/desired/had the strongest motivation for. In Anscombe's view, the fact that this statement can be (and indeed is) employed in meaningful conversation indicates that it is possible to have an intention *in the face of one's motivations*.

A good example that I can think of that would support her position is that of kicking a stone down a road. I can certainly do this intentionally, but it does not seem necessary to have a motivation to do so, where motivation is an expression of a want or desire that is ulterior to the actual kicking. If asked the relevant "Why?" question regarding my action, my response is not a causal explanation, but the provision of the reason: "Because I chose to." It does not seem that more is necessarily demanded. If this is the case, then it would strengthen

Anscombe's point. In a similar vein, Bayne (2010: 56) gives the example of catching a ball. This action, too, can be undertaken intentionally without a motive, or at least so Bayne thinks. It is immediately obvious that any example of a non-motive originating intentional action, if it exists at all, will be of the kind described above: impulsive and brief. Whether this is of any importance is not clear to me, but it does seem to indicate that if such actions exist they would all be examples of the intentional action usage of intention, as it appears to be impossible to divorce either of the other two categories – intention-for-the-future or intention-with-which – from motives of some sort. This association with intentional action is perhaps unsurprising, considering Anscombe's approach to the unity of the three uses. It might be the case that intentional action is intention in its unadorned state, whereas the other two uses of intention are expressions of intention that include motives in their structure. And, as Anscombe claims that all cases of intention-for-the-future and intention-with-which can be described in terms of intentional action, motives are extraneous to the meaning of intention. That being said, it is also the case that in many instances the acceptable answer to the "Why?" question will entail a reason that cites motives. The point is that this is not always the case, whereas all acceptable answers require the reason component.

Needless to say, Anscombe's position has come in for vigorous opposition, most crucially by Davidson. The seemingly obvious problem is the question: "If your choice was not directed by motivations, what directed it?" – a question to which Anscombe does not provide a direct answer. I return to this issue in Section 4, as it entails an understanding of her approach to the role of practical reasoning in intention and vice versa.

Anscombe's second argument for dissociating intention from motive flows directly from her understanding of the unity of the three uses. Since intention and action are necessarily linked, there can be no point prior to action where one had an intention; the advent of intention is the advent of action. Motives, by contrast, must necessarily come before the action that they motivate. I cannot declare that I was only motivated to do something when I do it; the "to do" implies some form of temporal ordering, or causal association. This argument relies almost entirely on Anscombe's argument for the unity of the three uses under intentional action, since it identifies intention with the commencement of action. As I have already demonstrated, her argument for unity suffers from a number of fatal flaws, not least of which is the existence of cases of pure intending. Hence, just as the limb perishes with the body, so

too the temporal distinction between motives and intentions must fail with the argument on which it is founded.

To sum up the conclusions of this section: Anscombe argues that intentions are necessarily related to reasons, but are not necessarily reasons themselves. She argues against the view that intentions are to be understood as causes, a position she defends by arguing for a strict distinction between reasons and causes. This argument builds into her position that intention should not be identified with motivation, as she takes motives to be causes. I have argued that this position is predicated on her arguments for the unity of the three uses of intention through intentional action. As I have found this underlying argument to be fatally flawed, I conclude that Anscombe has not provided convincing arguments for why intention should not be considered a motivation, and by extension, a cause.

#### **4. The relationship between intention and practical reasoning**

In line with the fourth requirement for a convincing account of intention that I outlined in Chapter 1, this section will investigate how Anscombe answers the three questions about intention's relationship to practical reasoning. These are: (i) whether intention is a component of, a result of, separate from or contingently related to practical reasoning, (ii) how we are to understand the relationship between intentions, and (iii) how she resolves the charge of cognitivism. On the first question, I show that Anscombe's response is that intention is not necessarily related to practical reasoning. I will be criticising this position. Regarding the second question, I show that Anscombe is unable to explain the relationship between intentions, due mostly to a lack of engagement with the question of such relationships. On the third question I argue that Anscombe's account is not susceptible to charges of cognitivism.

##### *4.1 Intention's role in practical reasoning*

Recall that Anscombe considers motives and intentions to be separate and distinct. The process of practical reasoning, as Anscombe presents it, explains the movement from a want (which presumably includes the notion of motive) to a choice (which we know she closely associated with intention). As far as she is concerned, all the results of practical reasoning are intentions, although not all intentions are the result of practical reasoning (otherwise all intentions would be linked to motives). Examples of intentions not following from practical

reasoning would be cases such as those I have described in Section 3, namely kicking the stone down the road and catching the ball, all of which illustrate Anscombe's claim that intentional action can occur without motives. If we accept her argument in this regard, we would arrive at the conclusion that choices can occur without necessarily involving any practical reasoning. This is one of the aspects of her account of intention with which I strenuously disagree.

However, it is very difficult to discuss her take on the relation between intention and practical reasoning, as she was never explicit about the link between intention and motives. She seems clear that intentions can exist without motives and vice versa, but there is little more to go on than that, other than that motives must precede intentions, *if* and when the two are connected, which she takes to be a contingent matter. It is also indubitably so that the action that is the conclusion to a practical reasoning process must always be an intentional action. The chief difficulty is whether it is ever really the case that we make choices or hold intentions *without* motives (wants). If we cannot – and given that, for Anscombe, intention is inextricable from reasons – it would seemingly require that all intention must be linked to a motive *and*, by extension – practical reasoning. Anscombe's defence against this claim seems to lie purely in the realm of examples, and these examples, I argue, are only effective bulwarks in so far as they are incorrectly understood.

When I kick a stone down the street, and I am asked the relevant "Why?" question, then, for Anscombe, my action will count as intentional if I justify my action through the provision of a reason as opposed to the explaining my action through the provision of a causal explanation. In a case such as this, what kind of reasons could I provide that are not linked to motives? I can provide a reason such as "I chose to." However this seems very weak. After all, if somebody asked me, "Why did you choose to?" I would generally respond with, "Because I wanted to." And if "wanted to" is not a motive, then I am not sure what it is supposed to be. It seems to me that "Because I wanted to" expresses the recognition of having a motive, but one that I have not thought out or "modelled" through the process of practical reasoning. What is at stake here is not a difference between an intention related to a

motive and one that is not, but rather the difference between an intention related to a reasoned motive and one with an unreasoned motive.<sup>28</sup>

An unreasoned motive is one where my response to “Why?” is often something akin to, “I wanted to.” By contrast a reasoned motive has more content when faced with the same enquiry; I can add something to the explanation, and indeed an honest explanation would require me to do so. Now, if “I wanted to” is understood in this way – as referring to an unreasoned motive – then Anscombe cannot maintain the position that such an answer is the correct type of answer to the relevant “Why?” question, as it does not provide a reason. This would entail that, unless we are to think of an answer such as “I want to” as an acceptable example of a reason rather than the recognition of an unreasoned motive (which I find exceptionally doubtful), we must conclude, given Anscombe’s conception of intention, that such actions are not intentional. It is also at least possible that any reasoned motive can be reduced to an unreasoned motive (or a complex of unreasoned motives) through repeated application of the “Why?” question, which raises a conundrum for Anscombe’s account of intention.

Are we then to conclude that there is no meaningful distinction between intention as the result of practical reasoning and intention that does not result from such reasoning, other than that the former entails reasoned motives while the latter entails unreasoned ones? And would this mean that cases of the latter type, such as catching the ball, should be viewed as unintentional? I would answer no to these questions, precisely because, as I argue later on, *all* intention is the result of practical reasoning, even in such cases as the catching of the ball, given certain provisos. The missing ingredients in the above account of intention are the roles of belief and uncertainty.<sup>29</sup> These two concepts are intimately entwined – a claim for which I argue directly in Chapter 5.

---

<sup>28</sup> Reasoned and unreasoned motives should not be understood as equivalent to rational and irrational motives. An unreasoned motive is not necessarily irrational; it is simply a motive upon which I act without having reasoned about it. When I catch that ball thrown at my head, I have a motive in doing so, but I have not “thought it through,” so to speak. If I sit down and engage my reasoning, it is probable that I will be able to arrive at a reason for why I acted thus, and if I cannot – if the motive is “unreasonable” in the sense that I am unable to reason it out – this would seem to problematize the assertion that this is in fact a case of intention. Certainly Anscombe would have to accept this implication.

<sup>29</sup> As I will show in Chapters 3 and 4, both Davidson and Bratman are in agreement that intention is the result of practical reasoning, though their arguments as to why differ in some aspects both from each other and from my own arguments.

## 4.2 *The problem of plural intentions*

At this point, I want to pick up the thread of the discussion in Section 1 on the problem of the relationship between multiple intentions, since this discussion forms part of question (ii). Consider the following case provided by Setiya, “I intend to drink wine or beer with dinner.” The trouble with this case was that it brought into question the direct link between action and intention, as the statement does not have an unambiguous answer to the question, “What action is being intended?” For Anscombe, of course, any intention must be connected to a discrete action. This results in cases like the one above to count as objections to her account. Although I will not be following Anscombe in arguing for the primacy of intentional action, I do think that she could have mustered a defence against cases such as these. She could have claimed that they are statements not of a single intention to an ambiguous action, but rather that they are statements of *plural* intentions (a position Bratman (1984) adopts). Anscombe, however, seems to assume that there is never more than one intention at play at any moment. Whether this was actually her position, or simply an accidental result of her style of discussion on the matter, the outcome is to leave her account light on explaining about how intentions interact.

It seems common-sensical that I might have a plurality of intentions at any one time. Anscombe’s own definition of intention acknowledges this possibility: I am almost continually in the process of acting with reasons, and these reasons, even if they refer to the same action, are not identical. While typing this thesis, I intend to receive a degree for its completion; I also want to convince my readers, to make a telling point about intention and to satisfy the expectations of those whose opinions matter to me. These various wants, though not unrelated, *cannot* be deemed identical; neither is each one merely a step in the chain towards a single, supreme want – at least not without serious mental abstraction and cognitive gymnastics. This leads us to the point where it is reasonable to say that we can (and frequently do) have multiple intentions at any given time. Not only am I motivated to convince my readers, etc., but I also *intend* to do so. I am able to claim this in a straightforward and unproblematic way because I know that I could not identify a supreme intention from the above list; each one has value separate from the others, and I can accomplish one without accomplishing another. The only absolute commonality they share is that they all require the action of me writing this thesis as a necessary, though in itself insufficient, condition for achievement.

This shift in thinking has major implications for Setiya's example; however, they are not ones that Anscombe's supporters can easily use in her defence as she herself never discussed the possibility of plural intentions. This is perhaps an unavoidable result of her commitment to intentional action as the primary use of intention, and if so, the criticism of her position would be justified. On the other hand, if it was merely an omission on her part then grounds for defence do exist. She could claim, for instance, that in this case the action in question can be described as "I plan to drink *something* with dinner," hence rendering an individual action identifiable. This strategy won't really work, however. After all, I am still set on beer or wine; water or soda will not suffice. To shift the description to "something" seems an inaccurate way of getting to the intention. The easiest escape would have been to say that an intention is the best (in terms of enabling satisfaction, and to the best of the agent's knowledge) concatenation all the possible intentions at a point in time. In other words, if I have a complex intention to drink beer or wine, the resolution of this choice need not be made until the temporal horizon of choice is reached, when the complex must resolve itself. Until then, my practical reasoning leads me to the intention and action that best allows the potential fulfilment of my various intentions.

This response is, however, beyond Anscombe's reach. Not only because of her lack of discussion on the matter, but also for theoretical reasons. This approach would require that she change her thinking about the relationship between motivations and intention, as well as renounce the absolute connection of intention to action in the form she espouses it. My conclusion is that Anscombe cannot escape the problem of our inability to identify the discrete action at play in the example, and that this is yet another blow to her account of the unity of the three uses. My further view is that, finding a solution to the explanation of intention in cases similar to Setiya's example would enable us to devise a better account of intention, particularly with reference to the relationship between practical reasoning and intention. I attempt to develop such a solution in my own account of intention in Chapter 5.

#### *4.3 The charge of cognitivism*

Within the recent literature on intention, arguments relating to practical reasoning have tended to centre on the issue of "cognitivism." Recall that cognitivism is the label that Bratman uses for those thinkers (Velleman and Setiya chief amongst them) who collapse practical reasoning into theoretical reasoning – which Bratman considers a major mistake.

Anscombe was vehemently opposed to such a collapse, she maintains that there is “a difference of form between reasoning leading to action and reasoning for the truth of a conclusion” (Anscombe, 1963: 60). Intention is justified by the former, not the latter: by practical not theoretical reasoning (ibid.: 57–62). In line with this, her account is definitively not cognitivist. However, the issue of cognitivism remains relevant for her theory of intention in so far as it crops up in the work of those of her contemporary adherents who have adopted an intention-as-belief account of intention. I discuss the intention-as-belief account in Chapter 4, Section 4.3. My goal in this section is to show that Anscombe’s account is not a cognitivist account. Before we continue on, however, it is first necessary to grasp certain points about practical versus theoretical reasoning.

This distinction between these two forms of reasoning has a very long history, and Anscombe herself holds a fairly orthodox position on the matter. Her thinking on the topic was heavily influenced by Aristotle, although she certainly didn’t agree with all of his views. The contemporary approach to the distinction between practical and theoretical reason stems from Kant. That being said, although Kant did distinguish between the practical, theoretical and speculative reason, he did not intend that these be considered to be two (or three) wholly exclusive conceptions. In fact, he quite clearly states in the *Critique of Practical Reason* that the difference between theoretical and practical reasoning is the object of reason, not the type of reason itself. Hence his preference for the title *Critique of Pure Practical Reason* to better represent that he did not intend practical reason to be read as an alternative to pure (theoretical) reason. He informs us that he decided against it because the title would then be tautological, as one of the key purposes in his writing the *Critique of Practical Reason* was to show that practical reason was an extension of pure reason (Kant, 1788; 1996: 13-14).

Anscombe’s own view, though not derived from Kant, was remarkably similar in that she concluded that the difference between practical and theoretical reason was the purpose they served, not the nature of the reasoning itself (Anscombe, 1974: 132). If we recall, the two primary points of divergence between Setiya’s intention-as-belief account of intention and that of Anscombe are the following: that the unity of the three uses is not found in intentional action but in intention-for-the-future, and that belief regarding our actions must supplant knowledge as an indicator of intention. These two points of difference are also the major reasons why Anscombe is not accused of cognitivism, while Setiya is.

For Anscombe intention is simultaneous with action, and the knowledge we gain of our intentional actions is of a type that I have it as I do it and not before. By rejecting both these positions in Anscombe those who argue for intention-as-belief introduce cognitivism into their accounts (see Velleman (1989) and Setiya (2010)). If intention is prior to action and this intention is a certain belief regarding future action, which then generates that action, then there seems to be no room for practical reasoning. Anscombe's account, however, does not share these elements with the intention-as-belief account, and so is not cognitivist in nature and avoids the criticisms typical of a cognitivist account. However, these very same elements – intention being simultaneous with action and knowledge of intention being present only when the action is performed – also result in fatal flaws for Anscombe's account, as I have argued previously. For this reason, Anscombe's position on this topic is not convincing.

## **5. The relationship between intention and (moral) responsibility**

My aim in this section is to explicate Anscombe's understanding of the relationship between intention and moral responsibility. Keeping in mind the two questions about responsibility I stipulated as being necessary for a convincing account of intention – Am I morally responsible for all the consequences of intentional action? and Am I morally responsible for anything I do unintentionally? – Anscombe answers *no* to the first question, and *yes* to the second.<sup>30</sup> To understand this position and its implications requires us to examine Anscombe's view of what makes certain consequences intentional and others not – and how this relates to the moral relevance of the consequence.

### *5.1 Three types of consequences of intentional action: intended, foreseen and unforeseen*

In so far as all intentional actions have consequences, we have to decide which of these consequences should be considered morally relevant. For Anscombe, the answer is that there are at least three types of consequences (and these are categories that will reoccur throughout the discussions in this thesis): intended, foreseen and unforeseen (Gormally, 2011: 4). She does not distinguish between intended and intentional consequences, using the two terms interchangeably (Anscombe, 1982). Unforeseen consequences are never intended consequences, while her use of the term foreseen consequences applies to those consequences

---

<sup>30</sup> Note that the “anything” in the second question is purposefully not “everything.” What I mean in full is: “There is at least one thing I do unintentionally for which I am morally responsible.”

which are foreseen, but not intended.<sup>31</sup> Further, she claims that we are morally responsible for all intentional (intended) consequences, and some unintentional (that is, foreseen or unforeseen) consequences of our intentional actions.

The everyday norm is for all intended consequences to be considered morally relevant. I am unable to identify a case where an individual's intended actions and its consequences were considered morally irrelevant. By this is not meant that any intentional action is necessarily moral or immoral. Many actions are amoral, in the sense that they are neither morally laudable nor immoral. What is meant is that if an agent performs an intentional action then the agent would be morally responsible for it *if the action proved to be moral or immoral*. On the other hand, if the action was unintentional, then the general view is that the agent does not bear this moral responsibility, or bears it in a different way, if her action is proved to be moral or immoral. More colloquially, if I say that I did a thing intentionally, that seems in of itself to be an admission that I am morally culpable for it. It is when I claim that the action was unintentional, accidental or involuntary that doubt arises. A common view is that I cannot be held morally responsible for an action that I did not perform intentionally, and vice versa.<sup>32</sup>

Anscombe is in line with the common-sense view, as she considers intentional consequences to be obviously worthy of moral consideration and unforeseen consequences to fall outside such consideration. The tricky area is that of foreseen consequences. The consensus here is harder to locate, and common-sense reactions are not always in agreement. If foreseen is not a sufficient condition for moral responsibility in of itself, then we must identify what are the criteria that determines why a given foreseen consequence should be considered morally relevant while another is not.

## 5.2 The “principle of side-effects”

Before we get to criteria for judging the moral relevance of consequences, a word on uncertainty. There is no ignoring the fact that the consequences of my actions are not wholly predictable, a fact that holds for *all* types of consequences identified by Anscombe. As we

---

<sup>31</sup> She does not mean by this that intentional (intended) consequences are not foreseen, merely that that her use of the term foreseen consequences does not include this group.

<sup>32</sup> There is, of course, an exception to this view in moral theory, namely utilitarianism (along with certain other forms of consequentialism such as welfare consequentialism, perfectionist consequentialism and eudaimonistic consequentialism). What is of particular interest to me in this thesis is that utilitarianism does not link moral responsibility to intention (or indeed voluntariness) in any necessary way; that is, for utilitarians, any such connection is utterly contingent.

have seen, she attempts to avert this uncertainty with regard to intentional action by setting a very stringent epistemic requirement, but this is a largely untenable position (or at the very least would have to explain too many exceptions). Given the unpredictability of consequences, there is a temptation to adopt Kant's approach, and simply limit the scope of moral responsibility to the pure intention itself. However, this goes against our everyday experiences of successful prediction and moderate certainty. For our discussion of Anscombe's approach, I assume that pragmatically useful degrees of certainty are indeed possible, and that these can be applied to the consequences of our actions.

The very terms "foreseen" and "unforeseen" seem to refer to different levels of certainty. But if the difference between these levels of certainty is obvious, what is less clear is the distinction between intentional and foreseen consequences. Anscombe discusses this matter directly in the "Medalist's Address: Action, Intention and 'Double Effect'" (1982), in which she presents a defence for a bare-boned version of the doctrine of double effect which she names the "principle of side-effects" (Anscombe, 1982: 60-61). She takes the intentional consequences of my actions to be those that I intended to bring about, in that they form part of my answer to the relevant "Why?" question; they are always the direct result of intentional human action tied to my reasons for acting. Foreseen, but unintended, consequences are those which the agent would not raise as part of the relevant reasons for acting. So, for example, if I slice a tough substance with a knife, one of the foreseen consequences is that the knife will blunt, but this is not my intention in performing the action. The intentional (intended) consequences of the action are for the hard substance to be cut into several slices.

At this juncture, it is already interesting to note that it isn't all that clear whether most of us, if asked the question, "Did you blunt the knife intentionally?", would reply in the negative. After all, if I fire a bullet to kill an individual threatening my life, but am able to foresee that the bullet will pass through the opponent and strike an innocent bystander, it seems questionable whether an observer will not call this consequence intentional.<sup>33</sup>

The more immediate question, however, is what could make a given foreseen consequence morally relevant. In dealing with this question, Bayne (2010: 217) asserts that what determines the moral value is not the applicability of the "Why?" question, but rather the

---

<sup>33</sup> The use of the notion "I foresee that X will..." is problematic, as it seems to imply a degree of certainty in excess of that which is commonly possible (or possible at all). So when I employ the notion I use it in what I take to be the actual meaning of its use in conversation, specifically: "I foresee that it is *likely* that X will..."

question, which he quotes from Anscombe, “What you do it for?” It is not the intention or the voluntariness of an action that determines its moral relevance; it is the motive that underlies it. However, I disagree with Bayne on this point. If Bayne is correct, this would imply that, for Anscombe, not all intentional actions are morally relevant, since, as we have seen that she thinks that at least some intentional action can be unmotivated. This conflicts directly with her position, in the “Medalist’s Address,” that there *is* a morally meaningful distinction between intended and at least some foreseen consequences. Nowhere does she talk about motives being an underlying criterion, and she certainly brooks no exceptions to the rule that intended consequences count as morally relevant.

A more accurate interpretation of her position, as I understand it, is that the intentionality of an action plays a role in the ascription of moral responsibility, but that this is not the final word on the matter. She does not want to characterise all intentional action as morally relevant and then anything foreseen but unintentional as morally irrelevant. What we could say, rather, is that although intentionality ensures that a given action is considered morally relevant, this is not the only measure of such relevance. Another measure of moral relevance, applicable to foreseen consequences, is that of “what you do it for?” So instead of a single measure of moral relevance, such as Bayne proposed, we have two. One is applicable to intentional consequences, the other to foreseen consequences.

Moral responsibility is applicable to intentional actions, and in particular to the intentional consequences of such actions, and *also* to any foreseen consequences that are motivated – i.e. justified – in an unacceptable way. This latter group can be better understood with reference to practical reasoning. If I choose to take an action that earns me a moderate sum of money (the intentional consequence) but sees a moderate number of children die (the foreseen consequence) then I am morally responsible, as it can be said that my valuations made during the practical reasoning process were corrupt. I should not evaluate these two events in such a way that I would accept the death of the children as a consequence of an action of any sort (or at least of this sort). However, if there is sufficient distance and doubt about the consequence, then Anscombe (1982: 64) says “there is indeed room for saying that you did not intend that result, even though you could foresee it.” Anscombe never spells out in any great detail what sufficient distance and doubt would entail; her purpose was only to assert that we are certainly more directly morally responsible for our intentional consequences as opposed to some of our foreseen ones. This is the reason that her answer to the first question

can be taken to be an unequivocal no, whereas her answer to the second, though a yes, is more unclear in its implications. Although my approach is more similar to Anscombe than to Davidson or Bratman on the matter of intention's relationship to moral responsibility, I oppose, in Chapter 5, Section 5, her view that there is a distinction between the intentional and foreseen consequences of an intentional action.

## **Conclusion**

My goal in this chapter has been to examine Anscombe's account of intention in the light of the five requirements for a convincing account of intention, which I stipulated in Chapter 1. For my purposes, the most significant conclusions that this examination has yielded are the following: Anscombe's attempt to unite the three uses of intention through a reduction to intentional action is not successful. Since I take Anscombe's attempt to be the most thorough such attempt yet produced, I conclude that trying to achieve the unity of the three uses through intentional action is not a viable approach. Anscombe argues for a very stringent epistemic requirement for intention, which I have shown to be unsustainable. A less stringent requirement, such as that based on belief, seems more plausible. Anscombe's arguments for the distinction between reasons and causes, as well as for a special relationship between intention and reasons in the form of the "Why?" question, have been shown to be some of the strongest elements in her account. Indeed, all the other accounts I discuss in this thesis – those of Davidson and Bratman, as well as my own – will adopt this position. That being said, I have also demonstrated that Anscombe's position that intention is neither a cause or a motivation leads to a number of insurmountable difficulties. I have further argued that Anscombe's view that it is possible to have an intention without practical reasoning is flawed, at least partly because this view depends on the mistaken assumption that intentions cannot be motives. Finally, I have discussed Anscombe's distinctions between intentional, foreseen and unforeseen consequences, and the moral relevance of each. Although I oppose her conclusions in this regard, I consider it important for any account of intention to be able to explain the relationship between consequences and moral responsibility; consequently, it is a question I will seek to resolve in my own account of intention in Chapter 5.

On a general point, it should be clear from the discussion in this chapter that Anscombe can be credited for having introduced many important concepts and conceptual distinctions into the debate about intention. These concepts and distinctions have played and continue to play

a significant role in the on-going deliberation about intention. It is thus with good reason that Davidson says of Anscombe's account that it is "the most important treatment of intention since Aristotle" (Gormally, 2011: 3).

## CHAPTER 3: DAVIDSON'S ACCOUNT OF INTENTION

### Introduction

Donald Davidson's work on intention is strongly influenced by Anscombe's views. Nevertheless, it is Davidson who has come to represent the orthodox position in the field, largely because his account resolves many of the intractable problems faced by Anscombe's theory of intention. In this thesis I use the terms "early Davidson" and "later Davidson" to refer to the shift in his thinking on intention, as he himself explains in *Intending* (1978). Since I think that later Davidson presents a distinct improvement over the early Davidson, and that all that was of value in early Davidson is still present in later Davidson, I will not spend too much time on early Davidson. That being said, it is still important to understand how Davidson sets out to resolve the problem of intention. I therefore devote the first section of this chapter to a brief analysis of the early material, particularly his views on the unity of the three uses of intention, reasons, epistemic requirements, and practical reasoning.

After this overview of early Davidson, I turn to an analysis of how later Davidson's account fares in meeting the requirements for a convincing account of intention that I had stipulated in Chapter 1. I begin by examining, in Section 2, the difficulty that leads Davidson to revise his position on the unity of the three uses. Here I show that, after having come to the conclusion that pure intending cannot be explained in terms of either intention-with-which (as he had thought earlier) or intentional action, Davidson then formulates a new position that makes prospective intention, or intention-for-the-future, paramount. This leads him to adopt a view of intention as being a mental entity, one that can exist whether or not any action towards its fulfilment actually takes place. His verdict is that an intention is an "all-out judgement" in favour of a certain action, though this action need not eventuate (Davidson, 2001: 99). While this shift does not bring about significant changes to his views on motivation or the relationship between intention and causation, I will show that it does have dramatic implications for his views on the epistemic requirements and practical reason.

In Section 3, I examine the implications of his account for the relationship between intention and practical reasoning. It will be shown that later Davidson considers intention as a rational judgement, in particular as an all-things-considered judgement. I argue that this position is fraught with difficulties that Davidson's account cannot overcome, namely: explaining the

rules applicable to plural intentions, the problem of bootstrapping in practical reason and the problem presented by the existence of irrational intentions (such as in some cases of *akrasia*).

Section 4 starts off with an analysis of Davidson's conception of the relationship between intention and motivation. Specifically, he holds that intention is a type of pro-attitude with additional requirements. I argue that this is one of the strongest points in his account of intention, although I disagree with him on the nature of the additional requirements. I then examine Davidson's claim that intention should be understood as a cause. I show that although this position has merit, it fails to resolve the problem of causal deviance. Finally, I show that Davidson definitely views intention as a reason. This follows from his commitment to intention as a rational judgement. However, in the face of some of the objections discussed in Section 3, he changes the kind of rational judgement he associates with intention from an all-things-considered judgement to an unconditional judgement.

Section 5 is devoted to Davidson's understanding of the epistemic requirements for intention. I argue that Davidson's commitment to intention as a rational judgment results in specific epistemic requirement for intention, namely that the rational process resulting in the intention be knowable by the agent. However, this requirement falters in the face of problems of irrational behaviour, such as in some cases of *akrasia* and pathological behaviour. Davidson's attempted solution (the shift to intention as an unconditional judgement), although a step in the right direction insofar as it makes allowance for the opacity of reasons, is shown to be inadequate to overcome the problems.

Lastly, in Section 6, I try to reconstruct Davidson's understanding of the relationship between intention and moral responsibility. The reason why this is a *reconstruction* is that Davidson did not write on this topic directly. My primary focus will be on the way in which his account is able to explain the relations between the different types of consequences of intentional actions (intentional, intended, foreseen and unforeseen) on the one hand, and moral responsibility on the other. I argue that, due to Davidson's commitment to intention as a rational judgement, his account is incapable of making sense of the relationship between intention and moral responsibility.

## 1. Davidson's early account of intention

In this section, I explain early Davidson's conception of intention as a primary reason, composed of a pro-attitude and an epistemic component. Further, I examine early Davidson's argument for the unity of the three uses of intention through reduction to intention-with-which. I then discuss his view that intention is the result of practical reasoning, which he mostly understands in a straightforward way. After this, I investigate the consequences of his theory of intention for understanding intention's relationship to reasons, motivations and causes respectively. Finally, I look into the epistemic requirement in early Davidson's definition of intention – which, I will show, is significantly less stringent than that required by Anscombe.

### *1.1 Intention as a primary reason*

Davidson's early position is built around two key assumptions: (i) Rationalisations of actions provide primary reasons for actions, explicitly or implicitly and (ii) such rationalisations are *causal explanations*. The term "rationalisation of action" refers to "the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did" (Davidson, 2001: 3). For Davidson, the rationalisation of an action is the intention with which it is performed. For a person to have such a rationalisation for their action it must contain two elements: "(a) having some sort of pro attitude toward actions of a certain kind, and (b) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind" (ibid.: 3-4). In other words, for something to be counted as an intention it must have a *motivational component* (a "pro-attitude," understood as any species of "wanting," ranging from desires, obligations, appetites, etc.) linked to a particular action, and it must have an *epistemic component* (Davidson lists knowing, believing and remembering, among others) which means that the agent must be aware that the action undertaken is of the particular sort indicated by the motivational component. For example: I have the intention to impress a certain woman. The motivational component of this intention is my *desire* to impress this woman. However, my impressing this woman while having this desire is not sufficient for intention to be present. I could have the desire to impress this woman while walking down the street singing aloud while she, without my knowledge, overhears me. She could be impressed by my singing, and therefore I would have impressed this woman. However, I would not have done so intentionally, as my intention did not have the required

epistemic component, namely knowing that singing aloud while walking home is in fact impressing the woman in question.

Together, the motivational component and the epistemic component give us the “primary reason” with which an individual acts (Davidson, 2001: 4). So, for an agent to have a primary reason means that agent has to have a motivation (pro-attitude) towards a certain action and must undertake that action under the awareness that it is that sort of action. Davidson’s second assumption, that such rationalisation is a species of causal explanation of action, is not particularly focussed on explaining what intention is, but it does describe an attribute that he proposes intention must have, namely *causal power*. If Davidson is correct about this, it would mean that to have an intention X to do Y (to bring about event Z) and then to do so, intention X is the *causal explanation* of the action Y (and/or event Z by extension), or at least part of the causal explanation. This, as Davidson (2001: 3) declares, is an “ancient – and commonsense – position.”

Rationalisation in the context it is used here means simply to be able to give a reason. So to rationalise an action X undertaken by an agent means to give the agent’s reasons for undertaking said action. Davidson’s argument dictates that such a rationalisation is necessary for any claim of intention to hold. No agent can claim to have undertaken an action intentionally unless that action can be rationalised. Thus, although rationalisation is not a sufficient condition for having an intention, it is a necessary one.<sup>34</sup>

### *1.2 Explaining the unity of the three uses of intention*

Early Davidson further argues that his account of intention understood as a primary reason solves the problem of the unity of the three uses by reducing all of them to intention-with-which. His reasoning in this regard can be explained as follows: A description of an intention, understood as a primary reason, is identical in meaning to the description of intention-with-which. For example: (i) “My primary reason for knowingly doing Y (and remember that I cannot have a primary reason if I do not know that I have it) is that I have a motivation to do X” is identical in meaning to (ii) “I am reading an article by Davidson because I intend to better structure my analysis of his arguments,” if intention is understood as primary reason. He then seeks to incorporate intentional action by arguing that “acting intentionally...was just

---

<sup>34</sup> This view comes close to Anscombe’s employment of the “Why?” question (Driver, 2011).

acting with some intention” (Davidson, 2001: xvii). This means that the use of intention as intentional action is merely a case of reporting “Y,” from (i) above. What makes “Y” intentional is precisely the fact that it can be accurately described as part of an intention-with-which, by which Davidson means that it is simply a case of acting with a primary reason – as is the case in (i). No action can be taken to be intentional unless it can be accurately described as part of an intention-with-which. Finally, he attempts to reduce intention-for-the-future to intention-with-which. However, this proves to be a far more challenging enterprise, and one that, by Davidson’s own admission, ends in failure. Early Davidson’s attempt to incorporate intention-for-the-future is thwarted, as is Anscombe’s, by his failure to recognise the existence of pure intentions – intentions that are only “in the head.”

With regards to practical reasoning, early Davidson views intention-with-which as the fundamental form of intention to which the other uses can be reduced. This has certain implications for his understanding of practical reasoning. However, he does not deviate from the general view that practical reasoning is a process by which a human moves from a motivation to an intention. Intention is, therefore, the *result* of practical reasoning. Considering what we already know about his account of intention, this should not be surprising: for Davidson, an intention is synonymous with a primary reason, which is in turn the combination of a pro-attitude and an epistemic awareness that the action to be performed will fulfil this pro-attitude. This means that, for Davidson, intention is necessarily related to practical reasoning. Further, due to the central position of intention-with-which, such reasoning can never be futile, so to speak. If I have an outcome to practical reasoning, an intention, then I must be on the way to action, even if only at a very incipient stage of progress. On this point, early Davidson agrees with Anscombe, and it should therefore come as no surprise that the very problems that plagued her account would eventually lead Davidson to revise his position.

### *1.3 Intention as a reason, motive and cause*

Turning to intention’s relationship to reasons, motivations and causes, Davidson’s commitment to rationalisation raises the following objection: Is it not possible to have an intention, or to act intentionally, without necessarily having a reason for that particular act? To return to an earlier example, I am walking home one day when I see a stone lying on the road. Without deliberation or having any determinate reason for it, I kick the stone and send it

skidding away. Now, few people would hesitate to call this an example of intentional action and, if asked, they would say that my intention was to kick the stone. However, would they say that I had a *reason* for kicking the stone? Remember that, according to Davidson, if an action cannot be rationalised then it cannot be called intentional, and we would therefore have to conclude that my actions in kicking the stone were unintentional. One might respond to this objection by pointing out that by “reason” Davidson means a co-occurrence of a pro-attitude for an action and the awareness that the action I perform is that kind of action. Thus, since I had a pro-attitude to kick the stone (a motivation to do so), and I was aware that kicking the stone was an action that would satisfy that pro-attitude, I did have a reason for the action. In simple terms: my reason for doing X was that I had a motivation to do so (Y).

This movement toward simplification is made possible, at least in part, by Davidson’s straightforward assertion that our intentions (composed as they are by both motivation and reasons) are in fact causes. Unlike Anscombe, who rejects understanding intentions as causes, Davidson sees no inherent difficulty in conflating intention as a rational explanation (a point on which he agrees with Anscombe) with intention as a causal explanation. In simple terms: our intentions are both rational and causal explanations for action. As mentioned before, Davidson takes this to be a perfectly intuitive and common-sense understanding of intention. In a similar vein, he asserts that intentions cannot be unmotivated. Indeed, the concept of primary reason necessarily includes what he calls a pro-attitude. When I ask somebody the relevant “Why?” question in Anscombe’s sense, it seems obvious that, in preparation of their answer, they will think about their motivation (pro-attitude), whether or not they were aware of having this motivation when they acted, and held the belief that acting thus would help fulfil the motivation.

#### *1.4 The epistemic requirement for intention and practical reasoning*

It is my contention that what Davidson takes to be the necessary epistemic requirements for intention are intertwined with the notion of practical reasoning. To explain: when I act on a pro-attitude with no epistemic awareness of this attitude, such action is not intentional. By contrast, when that pro-attitude has been the subject of a process of practical reasoning, then I am epistemically aware of the pro-attitude and of the plan of action to bring it about. The process of practical reasoning, and that of acquiring the necessary epistemic awareness, may not be identical, but they occur simultaneously and are inextricably linked. The one cannot

occur without the other. As this shows, early Davidson has far less stringent epistemic requirements for intention than does Anscombe. Let us also recall that it is that very stringent requirement that caused the most problems for Anscombe's account of intention, second only to the difficulties surrounding cases of "pure intending." The early Davidson's approach also has the benefit that it seems to fit many of our common-sense intuitions about the use of intention. In light of this we can therefore do away with any presumption about intention reflecting special or privileged knowledge.<sup>35</sup>

However, the difficulty of reconciling pure intending with a unity of the three uses of intention centered on intention-with-which eventually led Davidson to revise his theory of intention. I discuss this revision, which I take to be the turn from early Davidson to later Davidson, in the next section. I argue that this revision, in so far as it is an attempt to overcome the flaws of his earlier theory, is partially successful, though not sufficiently so for the later Davidson's account of intention to be considered wholly convincing.

## **2. The unity of the three seemingly irreconcilable uses of intention**

Any discussion of later Davidson's account of the unity of the three uses of intention must begin, as he does, by considering the problem of "pure intending." We have already encountered the difficulty presented by this class of intention in the discussion of Anscombe's account of intention. In its barest form, this difficulty can be expressed as a question: Is realised action necessary for intention? Anscombe, as we should recall, answers yes, as does early Davidson. The later Davidson answers no, after concluding that pure intending cannot be explained in terms of either intention-with-which or intentional action. Following this, Davidson then formulates a new position that gives pride of place to intention-for-the-future (or prospective intention). In his own words:

When I wrote Essay 1 [I] believed that of the three main uses of the concept of intention distinguished by Anscombe (acting with an intention, acting intentionally, and intending to act), the first was the most basic. Acting intentionally, I argued in Essay 1, was just acting with some intention. That left intending, which I somehow thought would be simple to understand in terms of the others. I was wrong. When I finally came to work on it, I found it the hardest of the three; contrary to my original view, it came to

---

<sup>35</sup> See Mackie's commentary on an article by Hintikka regarding a very similar point (Mackie, 1974: 103-112).

seem the basic notion on which the others depend; and what progress I made with it partially undermined an important theme in Essay 1—that ‘the intention with which the action was done’ does not refer to an entity or state of any kind. (Davidson, 2001: xvii)

This leads the later Davidson to adopt the view that intention is a mental entity, one that can exist whether or not any action actually takes place. His verdict is that an intention is an “all-out judgement” in favour of a certain action, though this action need not eventuate (Davidson, 2001: 99). This shift did not bring about significant change in questions of motivation and causation, but it did necessitate a number of changes to the epistemic requirements for intention and intention’s relationship to practical reasoning. I discuss these changes in the sections to come.

### *2.1 The turn to intention-for-the-future*

The later Davidson argues that intention can be said to exist even when the only use of intention is intention-for-the-future. An intentional action A is “the execution of an intention to do A, and doing A with the intention of doing B is intending, of one’s doing A, that it promotes or constitutes one’s doing B” (Setiya, 2011a), where A and B are both mental entities. The essential characteristics of intention are then those that are present for intention-for-the-future, and the other two uses are simply classes of intention with added features. In the case of intentional action, the element that makes the action intentional is the presence of the preceding intention to do A. The fact that the action indeed takes place in no way influences the intentional status of the action one way or the other. In the case of intention-with-which, the element that enables the attribution of intention is once again the presence of the intention A, which is brought closer to completion by the agent’s B’ing, but this latter fact does not have any significance as far as the presence of intention is concerned.

This revised position of Davidson’s immediately runs into the difficulty of addressing questions about actions that are intentional but not “pre-planned,” or rather, where no mental entity precedes the action. Davidson does not provide a solution to this difficulty in his writing. However, one way of trying to resolve the difficulty is to introduce a distinction between two instances of intention-for-the-future, namely “prior intention” and “intention in action” (Searle, 1981: 728). The introduction of this distinction is often attributed to Searle, expressed most fully in his book *Intentionality: An Essay in the Philosophy of Mind* (1983:

84-85). The advantage of this distinction is that it explains how intention-for-the-future can be described as present even in cases of spontaneous action. Under this understanding prior (or prospective) intention refers to instances where one intends to A at a point in the future. On the other hand, intention in action is present “without [the agent] forming, consciously or unconsciously, an [*sic*] prior intention to do those things” (Searle, 1980: 52). Searle asserts that though some intentional actions entail prior intentions, all intentional actions entail intention in actions. As Setiya (2011a) explains, “one intends to be doing it now when S is doing A intentionally, she is doing it in execution of an intention in action, though except in very unusual cases she also intends to do A: to complete the action she is in the midst of performing.”

However, Davidson’s account, with or without Searle’s refinement,<sup>36</sup> runs into the problem of causal deviance, which I have already mentioned in Chapter 2. Causal deviance is best explained by noting that there are possible situations in which I might perform my intention, but where I do so unintentionally. A good example of this is if I intend to shiver, and then shiver involuntarily due to a cool breeze. In this case, my actions fulfil my intention, but I did not act intentionally. However, Searle’s refinement would require me to refer to these cases as intentional, as they meet the requirements of intention in action. This is an outright flaw in the argument as “improved” by Searle. However, even without the refinement, Davidson’s account struggles with causal deviance. As we have seen, Anscombe sidesteps this problem by describing intentions as being non-causal. This works for her simply because the problem of causal deviance only arises if we allow that intentional actions are the *causal* results of an intention – which is precisely Davidson’s position. I unpack this problem in greater detail in Section 4 below.

Another, more general, problem for later Davidson’s theory of intention is accounting for the actual content of a mental entity: “intention.” For example, if I fire a flare with the intention of signalling for help, does the resulting avalanche qualify as part of the mental entity that is my intention? The issue at stake here tend to revolve around questions of foreseeability and the rather crucial notion of an unintended consequence. In other words, does my awareness of the likelihood of triggering an avalanche affect the verdict about its intentionality? Is there a threshold of likelihood? Obviously, these are important questions, not least because of their implications for notions of responsibility, to which I return in Section 5. For the time being, it

---

<sup>36</sup> I will be addressing Searle’s arguments and the problems that they face in detail in Chapter 5, Section 1.

is worth noting that there is near universal agreement that what is rationally unforeseeable cannot be intentional. The more important question for my present discussion is the concept of *unintended* consequences. The most troublesome aspect of this concept is that it does not seem involve questions of epistemology, but rather of priority, in some sense. If I push my friend out of the way of an oncoming train with the intention of saving her, at the expense of my own life, it is obvious that this is an *intended consequence*. It would be perfectly in line with normal use of language to refer to my being hit by the train as an *unintended consequence* of my saving my friend, just as causing the avalanche would usually be described as an unintended consequence of firing my flare in the case above. However, while in the case of the flare I may or may not be aware of the danger of the avalanche, it seems strange to think that I could be aware of jumping in front of the train to save my friend, without being aware that I would be hit by the train. Being hit by the train was certainly a foreseen consequence of my action: I knew that I would be hit by the train. I knew this as surely, or at the very least no less surely, than I knew I was going to save my friend. This means that in terms of epistemic certainty these two consequences are equal.

If the two consequences cannot be differentiated in terms of a quality of their epistemic content, then what separates them? What makes the one an intended consequence of my action and the other an unintended consequence? Recall that, according to the view of intention under discussion here, intention is a mental entity that precedes action, and that mental entity is an all-out judgement accounting for all variables that entered the practical reasoning process. This means that saving the friend and the being hit by a train cannot be clearly distinguished in terms of their intentionality, as both are elements of this all-out judgement. Let me expand on this. Remember that an all-out judgement is an *all-things-considered* judgement; it should literally be the result of my having reasoned through *all* the motivations, reasons and alternative actions and their potential consequences of which I am aware.<sup>37</sup> This means that any foreseen consequence of my action *has been considered*. I had a motivation to save my friend, but in the latter Davidsonian account this does not give that goal any particular status, as it had to be weighed against all my other goals, including possible routes to their fulfilment. This means that we must take every foreseen aspect of the action (including, obviously, consequences) as having been chosen by the agent. This creates

---

<sup>37</sup> Of course nobody is omniscient, so presumably there will always be some alternatives and consequences that an agent does not factor into her decision making. An all-things-considered judgement is never really an *all things considered* judgement in this fuller sense of *all*.

a problem for any attempt to distinguish between pushing the friend and being hit by the train. Both were necessary to achieve the intention; if they were not, I could have done otherwise. Once again, of course, this would only apply to those consequences I did, in fact, foresee.

At this point, two alternatives seem to present themselves. On the one hand, we could take it that some foreseen consequences are intended consequences, and that these are somehow different from those foreseen consequences that are taken as unintended. On the other hand, we can take the position that any and all consequences that are foreseen are intended and that the term “unintended consequences” has exclusive application to unforeseen consequences, where it is already applied in an unproblematic way. Davidson argues for the first position, taking the foreseen consequences that count as intentional to be those linked to the relevant pro-attitude. As this discussion is of particular concern for the relationship between intention and responsibility, I am leaving it here and taking it up again in Section 6 below and in the final section of Chapter 5.

## *2.2 Analysis of Knobe and Burra’s criticisms of Davidson’s account of intention*

At this point, it should be apparent that the biggest hurdle facing Davidson’s attempt to unify the three uses of intention is finding a way of uniting intentional action with intention-for-the-future. Anscombe, of course, had the same problem only in reverse; that is to say, she struggled to account for intention-for-the-future as part of intentional action. Indeed, Knobe and Burra (2006) argue that any attempt to unify the three uses of intention will fail as the concepts “intention” and “intentionally” are not, in fact, directly related terms. Examining Knobe and Burra’s arguments, as well as the counterarguments that Davidson’s account permits, will give us a clearer idea of whether or not a unity of intentional action and intention-for-the-future is possible under Davidson’s account of intention. Knobe and Burra’s arguments are also important for the thesis as a whole, in so far as they concern the possibility for *any* convincing account of intention that relies on the unity of the three uses.

Knobe and Burra’s argument can be broken down into three premises: (i) The relation “between intention and intentionally is ‘radically different from’ the relationships between compassion and compassionately, love and lovingly, or lust and lustfully,” so that they are in fact “two separate words that happen to be morphologically related” (Harman, 2006: 2); (ii)

“intention” is a non-moral concept in terms of its usage in common parlance, whereas the word “intentionally” is morally weighted and (iii) in other languages the terms that best translate “intention” and “intentionally” do not share any particular relationship, which indicates that the underlying meanings of these terms in English do not share such a relationship either.

The first premise (i) assumes that there is a relational rule, common to all the listed examples of related terms, that does not govern the intention-intentionally pair. This relational rule is that “the meaning of the adverb in each of these pairs is derived from the meaning of the corresponding noun” (Knobe and Burra, 2006: 124) and following this that:

[in] each case, the adverb is used to indicate that a behaviour was performed in the manner of one who has the mental state denoted by the noun. Thus, the word ‘lovingly’ means roughly ‘in the manner of one who feels love,’ and ‘lustfully’ means roughly ‘in the manner of one who feels lust.’ The fact that all of these adverbs seem to follow the same rule gives us some reason to believe that the meanings of the adverbs truly were derived from the meanings of the nouns (ibid.)

The second premise (ii) is that attributions of intention are not greatly altered by moral consideration, whereas attributions of intentionality are. That is to say, irrespective of whether the intention is moral, immoral or amoral there is no noticeable or significant change to its attribution to an individual. By contrast, when I attribute intentionality to an individual’s actions, moral considerations seem to play a much more substantial role. Knobe and Burra reach this conclusion largely due to the results of a practical experiment undertaken by Knobe, in which a sample group were asked to read two vignettes, each accompanied by an alternating set of two questions about what they have read. Although the actual experiment was run slightly differently<sup>38</sup> a simplified summary of the key elements can be described as follows: the first vignette, or the “Harm vignette” (Knobe and Burra, 2006: 130) goes as follows: There is a CEO who is brought a proposal that, if implemented, would increase profits for his business and simultaneously harm the environment. He exclaims, “I don’t care about the environment, let’s raise profits!” and ratifies the proposal. The respondent is then asked either: 1.1a. Did the CEO have the intention to harm the environment? and 1.1b. Can the CEO be blamed for the harm to the environment?, or she is

---

<sup>38</sup> To read the original see Knobe’s *Intention, intentional action and Moral Considerations*, 2004.

asked 1.2a. Did the CEO intentionally harm the environment? and 1.2b. Can the CEO be blamed for harming the environment? The second vignette or the “Help vignette” (ibid.) is very similar: There is a CEO who is brought a proposal that, if implemented, would increase profits for his business and simultaneously benefit (help) the environment. He exclaims, “I don’t care about the environment, let’s raise profits!” and ratifies the proposal. The questions change accordingly, either: 2.1a. Did the CEO have the intention to help the environment? and 2.1b. Can the CEO be praised for helping the environment? or 2.2a Did the CEO intentionally help the environment? and 2.2b. Can the CEO be praised for helping the environment?

Knobe tabled the responses to questions 1.1a, 1.2a, 2.1a and 2.2a as follows:

Here are the percentages of subjects responding ‘yes’ to each of these questions:

	Harm	Help
Intentionally	87%	20%
Intention	29%	0%

(Knobe, 2004: 185)

The answers to these questions represent, to Knobe and Burra, compelling evidence for two conclusions: (iia), that there is a gap between the attribution of intention and intentionality, and (iib), that moral considerations form a significant aspect of this difference.

The third and final premise (iii) – that the most fitting translations for intention in various other languages have little or no relation (even morphologically) to the translations for intentionally – although interesting, falls outside my competence, and I am therefore unable to judge the cogency of this claim. While it is certainly a relevant inquiry for specialists in linguistics, I will not engage any further with this aspect of Knobe and Burra’s argument. In any event, if they are right, then Anscombe – and hence Davidson – was mistaken in trying to unify intention and intentionality, and indirectly responsible for a great deal of wasted time and effort expended on this fruitless endeavour. Unfortunately for Knobe and Burra, I do not

think that their arguments are sufficient to overthrow the Anscombean orthodoxy on this particular point. Following Harman (2006) and Adams (2006), it seems to me that intention and intentionality cannot be so easily separated. Let us examine the three premises a bit more closely:

We have seen that Knobe and Burra argue in premise (i) that acting intentionally is nothing at all like “acting in a manner of one who has an intention,” whereas loving, lustfully and the rest can be expressed with exactly this form, e.g. “to act lovingly is to act in a manner of one who feels love.” But this seems terribly limiting, as Harman (2006: 3) points out. After all, I can wash the dishes lovingly while concealing that I am doing so, yet this does not alter the fact that I am washing the dishes lovingly. It seems that rather than “in a manner of one who feels X” it would make more sense to say, simply, “with X,” that is: “I lovingly washed the dishes because I washed them with love” rather than: “I lovingly washed the dishes because I washed them in the manner of a person who feels love.”

It strikes me that Knobe and Burra’s suggestion is focused largely on how external observers ascribe intention to others, rather than how an agent self-ascribes intention. What I mean by this is that, when it comes to my ascriptions of loving or lustful or intentional behaviour to others, I might well operate under the guideline of “acting in a manner of one who feels/has X.” However, surely there are cases – such as the dishwashing example – where an action is undertaken in such a way that I cannot correctly identify the intention based on my operating guidelines for ascribing intention to others. To be sure, in such cases I would have to *ask the agent in question*, as Anscombe would remind us. The final verdict as to whether the agent was acting lovingly or lustfully or intentionally can surely only be arrived at by interrogating the feelings and thoughts of the agent in question. What matters is *self-ascription*. So, in order to correctly ascribe loving, lustful or intentional status to an agent’s actions, what matters is not the guideline “acting in a manner of one who feels/has X,” but rather whether the agent would self-ascribe the status to herself. If I am right about this, then intention appears to be no different than the other cases provided.

This brings us to the problems with Knobe and Burra’s premise (ii), concerning the practical experimentation performed by Knobe and their subsequent claim about the role of morality in the ascription of intention and intentionality. First off, as far as (iia) is concerned, there is the rather obvious fact that, in both the “Harm” and “Help” examples, respondents attributed

intention to the CEO far less frequently than they attributed intentionality to his actions. Knobe and Burra then claim that if Davidson were correct about the relationship between the intention-for-the-future and intentional action, then this should be impossible, as to act intentionally literally means to act with an intention. So either these respondents are wrong in their attributions of intention or intentionality, or Davidson is wrong. In my view, it is the latter. Although I *do* think there is a way to salvage the unity between intention and intentionality from Knobe and Burra's criticisms, in my view Davidson's account cannot provide it. What we are seeing here is a result of the fact that Davidson is not particularly thorough in working out the content of the mental entity called intention. When Davidson tells us that an action X is performed intentionally if it is undertaken with an intention, he unfortunately insists that that intention must necessarily be "to X." He still links an intentional action to X with an intending (or an intention) to X, even when there is a supervening intention Y. This being the case, Davidson's account cannot make sense of problem (iia).<sup>39</sup> Knobe and Burra, as an alternative to Davidson, argue for premise (iib), that this discrepancy in attribution is due to the role played by the moral valuation of an action. I will not be discussing premise (iib) in this section, however, as it deals with issues concerning moral responsibility to which I turn in Section 5, and attempt to resolve in Chapter 5, Section 5.

However, it is not inconceivable that Davidson, if pushed, would concede that an intentional action X, though requiring an intention to be present, need not require that an intention to X must be present. In other words, I can perform an action X intentionally while *only* having an intention to do Y, provided that my all-things-considered judgement indicated that performing X was part of the fulfilment of Y. My intention is not the same as my intentional actions. To expand on this point, consider a case where I wish to strike a target with a bullet. I have the intention of striking the target with the bullet, and part of the all-things-considered judgement that this intention entails is that I must raise my gun and fire at the target. When I hoist my gun and fire at the target I am doing so intentionally, as I have an intention to shoot the target. Note, though, that my intention is to shoot the target, not to hoist the gun and fire, which is merely an action that I must undertake in order to achieve my intention. My intention as an all-things-considered judgement does account for this action in the sense that the agent is *aware* of it, but the action is not itself the intention; it is a means towards realising the

---

<sup>39</sup> As we will see in Chapter 4, one of the main strengths of Bratman's account of intention lies precisely in the way in which he tries to resolve this problem.

intention, a means that is intentional due to it being a necessary step on the way to realising the intention. More formally, one could say: “I perform action X intentionally when I perform it with an intention Y, where according to my all-things-considered judgement X is necessary for the achievement of Y.” This unites intentional action with intention-with-which, and suborns both to intention for the future. This would explain the discrepancy in the attribution of intention and intentionality that Knobe and Burra identified without having to make reference to the moral value of the action in question.<sup>40</sup>

At this stage, the reader might claim that I am being overly charitable to Davidson and insufficiently so to Knobe and Burra, and to some extent this would be true. Davidson never explains his view of intention in the manner I have above, and we must therefore remain unsure as to whether he thought that the intention I have when I undertake an intentional action must have the same object as the action itself. We might find a possible clue about his thinking in this regard in his explication of intention-with-which:

If someone performs an action of type A with the intention of performing an action of type B, then he must have a pro-attitude toward actions of type B (which may be expressed in the form: an action of type B is good (or has some other positive attribute) and a belief that in performing an action of type A he will be (or probably will be) performing an action of type B (the belief may be expressed in the obvious way) (Davidson, 2001: 86-87)

It would be peculiar to think Davidson meant for A and B to be the same!

Knobe and Burra (2004: 121) recognise this as a possible interpretation, but dismiss it thus:

[S]uppose that an assassin is trying to shoot the president but believes that it is extremely unlikely that he will succeed. Here it seems a bit odd to say that the assassin has formed an ‘intention’ to shoot the president, but if he actually did succeed, we would surely say that he had shot the president ‘intentionally.’

---

<sup>40</sup> An alternative argument in favour of not separating intention and intentionality as starkly as Knobe and Burra propose is put forward by Adams (2006), who contends that the pragmatic considerations of the respondents taints the results of the studies. This argument is, to my mind, compelling, but deals more with the question of how do we attribute intention and intentionality as opposed to what is and is not intended and intentional respectively. As such I have elected not to discuss Adams’ critique here.

The point being made here is that, even when the action in question is directly related to the pro-attitude, *intention* is still ascribed differently from *intentionality*. To my mind, this appeal to intuitions is a weak argument. From my own position, it would not at all seem “a bit odd” to attribute an intention to the assassin; in fact, I would propose that most of us in our own lives form intentions to do things that we believe to be extremely unlikely to succeed. A simple example: I want to call a girl to ask her out on a date. I believe that my attempt has a dismal chance of success, but I work up the nerve to undertake the necessary actions to achieve my goal. To borrow Knobe and Burra’s language, *surely* in this case I can be said to have had the intention to ask the girl out, which makes my phoning her an intentional action. For Knobe and Burra’s argument to hold, the key factor cannot be that the agent believes *success* to be unlikely, but that the agent must believe that *trying* is unlikely.

The question then seems to be whether there is a difference between cases of intentional action and intention-with-which. After all, if we are saying that any intentional action must have a preceding intention, does this not already imply that intentional action is subsumed under intention-with-which? The three answers that present themselves most readily are that (i) intentional action is a term attributable to a component or aspect of intention-with-which, or (ii) that the term only applies to those cases where our intention-for-the-future is of the type designated by Searle as intention in action, or (iii) that intentional action implies a conception of intention that is not intention-for-the-future. The second response is unacceptable, as Searle’s distinction is still plagued by the inability to account for causal deviance, while the third would require forsaking a crucial part of Davidson’s account. Although this latter possibility cannot be discounted out of hand, I think we would be making a mistake in pursuing this view.

As I see it, we would be better off concluding that intentional action is in fact a component of intention-with-which. That is to say, whenever an action can be incorporated into a statement of intention-with-which, then that individual action can be deemed intentional. Hence, since I am typing with the goal of completing this thesis, with the *intention* of completing it, I can explain my intention-with-which in various ways, and these various ways of explaining it can lead to a variety of ways of identifying intentional actions. For example: “I am typing this example (X) with the intention of completing this thesis (Y).” Because my action of typing can be accurately incorporated into this intention-with-which it can be deemed intentional. Or I could do it differently: “I am typing this paragraph (X) with the intention of completing this

thesis (Y),” in which case my typing of the paragraph can be deemed intentional. The two sentences offer two (slightly) different descriptions of my actions. However, since in both cases the action in question is incorporated into an intention-with-which, both actions count as intentional. It is worth noting that these descriptions of my actions are not mutually exclusive, and that there is not necessarily a single definitive description here. That being said, one cannot take *any* description as being valid. For any description (and hence any attribution) to be correct, it must form part of the all-things-considered judgement of the agent in question. For example, if I am unconsciously tapping my foot as I type, the tapping cannot be correctly described as part of an intention-with-which of which the intention is to finish the thesis, and so cannot be intentional.

In light of these objections to premises (i) and (iia),<sup>41</sup> I submit that Knobe and Burra are mistaken in arguing that the intentionality of intentional action can be divorced from intention-for-the-future. Rather, I suggest that Anscombe was correct in demanding that a convincing account of intention must explain the unity of the three uses of intention. In addition, I conclude that later Davidson’s account of the unity between the three uses of intention is an improvement over Anscombe and his (Davidson’s) own early work. In particular, I think that his argument for the unity of the three uses of intention through the primacy of intention-for-the-future as a mental entity is the correct solution. I do, however, reject Searle’s refinement of this approach. More problematically, Davidson’s failure to provide a sufficient explanation of the content of intention-for-the-future as a mental entity leaves his account vulnerable to objections such as those of Knobe and Burra. Indeed, one of the most important endeavours in my own account of intention will be to provide a thorough explication of the content of intention-for-the-future as a mental entity, which I do in Chapter 5, Section 1.

### **3. The relationship between intention and practical reasoning**

As I have already mentioned, one of Davidson’s most important contributions to the discussion of intention is the notion of an “all-things-considered” judgement. In this section, I consider this type of judgement as part of practical reasoning. In light of the criteria for a convincing account of intention established at the beginning of the thesis, we need to ask three questions about the relation between intention and practical reasoning: (i) whether

---

<sup>41</sup> As well the objection to (iib) that I discuss in Section 5, and Chapter 5, Section 5.

intention is to be understood as a component of, a result of, separate from or contingently related to practical reasoning, (ii) what is the nature of the relationship between plural intentions and (iii) whether or not Davidson subscribes to cognitivism.

### *3.1 Intention as a rational judgement*

For Davidson, intention is certainly a result of practical reasoning. We start with a motivation (pro-attitude), and then work through a reasoning process (which entails epistemic awareness), the output of which is our intention / primary reason / all-things-considered judgement. The relationship between these steps is a necessary rather than a contingent one, and Davidson does not give credence to the idea of intentions that are formed without this reasoning process, even if it takes place on a level that is only clear when analysed retrospectively. In Davidson's (2001: 85) words:

One way to approach the matter is through a rather abstract account of practical reasoning. We cannot suppose that whenever an agent act intentionally he goes through a process of deliberation or reasoning, marshals evidence and principles, and draws conclusions. Nevertheless, if someone acts with an intention, he must have attitudes and beliefs from which, had he been aware of them and had the time, he *could* have reasoned that his action was desirable (or had some other positive attribute). If we can characterize the reasoning that would serve, we will in effect have described the logical relations between descriptions of beliefs and desires, and the description of an action, when the former give the reasons with which the latter was performed.

On a related note, remember the example of the stone being kicked in the road. For Anscombe, this was an example of an intention without a motivation; ergo without a practical reasoning process (since she maintained that practical reasoning is a movement from motivation to intention). For Davidson, this would be an example where the intention formed was still the result of a process of practical reasoning which we can explicate if need be, but one that occurred in a very brief time period and of which the agent was not necessarily aware at the time.<sup>42</sup> The difference stems from the fact that, for Davidson, to want to kick a

---

<sup>42</sup> Note, however, that this does not extend to awareness of the intention. I am still aware I intended to kick the stone, even if I was not aware of the practical reasoning process that preceded it. For Davidson, the epistemic criterion is that I must be aware that action X is in fact of the sort so as to bring about Y, where Y is what I intend (my goal). This does not require me to be aware of the process that led to the formulation of Y or the identification of X as the most appropriate (given my knowledge) means to Y.

stone is sufficient for it being counted as a motivation (this is the result of his very wide notion of pro-attitude), whereas for Anscombe this is not an adequate motivation.

### 3.2 *Apparent impossibility of plural intentions*

Furthermore, while Anscombe runs into difficulties because she fails to provide for the possibility of composite or plural motivations and intentions, later Davidson has a different, though related problem: his notion of an all-things-considered judgement reduces these composite or plural motivations and intentions to a single judgement. To illustrate, let us again consider Setiya's (2011a) example of someone who declares, "I intend to drink beer or wine." Now, for Anscombe this presents a particular dilemma on account of her commitment to intentional action as the paramount use of intention, which then requires that for any intention to exist there must be an identifiable, discrete action taking place. The problem with the above example is that it is unclear what action is actually being intended. Davidson's account, by contrast, does not depend on an identifiable, discrete action taking place, but rather asserts that an intention must be an all-things-considered judgement that can be wholly "in the head" (an intention-for-the-future only), so to speak. This allows him to escape the dilemma Anscombe's account could not. It does, however, raise its own difficulties. In the example given, we are to assume that "I will drink beer or wine" is an all-things-considered judgement, meaning that all foreseen alternative courses of action have been considered and that the one settled upon is indeed the most suitable method for the fulfilment of all considered motivations. This is obviously a matter of practical reasoning: I have a series of motivations, amongst which is the motivation to drink beer and the motivation to drink wine, but not to the motivation to drink beer *and* wine. Thus, while Davidson's account is an improvement over that of Anscombe in that he explicitly deals (or tries to deal) with the reality of plural *motivations*, he nevertheless still reduces any instance of plural *intentions* to a single intention. In fact, on Davidson's account it should be *impossible* to have multiple intentions at any given time, as every intention is the result of an all-things-considered judgement. Given that any such judgement must, by definition, consider all things, any conclusion formed would be final, without any space for alternatives. We should therefore ask whether "I will drink beer or wine" is a case of a single intention or a composite case where two intentions have been used in a single statement.

If we assume the latter – that it is a statement of two intentions – then it seems that I am declaring that I will be performing *one of two* mutually exclusive actions (and not that I irrationally intend to perform both of these mutually exclusive actions at the same time), although *I am unsure which*. It is this uncertainty that makes things difficult. If we take Davidson’s account seriously, we have to ask whether I can really be said to have an intention if there is uncertainty of this sort, since, for him, an intention is the result of a conclusive all-things-considered judgement. In that case, the above example would then be a case of an incomplete intention, or rather of an incomplete practical reasoning process. If this is correct, the moment the uncertainty is removed, the practical reasoning process can conclude and the two intentions collapsed into a single, clear intention.

But is this really how we think about our intentions: that we can only ever have one at a time? I would argue that failing to recognise the role of uncertainty in this aspect of human reasoning – not just in the process of forming intentions, but often also in their aftermath – can only lead to an inadequate understanding of the relationship between intention and practical reasoning. In this example, it seems much simpler to admit that, in fact, I have two intentions, and I haven’t yet decided between them. Given this uncertainty, I undertake (intentional) action in order to bring me closer to a situation where either goal can be achieved. One way of describing this might be to say that my practical reasoning (at this moment in time, given my current state and knowledge) could not give me a preference between beer and wine, and so the most reasonable path for me to follow is the one most likely to allow the possible achievement of both of these alternatives. Only when I reach a point where the game is zero-sum, and any action I take can be toward only one of the two options to the exclusion of the other, would it be irrational of me to still entertain plural intentions. For the purpose of identifying intentional action, I would propose that the intention-for-the-future that imparts intentional status to that action can be expressed as either a single intention or a composite statement of plural intention, such as “I am going to drink beer or wine.”

One potential defence of Davidson’s account might be that drinking beer and drinking wine are two different motivations, but not two different intentions, and that the intention I in fact form in this situation is still singular, namely: “I will act in such a way as to maximise the

possibility of drinking wine and drinking beer until this is impossible.<sup>43</sup> The fact that these motivations are mutually exclusive will only matter once this more provisional intention has been achieved. After this point, the intention will be realised, and a new intention will have to be formed by a new practical reasoning process, and in this case (assuming I don't suddenly crave a 10km run to clear my head from all my alcohol-related indecision) I will simply have to make a choice. This defence would allow us to retain the single intention approach.

My retort to this requires us to change examples, but before doing so I need to clarify something about the landscape of this argument. Davidson's account requires that there be only a single intention, whereas I argue that in at least some cases it is possible to have plural intentions. All that is required to overturn the argument that intentions must be singular is just one example where reduction to a single intention would fail to give an accurate account of the use of intention. Although I still consider the Setiya example to be a case of plural intentions, it is possible that a single intention theory could provide an accurate account of what is going on there. Hence, instead of trying to prove that the Setiya example is undoubtedly a case of plural intentions, I offer a different example where Davidson's possible counterargument would not apply. Consider this example: "I am buying groceries, with the intention of making a salad. I also intend to feed some to my pet rabbit, feed my pet hamster, as well as give some to my poor grandmother."

In this example, it is not possible for me to declare any one of the listed intentions as the primary or supervening intention. Each intention is, to some extent, discrete and their requirements for fulfilment differ. The only thing that seems to lend credence to the idea that a single intention is at play here seems to be the fact that I am undertaking only a single intentional action.<sup>44</sup> Furthermore, and equally important, each intention does not require the other, which makes this statement different from one such as: "I am sharpening my pencil with the intention of writing a paragraph, and the further intention of completing my essay." The latter statement is in fact one intention, as the final goal supervenes over the subordinate requirements. The former statement, however, cannot be construed in terms of subordinate

---

<sup>43</sup> Needless to say, it is unlikely that an agent will spell out the intention in as much detail, but it is conceivable that, if the agent were required to explicate her full intention, that this is the form it would take.

<sup>44</sup> On a side note, it is obvious, if perhaps unnecessarily pedantic, to point out that whether the action can be thought of as singular under scrutiny is itself not a question with an easy answer! That being said allowing for this does not problematize the arguments made here.

and supervening elements; each is distinct and stands or falls by itself. Their only similarity is that a certain action must be fulfilled as a necessary part of their achievement.

Crucially, the defence I provided for Davidson's account with regard to Setiya's example will not work in this latter case. My intention cannot be construed as singular without either claiming that it is impossible to maintain any of these intentions simultaneously, or treating all, or all but one, of the listed intentions as motivations (pro-attitudes) rather than intentions. As I show, the first solution leads to an untenable conclusion, while the second depends on an incoherent conception of the relationship between intentions and motivations.

The first solution would try to reconstruct the statement thus: "I intend to act in such a way as to maximise the possibility of achieving A, B, D and C<sup>45</sup> until this is impossible." On the face of it, this might seem quite convincing, but that is only until one realises that for this defence to hold, A, B, C and D must be mutually exclusive. This is because this solution must assert that at no time am I actually fulfilling two intentions simultaneously, as this would require the actual presence of multiple intentions! This would hold for the Setiya example, but not for my example. It is perfectly conceivable that some combination of A, B, C and/or D could be achieved simultaneously, in which case I might very well say: "I intend to act in such a way so as to achieve A and B." This formulation requires that two intentions be present. Remember Wilson and Sphall's example: "Veronica mopped the kitchen then with the intention of feeding her flamingo afterwards," which was shown to be a case of an intentional action and an intention-for-the-future. The statement, "I intend to act in such a way so as to achieve A and B" is a case where there might well be only one instance of intentional action taking place, but there are two distinct intentions-with-which in play.

As for the second solution – i.e. treating all or all but one of the listed intentions as motivations (pro-attitudes) rather than intentions – I deal with it thoroughly in Section 4. I will therefore not discuss it here in any depth, save to comment that an intention, as opposed to a motivation, involves a commitment to the future, and that it is at least arguable that A, B, C and D each involves such a commitment independent of the like commitments of all the others.

---

<sup>45</sup> Where A, B, C and D refer to: "making a salad; feeding my rabbit, feeding my hamster and giving some to my poor grandmother" respectively.

If this is the case, then Davidson's requirement that all variables be reduced to a single discrete intention is both unnecessary and inaccurate. Thus, while my analysis of Anscombe's account of intention showed that we do not necessarily require a single, discrete intentional action in order to have an intention, my analysis of the difficulties with Davidson's account shows that one can have more than a single discrete intention when performing an intentional action.

A further important problem with Davidson's account is that it treats an intention as essentially a species of rational judgement. This is different from Anscombe, who only requires that an intention be the correct answer to a certain kind of "Why?" question, namely one where the appropriate answer is the provision of a *reason* rather than a cause. For Davidson, it is an all-things-considered judgement, a judgement that he believes is the result of the process of practical reasoning. However, if we understand intention as the result of practical reasoning, this seems to dictate that only a single *rational judgement* should result: motivations are rationalised into reasons which I then reduce to a single rational judgement to guide my actions. I do not dispute that such a process takes place; I just do not believe that this process in of itself warrants the name "intention."

My reasons for thinking so have to do with the problem of irrational intentions, such as are present in some cases of *akrasia* and pathological behaviour. If intention is exactly the same as my all-things-considered judgment, then I must, by definition, intend what the judgment dictates. Yet this is not the way we think about intention. After all, it seems obvious that we can – and often do – hold intentions contrary to our rational judgements. Although I can rationally judge that reducing my consumption of unhealthy food would result in a better life situation for myself, the fact that I do not intend to do so is not a contradiction of the use of the term intend. Another way of making this point is that it is not a failure of the term "intention" to say "that man has an irrational intention to fly by flapping his arms." This being the case, my intention cannot be the kind of all-things-considered judgment that Davidson describes. Such judgements might well exist. But they are not equivalent to intentions. To give up the conception of intention as a rational judgement is to give up the belief that we can only possess a single intention in any particular instance, specifically when we think of intention as primarily a belief concerning action (as I argue in Chapter 5). Before we get to that point, however, there are still two aspects of Davidson's account of intention

that are relevant to this discussion of practical reasoning: the problem of bootstrapping and the charge of cognitivism.

### *3.3 The problem of bootstrapping and the charge of cognitivism*

The problem of bootstrapping once again originates with the conception of intention as a rational judgement. The problem runs like this: if an intention is a rational judgement, then it can be used as an input in future practical reasoning. It would count as a reason for the purposes of practical reasoning, alongside all the other reasons that contribute to the outcome of the reasoning process. Unfortunately, this has the effect of producing a circular argument, where a pre-existing intention (which may itself be irrational, if it resulted from a flawed practical reasoning process) can be a reason for itself. This is obviously problematic. I could, for example, have the intention to smoke. Now, if the reasons in my practical reasoning process concerning smoking were very finely balanced, then it is possible that my holding the intention to smoke could tip the balance of reasons in favour of smoking. So my practical reasoning would then conclude that smoking is the rational course, *merely because I have the intention to smoke*.

The general view is that in the course of practical reasoning, we consider various motivations as well as various reasons for or against a given conclusion. The goal of the process is to reach the most apt conclusion given the motivations to be satisfied and the reasons that I bring to the process.<sup>46</sup> So, for example, I can have the motivation “I want to eat a sandwich.” Now several reasons can come into play here, such as: “the closest source of bread is my kitchen.” The assorted reasons will be weighed and eventually the process should conclude on the “best” course of action to satisfy the relevant motivations.<sup>47</sup> The issue is that, if intentions are taken to be reasons for the purpose of further practical reasoning, then we have the ridiculous situation where any intention, once formed, becomes a reason for its own continuation.<sup>48</sup>

---

<sup>46</sup> These reasons are generally the results of pre-existing judgements of theoretical reason. However, this is a contentious matter, and one that I will not be intending to discuss in any detail.

<sup>47</sup> The choosing of given motivations over others also takes place during this process as through the analysis of the reasons it can become clear that certain motivations are exclusive, requiring a choice regarding the priority of motivations.

<sup>48</sup> As we will see in Chapter 4, Bratman recognises this problem, although he fails to escape it completely.

When it comes to the charge of cognitivism, there is not very much to say. Davidson is not a significant target for criticism from the anti-cognitivists, as he himself strongly refutes the notion of intention-as-belief – which, I remind the reader – is one of the key features of a cognitivist account of intention. As my own account employs the notion of intention-as-belief, I discuss Davidson’s arguments against intention-as-belief, as well as my counterarguments, in Chapter 5.

In conclusion, I still agree with Davidson that intention is a mental entity, but as the analysis of the relation between intention as he understands it and practical reasoning shows, his account of the *content* of this mental entity is not convincing. In particular, intention cannot be understood as a rational judgement *tout court*, even (or perhaps particularly) when defined as an all-things-considered judgment. Further, we have seen that a single intentional action need not have only a single supervening intention, just as Anscombe’s account showed us that an intention need not be linked to only a single discrete action. Lastly, we have identified a crucial unresolved difficulty: how to make sense of the role of pre-existing intentions in our practical reasoning processes while avoiding the problems entailed by intention-for-the-future being a reason for intentional action? This is a question that both Bratman and I attempt to answer, and I examine these answers in Chapters 4 and 5 respectively.

#### **4. The relationship between motivations, causes, reasons and intention**

For Davidson, the relationship and difference between motivation and intention is remarkably simple. As I have already mentioned, Davidson refers to motivations in relation to intention as pro-attitudes, which “are to be included desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind” (Davidson, 2001: 4). However, we can act upon a motivation without reasoning about it, and in such cases this is not an example of an intentional act, as no all-things-considered judgment was present. This is because, for Davidson, intention is a rational judgement and hence can only be present at the conclusion of the practical reasoning process. For example, if I am putting my hand on a hot stove-plate, and then involuntarily remove my hand, I certainly had a motivation, but I did not have an intention, as no reasoning process took place. How do we know that no such process took place? After all, the process of practical reasoning could simply have taken place

unconsciously, resulting in an intention. The answer is that if this were a case of intention then the subject would have been able to provide an account (even if only in hindsight) of the reasoning process that led to the hand's removal. If, as I presume will be true in most cases, the answer runs along the lines "I just did it reflexively" then clearly there was no intention present. What was present was a motivation, where motivation is taken as a pro-attitude.

#### *4.1 Intention as a pro-attitude and a cause of action: the problem of causal deviance*

One important difference between pro-attitudes and intentions is that, whereas a pro-attitude in some cases might lead to unmediated (hence unintentional) action, intention always entails a commitment to act. I can have a pro-attitude and not act upon it at all, whereas having an intention means having a commitment to act. This is the difference between saying "I want to" and "I intend to". Sometimes I act on my wants without thinking, but I am always at least committed to do what I intend, otherwise that is not what I intend. With this being said, it is important to keep in mind that for Davidson an intention always includes a pro-attitude as a component, so one way to look at an intention is as a very special kind of pro-attitude. An intention is a pro-attitude with certain additional requirements: in this case, the presence of the correct sort of rational judgement. This understanding of the relation between intention and motivation – that an intention is a pro-attitude with additional requirements – is the most accurate one, and I incorporate it into my own account of intention in Chapter 5.

Moving now to the question of whether intention is a cause. As I stated at the beginning of the discussion of Davidson's account, for him intentions are most certainly causes. To be sure, this is a controversial position which opens up Davidson's account to attacks from various directions, some of which lead down really deep rabbit holes. The obvious reason for this is that the concept of causality is itself a fertile source of debate, so that any light-hearted use of the proposition "X is a cause" easily runs into major difficulties – so much so that Russell famously said about the word "cause" that "its complete extrusion from the philosophical vocabulary [is] desirable" (Russell, 1912-1913: 1). As I made clear early on, it is not my intent to become embroiled in this debate, but it would be remiss of me not to acknowledge its significance. For the purposes of this thesis, however, I simply assume that causality functions in accordance with the "ancient-and commonsense-position [*sic*]" (to use Davidson's expression in a different context).

Having said this, I am not asserting that *intentions* should necessarily be considered causes. Recall that for Anscombe they most definitely are not causes, and her rejection of any identification of intention and cause was actually a boon to her account in some ways, as it allows her to sidestep the issue of causal deviance. This is precisely the issue that Davidson's account must meet head on, if it is to be taken as a convincing account of intention. To remind the reader one more time: causal deviance refers to the disjoint between achieving an intention and performing the action that that achievement requires. To use an example: I intend to blink to signal a friend, but because of my focus on my intention I am surprised by a loud noise, which causes me to blink involuntarily. The two questions this raises are: "Did I blink intentionally?" and "Did I achieve my intention?" For Anscombe, of course, this is a non-question. The only question we must consider is the "Why?" question. If somebody were to ask me, "Why did you blink?" then I would answer, "Because the noise startled me," and in Anscombe's view I would clearly be providing a *causal* explanation and not a *rational* explanation for my action, in which case there was no intention. Remember as well that for Anscombe the intention only exists at the point of action, which introduces all the complexities discussed in Chapter 2, Section 1.

But for Davidson, things work quite differently. In his account, I act intentionally when I am *caused* to do so by an intention. But in the blinking example it is still the case that I acted with an intention, and in fact I even acted thus *because* of the intention, but almost nobody would consent that the action in the example was performed intentionally! It appears that the "because" in this case is the wrong kind of "because;" in other words, my intention did not lead to the action in the right way. Despite recognising the problem of causal deviance, Davidson was, by his own admission, nevertheless unable to provide a solution (see Davidson, 2001: 79). If an account of intention is to maintain its commitment to intention as an effective cause – as my account will attempt to do – then this problem must be resolved or we run the risk of including too much in the denotation of intentional action. I present my solution to this difficulty in Chapter 5, Section 3.

#### *4.2 Intention as an unconditional judgement*

Turning then to the relationship between reason and intention, Davidson commits to the view that an intention is a rational judgement. Early Davidson identified intention with a primary reason, which was provided by a pro-attitude and a certain kind of knowledge

regarding action. By contrast, later Davidson equates reason with an all-things-considered judgement. This, however, is not the end of the story. In the face of criticism regarding the implications of his account of intention for questions related to *akrasia*, pathological behaviour and irrational behaviour more generally, Davidson altered his view concerning the kind of rational judgement with which intention should be identified. The new form of judgement he settles upon is called “unconditional judgement” (Davidson, 2001: 39). As the issues that lead him to this shift have to do with the epistemic requirements for intention, I discuss the shift in its entirety in the next section. For now, I will simply say that Davidson’s attempt to identify intention with rational judgement is one of the weakest aspects of his account, and that he would be better served to rather identify intention with a different kind of mental entity.<sup>49</sup>

At this point, I want to claim that Davidson’s account provides a solid base for understanding the relationship between motivation and intention; in fact, I consider this to be one of the strongest elements of his account. I also think that Davidson is right in declaring intention a cause. However, he clearly fails to address the problem of causal deviance. Finally, we have seen in this section that Davidson is committed to the view that intention is a type of rational judgement. In the following section, I analyse the problems with this view, in so far they relate to the epistemic requirements for intention.

## 5. Epistemic requirements

On the question of what I must know in order for this to count as having an intention, Davidson holds that I do not have to be able to identify the entire process that resulted in my forming the intention at the very moment of its formation, but that this process can be reconstructed retroactively. This position is far removed from that of Anscombe, for whom one of the defining characteristics of intention is that it requires a kind of special knowledge. Davidson does not think we require any kind of special knowledge about the reasoning process that delivered the intention. But in that case, what *do* we need to know? Davidson’s position certainly isn’t that we can possess an intention unknowingly; in fact, such a contention would probably be in contradiction with the very concept of intention as it is commonly employed. What we must know is that we have an intention.

---

<sup>49</sup> This is a view I share with Bratman (1984) and Broome (2001), although I disagree with the particulars of their arguments. I touch on this in more detail in Chapter 4, Section 3.3.

In Davidson's terms: I must be aware of my all-things-considered judgement, although not necessarily of what has led me to hold that rational judgement. I must, however, be able to reconstruct the process upon reflection. We have already seen that this manoeuvre helps us make sense of the difficult problem of seemingly spontaneous action. However, these positives notwithstanding, I would still argue that there is a major problem with Davidson's solution. This problem, perhaps surprisingly, is that despite moving away from Anscombe's argument for the necessity of special or privileged knowledge, Davidson still requires too great a degree of epistemic transparency.

For Davidson, our knowledge concerning intention must still be of a very definite sort. When I have an intention I must *know* that I am going to perform, am performing, or did perform it. And since intention is an all-things-considered judgement, when I have, or had, such a rational judgement, I cannot but act upon it. Otherwise it either was not an all-things-considered judgement, or intention has epistemic requirements different from such a judgement. This has very significant consequences for dealing with *akrasia* and pathological behaviour. In the case of *akrasia* – which Davidson understands as acting, and intentionally acting, against our rational judgement – Davidson's account does not seem as though it can give an accurate account of what is taking place. To return to the example I introduced earlier: I come to the rational, all-things-considered judgement that I should, for my own good, eat less unhealthy foods. According to Davidson, if I am aware of having this judgement, then I must also intend to eat less healthy foods. Note: not try to eat less healthy foods, or believe that I should eat less healthy foods. This is where cases of *akrasia* really throw a spanner into the works. After all, I can very well hold such a judgement, but it would not mean that I necessarily intend to act upon it.<sup>50</sup>

Davidson's (2001: 39-42) response to this actually moves him in a direction I will follow as well, although not in the same way. He comes to admit that there must be some degree of epistemic opacity in the process of practical reasoning, and therefore forsakes the attempt to identify intention with "all-things-considered" judgement and turns to "unconditional judgement" instead (ibid.: 39). He argues that there is no paradox in us acting against an all-

---

<sup>50</sup> Interestingly this situation has some resonance with the criticism Bratman makes against cognitivism, namely that the beliefs formed by the process of theoretical reasoning cannot be considered intentions as such beliefs do not entail a commitment to action. Such commitment is introduced by practical reasoning. As we will see, although I agree with Bratman that the conclusion of practical reasoning always yields a commitment to action, it does so precisely because it yields a "belief in trying."

things-considered judgement, as such action would still be in line with our unconditional judgement. In his words:

Every judgement is made in the light of all the reasons in this sense, that it is made in the presence of, and is conditioned by, that totality. But this does not mean that every judgement is reasonable, or thought to be so by the agent, on the basis of those reasons, nor that the judgement was reached from that basis by a process of reasoning. There is no paradox in supposing a person sometimes holds that all that he believes and values supports a certain course of action, when at the same time those same beliefs and values cause him to reject that course of action. If *r* is someone's reason for holding that *p*, then his holding that *r* must be, I think, a cause of his holding that *p*...[But] his holding that *r* may cause his holding that *p* without *r* being his reason; indeed, the agent may even think that *r* is a reason to reject *p* (Davidson, 2001: 40-41)

What he is getting at here is that, although an all-things-considered judgement is a judgement where “all truths, moral and otherwise” are taken into account, this actually means only “the sum of all that seems relevant to him [the subject]” (Davidson, 2001: 40). This would mean that even if all the reasons I take to be relevant indicates a certain choice, I could still choose to act otherwise based on considerations (reasons or motivations) that I have not *consciously* considered. These unconsidered considerations are opaque to me. In a rather delightful conclusion Davidson remarks on this: “What is special in incontinence [*akrasia*] is that the actor cannot understand himself: he recognizes, in his own intentional behaviour, something essentially surd” (ibid.: 42).

Although this defence succeeds in countering the original criticism, it opens up Davidson's position to a different problem. This problem is that, if Davidson is correct, an agent can never hold an intention opposed to the balance of reasons, taken to include unconscious reasons. This means that an agent can never perform an action that is irrational by her own reasoning, provided that we take her unconscious reasons into account. For example: After consideration of all the reasons I hold relevant I take *A* to be a better choice than *B*, but I still choose to intend *B* rather than *A*. According to Davidson this must be because I have an unconditional judgement that *B* is in fact better than *A*. This means that I may, at the time, *think* that I am acting irrationally in the sense that I intend *B* even though I have no reasons (that are transparent to me) for doing so. However, if the

unconscious reasons were explicated, then this would no longer be a case of me acting irrationally, as in fact I did – and had to – act in accordance with the balance of reasons that includes my unconscious reasons.

The issue is that Davidson cannot allow us to fail to intend in line with our rational judgement, as this would require that intention be something other than just such a judgement. This means he has to argue that I *must* intend in line with my unconditional judgement. This is not convincing, however, as it is entirely plausible that I can act in opposition to the balance of reasons (which is exactly what *akrasia* is taken to show), be this balance transparent to me or not. For this reason, Davidson's epistemic requirements for intention, which follow from his requirement that intention be a rational judgement, are not convincing. An easy remedy to this problem could be found if Davidson were prepared to shift intention away from being a reason judgement. Taking intention as a belief means that intending against the balance of reasons would no longer be problematic. There are difficulties reconciling this view with one in which intention is taken to be an outcome of practical reasoning, but I believe these to be surmountable, and will address them when I outline my own account of intentions and *akrasia* in Chapter 5.

I turn now to the issue of pathological behaviour. For the purpose of explicating the problems that Davidson's account suffers from in dealing with such behaviour, I rely strongly on the argument made by Galgut (2011). In this argument, Galgut provides three examples of pathological behaviour. These examples are taken to be cases of irrational intentional behaviour, characterised by the agent acting irrationally due to the presence of unconscious considerations. They are a problem for Davidson, as in these cases his account cannot explain how the actions can be irrational *and* intentional. I will describe the examples and then discuss the argument.

Example A concerns the case of Lady A, who has a jealous husband. Lady A dances at a party and consequently her husband accuses her of being a "harlot." The next day she throws herself from a cart and breaks her leg. She reports that she jumped because she was overcome with fear that the horses would bolt. Freud argued that she was acting with the unconscious purpose of punishing herself for her "unladylike" behaviour, and simultaneously removing

the ability to dance.<sup>51</sup> He noted her seeming lack of negative response to the injury as evidence to support his position. This was also taken as the event that triggered a series of neurotic episodes.

In Example B we encounter the Rat Man, another of Freud's patients. In this case the Rat Man is awaiting the arrival of his fiancée via carriage. While so waiting he notices a stone in the carriage's path. Filled with concern that the stone may cause the carriage harm, and so harm his fiancée, the Rat Man removes the stone from the road. As he walks away "he thinks to himself – "what a foolish thing to have done" – and replaces the stone where he originally found it" (Galgut, 2011: 134). Galgut points out that although it might be tempting to see the returning of the stone as a rectification of the pathological behaviour of its removal, this is misunderstanding the situation. After all, the Rat Man seems to be under a sort of compulsion to return the stone to its original place, as opposed to simply dropping it or tossing it aside. Galgut thus contends that the "explanation of the Rat Man's removal and replacement of the stone in the road was that the apparent change of mind was a surface *redescription* of a deeper unconscious process of which the Rat Man was unaware" (ibid.). This may seem to support Davidson's position, but, as we shall see, this is not the case.

In the final case, Example C, we have a Mr X who reacts to his mother's death by writing down his dreams about her in a notebook. Mr X's psychoanalyst, Hanna Segal, described this behaviour as "mental defecation" (Galgut, 2011: 135) and associated it with another of Mr X's pathological habits which was his habitual act of going to the bathroom whenever he felt hungry, as though the "perception of hunger was a bad thing inside him that he could get rid of by defecating" (ibid.). Mr X had numerous rationalisations for these actions.

As I mentioned previously, these three examples are all cases where the agents are considered to have acted intentionally, but also irrationally. This once again poses the question as to whether Davidson's account of intention, with its commitment to intention as a reason judgement, can make sense of these cases. Davidson might argue that there is no problem in assuming that there are unconsidered motivations and reasons that move the subject to a

---

<sup>51</sup> I am not, by using these examples, intending to endorse or support any aspect of Freud's work. I am using these examples as they are those employed by Galgut, their link to Freud is inconsequential to my arguments. In a similar vein, my use of the notion of the unconscious is not intended to relate to Freud. A number of thinkers that I discuss in this thesis make use of the notion of the unconscious, and it is in relation to their understanding of this concept that I employ it myself (see Davidson 2001, Bratman, 2009b and Galgut 2011 as examples).

certain action, such as the unconscious desire of Lady A to break her leg as punishment and prevention. What we have here is still a case where we have a pro-attitude (even if it is unconscious), which then goes through a process of practical reasoning (after all, it still had to be decided that jumping from the cart was the most effective means of achieving the end), which yields the unconditional judgement (and hence the intention) to jump from the cart. These cases seem to show that this process can be reconstructed in hindsight; ergo the epistemic requirements are met.

What, then, is the difficulty here? The difficulty is that any Davidsonian explanation of the pathological behaviour faces the same problems that plague his explanation of *akrasia*. The three examples were taken to be cases of irrational behaviour, so once again it seems that Davidson has to argue that this irrationality is only an *apparent* irrationality, and that if we properly explain the hidden variables then the agent's course of action can be rendered rational. This hidden rationality can be explained in terms of the distinction between all-(relevant)-things-considered and unconditional judgements. Galgut (2011: 136) refers to this as a "partitioning [of] the mind," whereby the separate, "coherent and rationally consistent" parts of the mind that are sometimes opaque to us give the appearance of irrationality. This is another way of saying that we can appear irrational when we act against our conscious best judgement, but if we were to know what our unconditional judgement was then the rationality would be obvious.<sup>52</sup>

In the face of this argument from Davidson, Galgut (2011: 136) then asks: "although, once we are provided with the explanations of the Rat Man et al.'s behaviour, it is certainly more *understandable* than it was before, is this sufficient to render the behaviour *rational*?" Her answer here is no, and I am in full agreement. Even once the so-called unconditional judgement of Lady A, Rat Man and Mr X are explained to me, it does not render their actions rational, merely understandable. This scuppers Davidson's position, as he requires the presence of a rational judgement of some sort in order to justify taking these actions to be intentional. Davidson's concession to some level of opacity about our decision making is insufficient for making sense of cases such as those discussed by Galgut.

---

<sup>52</sup> Note that the use of the term rational here means "rational given the information available to the subject." Obviously, even unconditional judgements can be irrational if they were founded on false information. However, they are not irrational in the sense that I know what the rational thing to do is, and then act against it. It is this latter type of irrational behaviour that characterises *akrasia* and pathological behaviour, and it is this type that Davidson's arguments are intended to circumvent.

As I mentioned, the issues surrounding pathological behaviour have certain similarities to those involving *akrasia*. In both cases we find instances of “irrational” intentional behaviour, which is obviously problematic for any account of intention that presents intention as a rational judgement. I have also argued that in both cases Davidson’s account falls short due to the epistemic requirements that his account of intention demands. I have further contended that these demands are the result of his commitment to intention as being a rational judgement. It will be my position that these problems can be overcome through understanding intention as a belief, which will have the effect of easing some of the stringent – and ultimately self-defeating – epistemic requirements inherent in Davidson’s account of intention.

## **6. The relationship between intention and (moral) responsibility**

Davidson never wrote directly on the relationship between intention and responsibility. In this section, I attempt to reconstruct a possible approach to this relationship that his account of intention would have allowed for. I then examine this reconstruction in order to gauge its ability to answer the questions I put forward in Chapter 1: Are we responsible for all the consequences of intentional action? and Are we responsible for anything we do unintentionally? In this section, I argue that, at best, Davidson’s account of intention only provides us with an ambiguous answer to both of these questions.

### *6.1 Reconstructing Davidson’s views on intention and moral responsibility*

The first step in my reconstruction is to consider the implications of Davidson’s account for understanding the moral status of different types of consequences to intentional actions. As we saw in the discussion on Anscombe, there are three broad types of consequences for an intentional act, namely intentional, foreseen and unforeseen consequences. As I show, Davidson’s account collapses the categories of intentional and foreseen into each other, while his understanding of unconditional judgements makes it very difficult to judge moral responsibility in either case. On the other hand, the status of unforeseen consequences still remains in line with the general intuition, that they are not things for which we can be held responsible.

As we have seen, Davidson's account requires any given intentional act to be the result of a rational judgement that, at the very least, contains all the things considered relevant by the agent, and, at the most, is an unconditional judgement that even accounts for (unconscious) considerations that the agent did not consider at the time of the action. Either way, it is not obvious what about any given consequence would make it count as the intended consequence to the exclusion of all the others. If we cannot make this distinction based on the epistemic status of the consequence, we are left with two possibilities: either we say there is a set of foreseen consequences that contains all intentional consequences, or we admit that all foreseen actions are intentional and only unforeseen consequences are unintentional.

The trouble with the first answer is that it still leaves us with the problem of distinguishing intentional foreseen consequences from the other foreseen consequences. Davidson might argue that foreseen consequences that are intentional are those consequences that relate directly to the pro-attitude responsible for setting the whole ball rolling. So if my pro-attitude was to kick a stone for pleasure, then my intentional and foreseen consequences would include the fact that the ball moves and the sensation of contact in my foot, as well as the impact of the stone with a target, provided I intended to strike the target. The rest of the foreseen consequences – such as that my friend who also wanted to kick the stone is no longer able to do so and that the kicked stone left an indentation on the target I was aiming for – would be unintentional. Another way of explaining this approach to the difference between those foreseen consequences that are intentional and those that are not, would be to say that the only consequences that count as intentional are those that were *intended*, the rest are merely foreseen.

However, this solution flies in the face of the everyday use of the terms involved. As Knobe and Burra's (2004) study shows, people generally do not use the terms "intentional" and "intended" to describe the same sets of consequences. Although intended consequences are always intentional, not all intentional consequences are intended. The difference is that, whereas the term "intended consequences" match the ones described in the preceding paragraph, the term "intentional consequences" has a broader application. This broader application appears to cover a lot of foreseen but not intended consequences. For example: When the CEO causes harm to the environment in the name of profit, this is an intentional consequence. It is also a foreseen consequence, but it is not an intended consequence.

The objections to the second approach once again take us to the discussion about irrational behaviour. Since Davidson cannot allow a decision to be both intentional and irrational, he creates a partition in our reasoning between the transparent (to the agent) considerations of our all-things-considered judgement, and the opaque and unconscious considerations (as well as the transparent considerations) of our unconditional judgement. By doing so, irrationality can then be described as the failure of an agent to be fully aware of our own reasoning, caused by the opacity of some of the considerations involved. The question that concerns us in this section is whether this opacity has any influence on whether a set of consequences are deemed foreseen or unforeseen. In other words, if my intention is an unconditional judgment, where much of the reasoning is opaque to me, then are the consequences of that action foreseen or not?

If by “foreseen” we mean “consciously foreseen” – as in: the subject could, if answering truthfully, attest that a given consequence was predicted when the action was undertaken – then clearly only those consequences that factored into my all-things-considered judgement can be deemed foreseen, while the consequences of the unconscious considerations of my unconditional judgement cannot. But remember that our intention is identical to the unconditional judgement, so if this line of reasoning is correct, it would mean that the consequences that are foreseen only match our intended or intentional consequences by happy chance. This means that there might be intended (not just intentional) consequences that are nevertheless *not* foreseen, a position that is strikingly counter-intuitive. In such a case, I would presumably not be responsible for that consequence. Indeed, it is likely that there would be many intentional consequences for which I would then not be responsible. This would be less of a problem were it not for the fact that in our everyday use of the term, the concept “intentional” almost always denotes responsibility.

## *6.2 The problem of irrational intentions*

As I mentioned before, Davidson never wrote directly on the problem of responsibility, which makes it difficult to see how he could have resolved this impasse. However, it might be possible to deduce a defence of the implications of his theory for responsibility from his extant writings. Let us recall that the reason for Davidson’s partition was that he sought a remedy to problem of irrational intentions (such as in cases of *akrasia* and pathological behaviour). Would Davidson not be justified in claiming that the above-mentioned view –

that there might be intended consequences that are nevertheless not foreseen by the agent and for which we would therefore not hold her (morally) responsible – provides us with just such an explanation?<sup>53</sup>

There is a difficulty here, however. Since in all cases, and not just in cases of pathological behaviour, our intention is our unconditional judgement, would Davidson's argument not then imply that we are always first required to give a detailed explanation of all the unconscious processes that might be in play before we would be justified ascribing any kind of responsibility to anyone at all? More to the point, would instances of *akrasia* also be absolved of responsibility on the same grounds as those who suffer from pathology? And if we try to move in the opposite direction and say that we are responsible for all the intentional consequences of our actions, including the ones that are unforeseen due to the opacity of elements of our unconditional judgement, then we encompass far too much. Surely a kleptomaniac is not responsible for the theft that she commits in the same way as a thief who steals out of sheer avarice. Indeed, it is not clear that the kleptomaniac is really morally responsible at all.

This leaves us with the conclusion that Davidson's account cannot provide a solid foundation for making sense of the relationship between intention and moral responsibility. The difficulty here, which is indeed the greatest difficulty with his account of intention as a whole, is that he remains committed to intention as a rational judgement. This leads him to attempt an abortive solution to problems of irrationality, which renders any attempt to make sense of moral responsibility in a way that conforms to our everyday use of intention impossible. Although I have shown that Davidson's account is an improvement on that of Anscombe in a number of aspects, and I recycle many of his insights in my own account, I submit that we cannot rely on his account of intention when thinking about responsibility.

## **Conclusion**

The most important conclusion to be drawn from this chapter is that intention understood as a rational judgement is an untenable position. It is Davidson's commitment to this view that

---

<sup>53</sup> Obviously it is not my argument that all varieties of pathological behaviour necessarily follow this course. In some cases it is probably the case that the actions are simply involuntary and ergo not intentional in any way. What is meant here are cases where we would normally say that intention is present in the action, but the intention attested by the agent is not in fact the actual intention acted upon.

resulted in difficulties for his account discussed in Sections 3, 5 and 6. Whatever else intention might be, it cannot be a rational judgement. The second important conclusion to take from this chapter is that the best chance for achieving the unity of the three uses of intention is through the primacy of intention-for-the-future. Early Davidson's attempt at a unity through intention-with-which failed due to its inability to account for cases of pure intending. Only a unity through intention-for-the-future can incorporate such cases. It should be no surprise that this position has become all but universal in the discussion around intention. Of lesser, but still crucial, importance is Davidson's identification of intention as a kind of mental entity – one component of which is a pro-attitude – as well as a cause. I take each of these to be a significant step forward. However, there are still several difficulties that must be overcome, but which Davidson's account fails to do. These are: (i) addressing Searle's argument for the necessary role of "intention in action," which is incompatible with a unity of the three uses of intention through reduction to intention-for-the-future; (ii) explaining the relation between pre-existing intentions and practical reasoning while avoiding bootstrapping; (iii) resolving the problem of causal deviance; (iv) resolving the problem of irrational intentions; (v) countering Knobe and Burra's alternative explanation for the three uses of intention and (vi) explaining the relationship between the different types of consequences of intentional action and moral responsibility. As will become clear, the account of intention I develop in Chapter 5 is meant to address all of the aforementioned difficulties. First, however, we need to delve into another influential theory of intention that follows on where Davidson left off. That is the theory developed by Michael Bratman.

## CHAPTER 4: BRATMAN'S ACCOUNT OF INTENTION

### Introduction

There has been no contemporary thinker who has made a more concerted attempt to provide a convincing account of intention than Michael Bratman. Beginning his engagement with the question at the time when Davidson's arguments, and its derivatives, were very much the orthodoxy, Bratman has severely criticised what he calls the "Simple View" of intention (Bratman, 1984: 377).<sup>54</sup> His definition of the Simple View is: "for me intentionally to A I must intend to A; my mental states at the time of action must be such that A is among those things I intend."

While brief, I consider this to be a fairly solid summation of Davidson's view of intention.. Having developed a critique of this Simple View, Bratman then tries to provide a viable alternative. He calls this alternative the "planning theory of intention" (Bratman, 2009b: 411), or as I will at times refer to it, intention as plans. Thinking of intentions as components in plans that have certain consistency and coherence requirements is probably Bratman's most significant contribution to the discussion around intention (Velleman, 1991: 277). This is evinced by the fact that the nature of these requirements has become one of the most contested aspects of the recent literature on intention. Much of this has taken the form of criticism of Bratman's version of the requirements. As I will show, while Bratman has proffered some responses to such attacks – most notably in his more recent work such as *Intention, Belief and Instrumental Rationality* (2009a) and *Intention, Practical Rationality and Self-Governance* (2009b) – these responses, though powerful, do not address all the underlying problems.

Out of the three accounts of intention I analyse in this thesis, it is Bratman's that places the least emphasis on resolving the issue of the unity of the three uses of intention. In Section 1 of this chapter, I consider his limited attempt at a solution in the form of his argument for the

---

<sup>54</sup> It is interesting to note that Bratman did not include Davidson himself in the list of thinkers he identified as advancing the Simple View. He explains that one of the requirements of the Simple View is that its advocates argue for too strong a link between intention-for-the-future and intentional action, and he does not believe that Davidson commits to such a view. In Bratman's (1984: 381) own words: "I did not include Donald Davidson among those who accept the Simple View, even though he comes close to endorsing the view that if I intentionally start my car then I must intend my particular act of starting it." I would argue that this is mistaken generosity on Bratman's part, as there is strong evidence that Davidson is in fact committed to such a view. Indeed, I believe that my arguments in Chapter 3 have shown this to be the case.

“Single Phenomenon View” (Bratman, 1984: 377), and most importantly, the explication of what he calls the “*motivational potential*” of an intention (ibid.: 395). I argue that the Single Phenomenon View is certainly an improvement over the Simple View, but that it does not successfully account for the three seemingly irreconcilable uses of intention – most notably, the link between intention-for-the-future and intentional action. On the positive side, I will show that Bratman’s Single Phenomenon View allows us to posit such a link nonetheless, as long as we are prepared to give up what Bratman terms the “*assumption of tight fit*” between intention-for-the-future and intentional action (Bratman, 1984: 394) and take into account the “motivational potential” of an intention. On the matter of the relationship between intentions, I will argue that Bratman delivers the best solution to the relation between these relations through his use of the notion of plans, although I argue against his adoption of the notion of “settled objectives” (Bratman, 2009a: 18-19) as well as for a clearer account of the role for intention-with-which.

Section 2 deals with Bratman’s take on the epistemic requirements for intention, which I will show is intimately linked to his views concerning intention and practical reasoning. This is due to the crucial role played by the notion of “belief”. Bratman introduces it as a requirement that our intentions must be consistent with our beliefs, but he strongly resists identifying an intention to X with a belief to X. I will show that this approach allows us to resolve the problem of epistemic opacity – including many of the difficulties related to *akrasia* and pathological behaviour – in a way that Anscombe’s and Davidson’s accounts could not. While this approach moves the discussion in the right direction, Bratman also argues that it is possible to have an intention to X without having a belief to X, although you cannot have the belief to not-X. I argue that if this claim is correct, it becomes impossible to justify any account of intention that requires belief in what is intended or identifies intention as a type of belief. Consequently, I argue that Bratman’s assumption on this score is incorrect.

Following my practice thus far, Section 3 deals with the relationship between motivation, causes, reasons and intention. Here I show that Bratman agrees with Davidson (and hence disagrees with Anscombe) that intentions are a kind of motivation and that they are causes for action. He does not, however, consider intention to be a reason for action. On the topic of motivation, Bratman follows Davidson in seeing intention as a pro-attitude, “distinct from the agent's desires and beliefs” (Bratman, 1984: 393). In particular, he argues that what

characterises this pro-attitude is that it is conduct-controlling and that it is open to a set of rational requirements that we do not usually require of other kinds of pro-attitudes (such as desires). The discussion of these requirements will form the bulk of this chapter.

I begin my discussion of practical reasoning in Section 4 by noting the influence of David Gauthier on Bratman's understanding of the rational requirements for intention – specifically the requirement of means-end coherence, understood as a deliberative procedure of practical reasoning. My main aim here is to criticise Bratman's argument that a commitment to “self-governance” is the substantive reason for why we, as agents, strive to adhere to means-end coherence in our intentions. Besides the issue of means-end coherence, this section further examines the validity of Bratman's criticism of cognitivism. Here I show that Bratman's critique of cognitivism can be reduced to three points: (i) intention can occur even when no belief to so intend is present, hence intention does not entail belief; (ii) it is possible to falsely believe the one intends an end when in fact one does not and (iii) cognitivism, in so far as it reduces practical reasoning about intention to theoretical reasoning about the truth of our beliefs cannot explain the motivational quality of intention (the fact that an intention is a pro-attitude). Having explained these three points, I demonstrate that (i) and (ii) are, in fact incoherent, while (iii) is substantially correct.

Bratman has written directly, if sparsely, on the nature of responsibility. In Section 5, which deals with the relationship between intention and responsibility, I begin by outlining some points made by Claire Finkelstein regarding the requirements for the ascription of moral responsibility to agents. In light of these requirements, it then becomes clear that Bratman's inability to explicate the nature of the relationship between intention-for-the-future and intentional action means that he cannot provide a convincing account of the relationship between intention and responsibility. I further demonstrate that, while a consideration of the motivational potential of an intention allows us to improve our understanding of ascriptions of moral responsibility in cases of *akrasia*, failed intentional action and irrational intentions, Bratman's account cannot make sense pathological behaviour. I also consider his seeming commitment to some degree of ethical rationalism as evinced by his worry about the possibility of rationally justified immoral intentions. Overall, I conclude that, although Bratman's account of intention provides a number of useful concepts for thinking about responsibility, his own use of these concepts – particularly the notion of self-governance – leaves us with much work to do.

## 1. Explaining the unity of the three seemingly irreconcilable uses of intention

As I have already mentioned, Bratman follows Davidson in seeing the key to unifying the three uses of intention in intention-for-the-future as a mental entity. As with Davidson, this means that intentional action and intention-with-which are combined into intention-for-the-future. On the simplest level, this approach requires that intentional action and intention-with-which can only be present if the mental entity intention-for-the-future is also present. However, the two theorists differ considerably on the question of the *content* of this mental entity.

Whereas Davidson does not spell out the content of intention as mental entity in any great detail, Bratman does provide an extensive account of what this mental entity is meant to entail. For Bratman, intention is a pro-attitude which, as opposed to other pro-attitudes such as desire, is subject to a series of conduct-controlling regularities. These regularities, which Bratman (2009b: 413) labels “norms of practical rationality,” are what structure our intentions. These norms are: (i) intention is a commitment to act (it *is* a pro-attitude itself), (ii) means-end coherence and (iii) consistency between intentions and beliefs (Bratman, 1987; 2009a). A mental entity can only count as an intention in so far as it aims at fulfilling these norms. That is to say that, while an intention might fall short of these requirements, it must *attempt* to satisfy them for it to count as an intention at all (Bratman, 2009a; 2009b). An intention that fails to satisfy these requirements can be criticised as an irrational intention. That being said, even an irrational intention makes an *attempt* at consistency and coherence, and so such requirements always constitute part of the mental entity of intention.<sup>55</sup> Finally, Bratman maintains that intentions are components of plans. By “plans” he means: “mental states involving an appropriate sort of commitment to action; I have a plan to A only if it is true that I plan to A” (Bratman, 1987: 29). Plans, in other words, are “intentions writ large” that function as “conduct controllers, not merely potential conduct influencers; and they provide crucial inputs for further practical reasoning and planning” (ibid.).

---

<sup>55</sup> The norms of means-end coherence and consistency between intention and beliefs will be examined in Sections 2 and 4 below. For the purposes of this section it is sufficient to know that they form part of intention.

### *1.1 Reduction of the three uses of intention to intention-for-the-future*

Having a more fully fleshed-out conception of intention as a mental entity gives Bratman's theory several benefits over those of Anscombe and Davidson, as well as leading to a number of potential weaknesses. There are two positive points to consider: First, the content of intention that Bratman provides brings us closer to resolving the relation between intention-for-the-future and intentional action – a problem that I argued Davidson's account could not resolve. Second, because intention is understood as part of a plan, Bratman is better able to deal with the issue of plural intentions, something that neither Anscombe's nor Davidson's theories was able to address. On the negative side, Bratman's proposed unity of the three uses does not go deep enough. His account does not make it clear how we are supposed to understand intention-with-which, as he never explains how this use of intention relates to the other two (i.e. intention-for-the-future and intentional action). Furthermore, although his solution to the unity of intentional action and intention-for-the-future is a step in the right direction, it serves only to explain how the two *can* relate; it provides no reason for thinking them *necessarily* related.

For Bratman, the problem of the unity of the three uses centres on resolving the gap between intention-for-the-future and intentional action. This should not be too surprising, considering that this is precisely the gap that causes problems for Anscombe's and Davidson's theories. In fact, Bratman's account is overcomes this gap. According to Bratman, the Simple View's explanation of the relationship between intentional action and intention-for-the-future requires too direct a connection between the two types of uses. In particular, this explanation leans on what he calls the "assumption of tight fit," which he describes as follows:

They both [the Simple View and the Volitional Thesis<sup>56</sup>] assume that if there is a distinctive pro-attitude involved in intentionally A-ing, it will be a pro-attitude specifically in favor of A – that there must be a tight fit between what is done intentionally and what is intended (willed). (Bratman, 1984: 394)

---

<sup>56</sup> The Volitional Thesis is another attempt at providing a convincing account of intention that Bratman criticises in *Two Faces of Intention* (1984). I will not be examining it in my arguments here, as I take Bratman's criticisms of this kind of account of intention to be accurate, and as such I take the Volitional Thesis to have already been shown to be unconvincing.

The assumption of a tight fit is that it seems to prevent many actions that should otherwise be considered intentional actions from being recognised as such. For example: I have the intention to make a salad, and as part of doing so I chop carrots. Now, conventional usage would dictate that my chopping of the carrots is described as an intentional action. The problem Bratman raises is that if we take the Simple View seriously, then it is not at all clear that we can do this. In order for me to be doing A intentionally I must have an intention to A, or A must be what I intend. In the example, however, what I intend and my intentional action do not share such a tight fit since, on the face of it, my intention is to make salad, not to chop carrots. If this is true then it represents a problem for the Simple View.

A possible defence of the Simple View is that the case described is a case of intention-with-which, and that the relation between such intention-with-which and intentional action is different from the relation between intention-for-the-future and intentional action. After all, I could say that my intention is not “to make a salad,” but that “I am chopping carrots with the intention of making a salad.” In this case, the action could be deemed intentional, provided that the action that is intended is widened to include the component steps that must be taken to accomplish it. In terms of this reasoning, chopping the carrots is subsumed under the overarching action: making a salad. This defence is convincing, I think, and it helps to show that the Simple View is not merely a flimsy stand-in of a theory, but a view with some depth. It is for this reason that the example that Bratman uses to refute the Simple View is quite an outlandish one – though not, I would argue, to the point where it should be considered unconvincing.

To make his point against the Simple View, Bratman uses the example of a gamer playing two games simultaneously, in each case trying to strike a target (T1 and T2 respectively) with a missile. The games are set up so that striking either of the targets makes it impossible to hit the other, so although the gamer is trying to hit both the targets, he cannot hit both. The problem here is that it seems that the gamer cannot intend to hit T1 *and* T2, as these are two mutually exclusive goals. In this case, I cannot formulate my intention as a statement of intention-with-which. In no way does approaching T1 bring me closer to T2 or vice versa; I cannot subsume them under one action or goal. It is obvious that the “beer or wine” example from the previous chapter would be an almost identical case to this one, with the only meaningful difference that the gamer can pursue both goals up until the absolute limit, whereas this is less likely for the indecisive drinker. Yet this is a crucial difference. Because

at no time in the process does progress toward T1 bring me closer to T2, I am always in a state of seeming contradiction if I intend both.<sup>57</sup>

How can we then make sense of the common intuition that as he is playing, the gamer is intending to hit both T1 and T2, and that if he successfully strikes one then he did so intentionally, and that in undertaking this course he is not being irrational? As I have already said, Bratman's first step is to point out that the Simple View cannot resolve this gamer example, as this view requires too tight a fit between intention-for-the-future and intentional action. The solution Bratman posits is to relinquish the requirement of a tight fit. He proposes instead that we understand an intention as (usually) part of an overreaching plan, and that such a plan can incorporate plural intentions. Actions undertaken in the context of this plan can be deemed intentional if they were undertaken due to the presence of an intention-for-the-future, even if this intention-for-the-future is not directly linked to the action. This leads Bratman to assert the following:

I propose to give up the assumption of tight fit and to distinguish between what is intended, and the sorts of intentional activity in which an intention may issue. *Making this distinction, we can say that when I A intentionally I intend something, but I may not specifically intend to A* [my emphasis]. Our notion of intentional action embodies a complex scheme for the classification of actions (or, perhaps, actions 'under a description'). To understand the relation between intention and intentional action we must recognize that the factors that determine what is intended do not completely coincide with the factors that, on this scheme, determine what is done intentionally. (Bratman, 1984: 394)

A simpler formulation of this same point would be: "An action X is performed intentionally (is a case of intentional action) if, and only if, it is undertaken due to an intention(-for-the-future) Y". In light of this formulation, we can describe the gamer example in a different way. If I have a plan to hit either target T1 or T2, and my intention-for-the-future is formulated as "I intend to hit at least one of the targets" (let this be Y<sub>1</sub>), I can then call the actions undertaken in attempting to hit T1 and T2 (let this be X<sub>1</sub>) intentional, as they take

---

<sup>57</sup> This would be the case regardless of whether my intention was singular or plural. If singular, I would be holding an intention that is impossible, while if plural, I would be holding two contradictory intentions, which would be irrational. As we would not normally call the gamer irrational in everyday life, it seems wrong to declare him so based on theory, since the theory is meant to reflect our everyday usage.

place under the auspices of  $Y_1$ . The differences between this and the Simple View can be subtle, but we should not underestimate their importance. Whereas on the Simple View, my action of trying to hit T1 could only be intentional if I had the intention-for-the-future to hit T1, on Bratman's view I can hit T1 intentionally even if my related intention-for-the-future is not so direct. This opens up possibilities for me to intentionally strike either target without having to have the particular intention of striking either, which avoids the entire line of reasoning that leads us down the path towards irrational outcomes.

### *1.2 Two criticisms of Bratman's reduction of the three uses of intention*

The above-mentioned solution to the problem of the three uses is one that I find convincing to a degree, although I believe that it could be improved upon and will attempt to do so in Chapter 5. Others have not been so convinced.<sup>58</sup> Indeed, a good number of my proposed improvements are meant to resolve one of the most significant problems with Bratman's account. This is the problem that he widens the gap between the intention-for-the-future and the intentional action to such an extent that it is not at all obvious whether and how the connection between them is supposed to work. This means that Bratman's account can be criticised in one of two ways: either he includes too much in his definition of intentional action, he includes too little.

Given Bratman's statement of the relation between intention-for-the-future and intentional action which I have quoted above, the "too much" criticism is that having any intention while performing an action would qualify as sufficient for the action to be deemed intentional. In effect X could be intentional if there is an intention Y, regardless of the content of X or Y. This is obviously unsatisfactory in terms of the provision of reasons. If I have the intention to pay my taxes and while having this intention change the channel on my television, it seems absurd to think that the action is intentional because of this intention, let alone that I could provide "I intend to pay my taxes" as a reason for the intentional action. We would most likely think that the relation between our intention-for-the-future and intentional action depends on more than their mere coexistence in time; that is, there should be a *rational* rather than a purely *temporal* connection between them.

---

<sup>58</sup> See McCann (2005) as a clear and unambiguous example of defence for the Simple View. For criticisms of Bratman that do not entail defences of the Simple View see Velleman (1991) and Setiya (2011b) amongst others.

Bratman is aware of this problem and tries to resolve it by positing the “*motivational potential* of an intention” (Bratman, 1984: 395). By this he means that whenever we possess a given intention-for-the-future, there will be a certain set of actions that can be correctly described as intentional. This set will fall within the motivational potential of the intention-for-the-future in question. Taken on its own, this strategy is a delay, not a solution, as it merely shifts the question of the relationship between intention-for-the-future and intentional action to the question of the rules that govern motivational potential. However, the lack of content to motivational potential is not accidental. Bratman does not intend for motivational potential to solve the problem in its entirety. In his words:

The notion of motivational potential is intended to mark the fact that my intention to B may issue in my intentionally A-ing, not to explain it [the account of how this issuing occurs]. It is a *theoretical placeholder*: it allows us to retain theoretical room for a more complex account of the relation between intention and intentional action while leaving unsettled the details of such an account. Such an account would not itself use the notion of motivational potential but would, rather, replace it with detailed specifications of various sufficient conditions for intentional conduct. (Bratman, 1984: 369)

Bratman does, however, go on to provide a deeper explanation of how he envisages these “various sufficient conditions” to work in his 1987 *Intention, Plans, and Practical Reason*. Finkelstein (2005: 588) summarises the solution offered up in this book as follows: “Only those effects that lie within the “motivational potential” of one’s action should be thought of as done intentionally, meaning that a person must have consciously adverted to and actually deliberated on an effect for it to count as something done intentionally.” This allows Bratman to overcome the criticism that he includes “too much” under the rubric of intentional action.

However, the solution to the first criticism does not help in resolving the second – that is, that Bratman includes too little in his account of intention. There are two possible ways in which this criticism can play out: either some actions we would usually call intentional are ruled out by Bratman’s account, or there is no obvious reason for thinking that intention plays a role in intentional action at all.<sup>59</sup> Resolving these two difficulties requires us to remember a part of Bratman’s conception of the content of intention as mental entity, namely that such content

---

<sup>59</sup> Finkelstein raises yet a further problem about the implications for the relationship between intention and reasons, and intention and responsibility. I will deal with her criticism in this regard in Section 5.

must form part of a plan in such a way that it fulfils the requirements of means-end coherence and intention-belief consistency. In other words, for an action to be part of the motivational potential of an intention-for-the-future, that action must be considered during the deliberation so as to ensure that the plan – of which the intention-for-the-future is a part – meets the requirements of coherence and consistency. Indeed, another way of looking at this is to say that these deliberative processes of coherence and consistency are part of intention-for-the-future, and that the action is part of the deliberation. That is to say, the action is intentional in that it “partakes” in the content of the intention. Although this interpretation is effective in staving off one aspect of the “too little” criticism by extending intentionality to actions that are part of the deliberative process (even as implicit assumptions), there is a second aspect of this criticism that has proven more difficult to refute. At issue here is whether this account of intention indicates to us *why* intentional action should be considered related to the intention-for-the-future. Since intention is part of a plan, and any deliberation that takes place with regard to such intention is rationally subordinated to the deliberations about the plan entire, then surely the intentional action need not have a direct connection to any particular intention-for-the-future, but merely to the plan which stands in a meta-position to both the relevant uses of intention? In other words, Bratman provides no evidence other than our initial intuition for why intentional action necessarily requires intention-for-the-future when the concept of plans is employed. I deal with this problem more fully when I discuss Bratman’s understanding of plans in Section 3. For now, it is worth noting that I consider this to be a serious problem for his account of intention.

### *1.3 The problem of plural intentions*

It must be said, however, that very element of Bratman’s account of intention that leads to the above-mentioned difficulties is also the reason he is able to handle plural intentions in a vastly superior way than either Anscombe or Davidson. Since an intention is to be understood as an element in a plan, it opens the possibility that a given plan can contain multiple intentions. The relationship of these intentions to each other will be governed by a set of regularities akin to those that are present in the intentions themselves. So a given intention must be internally coherent and consistent, and in addition, my various intentions must be coherent and consistent with the other intentions within my plan or risk irrationality. The only quality not present in the regularities between intentions in a plan, as opposed to intentions themselves, is that only the intentions themselves *must* have volitional force. Intentions are pro-attitudes

whereas plans are not. Although I might hold an intention as part of a greater plan, it is the intention that motivates to action, not the overarching plan. It is exactly this that makes intentions necessary in a plan.

A problem with this attempt to explain the relationship between intentions is that Bratman finds it necessary to introduce talk of “settled objectives” (Bratman, 2009a: 18-19).<sup>60</sup> He does this in order to resolve what he sees as a continuing threat of rational contradiction in cases like the gamer example. On this view, in cases where I have a plan that seemingly entails two intentions that are mutually exclusive, such as “I intend to hit T1 and T2” where I cannot achieve both, each individual element should be considered to be a “settled objective” rather than an intention (Bratman, 2009a: 19). So, I can say that my plan entails the general intention-for-the-future “I intend to hit at least one target,” or even “I intend to hit T1 or T2,” but this overall intention entails that I have a settled objective to hit T1 and a settled objective to hit T2. Since settled objectives do not have the same rational requirements that intention does, the fact that these objectives are mutually exclusive is not a concern. Also, any subplans I might form to achieve these objectives would still count as intentional. For example: “I intend to press the fire button to hit target T1.” By distinguishing settled objectives from overall intention in this way, Bratman hopes to avoid the problem of rational contradiction.

However, I consider “settled objectives” to be an unnecessary addition to his account of intention. It hinges on the creation of a placeholder concept, “settled objective,” that seems to have no content of its own and is introduced purely to make a perceived contradiction go away. I am reluctant to accept this line of reasoning precisely because in our everyday talk I would not refer to my “settled objective” to hit T1 or T2; rather, I would refer to my *intention* to hit T1 or T2.

In conclusion, Bratman’s attempt to unite the three seemingly irreconcilable uses can be seen as a direct improvement over Davidson. His most important contribution is to provide a means to finally unite intentional action and intention-for-the-future, a project that has thus far stymied other theorists of intention. In addition, his introduction of the notion of plans allows us to deal with the previously neglected problem of the relationship between intentions in a meaningful way. Although there are still problems here, I take only two of them to be related to the unity of the three uses, namely that Bratman does not offer us a

---

<sup>60</sup> Bratman adopts this the notion of “settled objectives” from McCann (1991: 26).

completely convincing solution to the problem of mutually exclusive intentions, nor does he provide sufficient grounds for thinking that intentional action *must* be rationally related (or related by reason)<sup>61</sup> to intention-for-the-future. Finally, Bratman still does not provide us with a proper explanation of the relationship between intentions and plans. It will try to overcome these hurdles in my own account of the unity of the three uses in Chapter 5.

## 2. Epistemic requirements

When it comes to the epistemic requirements for intention, the first thing to note is that Bratman claims that the consistency of my actions and my beliefs is one of the critical requirements that differentiates intentions from other forms of volition, such as desire. His point is that when I am in the process of intention formation I must ensure that my intention is “*strongly consistent, relative to my beliefs*” (Bratman, 1987: 31) to avoid involving myself in “a criticizable form of irrationality” (Bratman, 1984: 383).<sup>62</sup> The consistency requirement is an epistemic requirement in that it concerns how an agent is supposed to incorporate her knowledge (in the form of her beliefs) into her intention. If there were no connection between an agent’s intention and what she knows (believes), there could also not be any link between my possession of a certain set of beliefs and the quality of my intentions.

### 2.1 The intention-belief consistency requirement

Bratman has phrased requirement for consistency between intention and beliefs in several ways: “*Intention Consistency*: The following is always pro tanto irrational: intending A and intending B, while believing that A and B are not compossible” (Bratman, 2009b: 413); alternatively, “it should be possible for my entire plan [or intention] to be successfully executed given my beliefs are true. This is the demand that my plan [or intention] be *strongly consistent relative to my beliefs*” (Bratman, 1987: 31). Both formulations are meant to convey that any intention must endeavour to be consistent with the agent’s beliefs. To intend what I believe is impossible is irrational, and part of what makes a given mental entity an intention is that it should strive to avoid such irrationality. Of course, this does not prevent me from having irrational intentions, but it indicates that any intention must strive to meet this

---

<sup>61</sup> This is of particular concern to Finkelstein, who is concerned that understanding the relationship between intention and intentional action as something other than a rational relationship undermines our understanding of moral responsibility. More on this in Section 5.

<sup>62</sup> As I have already said, there are requirements other than that of consistency relative to belief, most notably means-end coherence, which I discuss in Section 4.

criterion, even if it fails. This striving is an essential component of intention. Although Bratman claims that I do not have to have a belief to A in order to have an intention to A, this still means I must “check” my beliefs whenever an intention is formed. Or, more accurately, I would run the risk of being justifiably criticisable as irrational if I did not. There are certainly benefits to Bratman’s argument in this regard. It helps us to make sense of certain difficult cases, such as where I write a will with the intention of ensuring a better life for my children, but I do not believe that I will succeed, provided that I do not believe that success is impossible. Bratman’s argument also frees us from the charge cognitivism, as intentions cannot be the result of theoretical reasoning if they are not reducible to beliefs (Bratman, 2009a).

## 2.2 Epistemic opacity and irrational intentions

Recall that Davidson, because he has described intention as being an all-things-considered (and later an unconditional) judgement, is forced to maintain that it would be impossible for me to intend against my rational judgements. When I intend it must always be in line with my rational judgement, as that is what my intention is. I argued in Chapter 3 that this assumption causes difficulties for Davidson’s theory, as it hinges on how opaque or transparent my intention needs to be to me. This same issue of transparency crops up in Bratman’s account.

It is obvious that Bratman thinks that our intentions are much more transparent to us than Davidson is willing to concede. This is because, unlike Davidson, Bratman requires that my intentions must be consistent with my beliefs, which *are* transparent to me (at least on reflection). If this is the case, and given that Bratman still wishes to retain some link between intention and rational judgements, then how are we to make sense of those instances of pathological behaviour and *akrasia* already mentioned in relation to Davidson? The answer is twofold and surprisingly simple: Firstly, Bratman denies that intention is a rational judgement, and argues instead that intention only contains a *striving* toward rational coherence and consistency. Secondly, he asserts that his account does not claim that any intention to A must coincide with a belief that I will A, merely that the intention not be *inconsistent*. In other words, I can have an intention to A, even in the absence of a belief that I will A, provided that I do not believe that I will *not* A. This last possibility would render me inconsistent, as I would be intending what I believe I will not do. Another way of saying this

is that, although a belief to the contrary would be unacceptable, intention in the face of agnosticism is perfectly acceptable (Bratman, 1987: 32-34).

The first part of his reply allows us to recognise that an action can be irrational and still intentional. The measure for the attribution of intentionality to an action would depend on whether or not the agent could, if asked, answer the question of “Why?” that particular action was undertaken and the explanation is subject to the rational regularities of means-end coherence and intention-belief consistency.<sup>63</sup> As is undoubtedly apparent, much of this has to do with practical reasoning, and for that reason we encounter this matter again in Section 4. As far as the epistemic requirements for intention is concerned, we can say that the move from intention as a rational judgement to intention as open to rational regularities reflects an increase in for the amount of epistemic opacity allowed. While we have seen that Davidson also increases the amount of opacity allowed in order to deal with the problems of *akrasia* and pathological behaviour, his approach is flawed. The opacity that Bratman includes is of a different kind. For Davidson, what is opaque is the submerged “real” rational judgement that has led to my intention, whereas for Bratman what are opaque are merely aspects of the practical reasoning process. If it were a requirement for having an intention that it *had* to achieve coherence and consistency (in the sense that Bratman uses these terms), then I would have to be crystal clear about every aspect of my intention. By changing this requirement to *striving* for as opposed to *achieving* coherence and consistency, it becomes possible that aspects of my decision-making process are not clear to me. This opacity when present<sup>64</sup> will likely lead to irrationality, but it will not prevent the presence of intention.

In the case of *akrasia*, we can describe this as a case of an agent failing to intend rationally, but certainly intending nonetheless. The same can be said of the cases of pathological behaviour, although I hesitate to simply resolve the entire issue in this way – particularly where the relationship between intention and responsibility is concerned. For example: although it might be fine to say that the slacker who avoids writing his thesis due to *akrasia* is

---

<sup>63</sup> Crucially this does not mean that the explanation will *be* rational; it means that rationality is an appropriate criterion for judging such an explanation. For example: I sit down on a thumbtack and involuntarily leap up. In leaping up I hit my head on a light fitting. When I am asked “why” I did this, my explanation would be a wholly causal one. Judging the rationality of the action would be inappropriate (for the reason that no reasoning in fact took place). This is opposed to a situation where I, fearing a thumbtack on my chair, leap up to avoid it and then hit my head like in the previous case. In this second case, my explanation would have to include references to reasons, and judgements of rationality would be appropriate.

<sup>64</sup> Needless to say that just because the possibility of this opacity is recognised does not mean that Bratman’s account claims that all, or even most, instances of intention involve such opacity. Often it will (may) not.

responsible for the consequences of this intention, it is often the case that we hold those who act out some kind of pathology responsible in a different way. After all, it would seem contrary to our everyday understanding for me to hold Lady A, from the Galgut example, responsible for breaking her leg in the same way that I would hold the slacker responsible for failing to write his thesis. We normally think that the fact that the action was pathological somehow mitigates responsibility. I will discuss this in more detail later on. What we can say at this point is that Bratman's treatment of epistemic opacity is convincing in so far as it solves the problem of irrational intentions.

As Bratman doesn't pay much attention to the epistemic requirements for intention, and since responses to his account deal with the matter almost wholly tangentially, I leave this discussion here for the moment. I will state that Bratman provides us with a clearly superior solution to the issue of opacity to that of Anscombe or Davidson. As far as the epistemic requirements for intention are concerned, my own account of intention is closer to Bratman's than to either of the other theorists I have discussed here, although my argument for the necessary link between intention and belief follows Anscombe rather than Bratman. I take up the epistemic requirements for intention again in the next chapter, where I will endeavour to overcome Bratman's arguments against belief being a necessary component of intention.

### **3. The relationship between motivations, causes, reasons and intention**

This section deals with Bratman's understanding of the relationship between intention on the one hand, and motivation, causes and reasons on the other. I will show that Bratman considers intention to be a type of pro-attitude, as well as an effective cause for intentional action. On the other hand, he argues strongly against the view that intention should be counted as a reason for action. I will examine each of these relationships in turn. As far as motivation is concerned, I find Bratman's argument convincing. With regard to the relationship between intention and causes, while I agree with Bratman's conclusion here, his failure to resolve the problem of causal deviance prevents his arguments from being convincing. In a similar vein, I agree with Bratman's claim that intention should not be taken as a reason for action, but find his arguments in support of this view to be unconvincing.

### *3.1 Intention as a conduct-controlling pro-attitude*

Let us start off by looking at Bratman's treatment of the relationship between motivations and intentions. For Bratman, an intention certainly represents a motivation; to be precise, he refers to intentions as "conduct-controlling pro-attitudes" (Bratman, 1987: 20). However, this is not meant to equate intention with desire. While that intention involves desire, even predominant desire is insufficient for intention, since it need not involve a commitment to act. In addition, unlike desires, intentions have an assortment of commitments to coherence and consistency that our practical reason compels us to fulfil (ibid.: 32). In other words, Bratman, like Davidson, considers intention as representing a particular kind of pro-attitude. Or, as I would say, intention includes a pro-attitude in its composition, together with a commitment to certain rational regularities, as discussed in Section 2 above.

An interesting contrast can be drawn here with Anscombe's argument that it is possible to have an intention without a dominant desire or in fact any desire at all. Both Davidson and Bratman have advanced the idea that intention necessarily includes a motivation, so the two camps appear to be at loggerheads. However, Anscombe's arguments in this regard are extremely problematic, and her line of reasoning taken to its fullest conclusion results in unacceptable outcomes (as I have demonstrated in Chapter 2, Section 3). As such, it seems reasonable to assume that the movement towards understanding intention as a pro-attitude is a movement in the right direction.

### *3.2 Intention and causes*

On the matter of the relationship between intentions and causes, Bratman, again in line with Davidson, treats intention as an effective cause, even though explanation via intention is a rational explanation. This works as follows: say that I have the intention to kick a ball in order to strike a can. I then intentionally kick the ball striking the can. If I am asked the question, "Why did you kick the ball?", my answer would take the form of a rational explanation as I unpack my intention. By this stage this should seem very familiar. On the other hand, if we examine the situation in terms of causality, then we can equally say that my intention to kick the ball was the cause of me kicking the ball, recognising also that that intention was the effect of its own cause, namely my preceding reasons and pro-attitude for kicking the ball. What is characteristic of intentional action is not that it cannot be explained

by causality, but that it can *also* be explained by reasons, whereas unintentional actions cannot be explained in terms of reasons.

As discussed before, there is one significant objection to viewing intentions as causes: the difficulty of causal deviance. To recall, causal deviance refers to the difficulty in identifying what is meant by “with an” in the case of a statement such as “I did X intentionally because I did it *with an* intention-for-the-future to X (or Y).”<sup>65</sup> The problem is that there are clear cases where I could act because of an intention-for-the-future, but where we would not normally attribute intentional status to that action. An easy example is to imagine that I am intending to blink, perhaps to signal somebody, but my intention causes me distress such that it *causes* me to blink involuntarily. In this case, I have blinked because of my intention to blink, but somehow not in the right way for this to be considered an intentional action. Bratman does not attempt a solution of this problem, which is a weakness in his account of intention.

### *3.3 Intention as not counting as a reason for action*

Bratman certainly thinks that intention-for-the-future is concerned with reasons, even though he does not follow Davidson in defining intention-for-the-future as a rational judgement. By introducing the notion of plans, Bratman is able to differentiate the rational judgement that constitutes the conclusion to a given practical reasoning process from the intention to act on that judgement. In other words, my practical reasoning can yield a judgement in favour of a certain plan, the fulfilment of which requires that I hold a series of intentions. Each intention is a pro-attitude that commits me to act, and which must strive to be intention-belief consistent and means-end coherent with my overarching plan. In this way, the intention can be understood as a candidate for rational evaluation, but is not as a rational judgement itself.

I have already discussed what I take to be the benefits of this view, particularly the fact that it helps to resolve difficulties relating to the epistemic requirements for intention. That being

---

<sup>65</sup> Another possible formulation is “I do X intentionally when my doing X was caused by my intention to X (or Y).” This formulation has the benefit of clarifying the role played by causation in this dilemma. I have chosen not to use it as my primary formulation because it is a less in line with the way we normally talk about intentions and intentional action. If asked, I will usually say “I did X (intentional action) due to (or because of) my intention to X (or Y),” rather than “my doing X (intentional action) was caused by my intention to X (or Y).” If I were to hazard a guess as to the reason for this everyday preference, I would suggest that the second formulation places me as the agent in too much of a passive role. In effect, the second formulation feels like reported speech, and it is probably in the report of the actions of another that I might employ this formulation. Ultimately, these two statements carry the same meaning: that intentional action is causally linked to intention-for-the-future.

said, there are certain weaknesses in Bratman's approach— specifically, his understanding of the role played by reasons in relating intention-for-the-future and intentional action is untenable. For Bratman, an intention to X does not generally count as a reason for the necessary means to X, which means that it does not count as a reason for intending the means to X (Bratman, 2009b: 416). This view is motivated by his desire to avoid two problems: unacceptable bootstrapping in reasoning and the issue of having an immoral intentions count as a reasons for intending an immoral means to that intention.

We have encountered the first problem in Chapter 3, where we saw that it involves the unacceptable circularity in reasoning that results from understanding intention as necessarily being a reason for action. This circularity results from the ability of an intention – when counting as a reason for action – to count as a reason for itself. Thus, suppose I have an intention to X, but my intention is *akratic* – that is to say, I intend against the balance of my reasons. My practical reasoning makes me aware that on the balance of reasons it would be preferable to intend Y. If my intending X counts as a reason for me to intend the means to X, then it is possible that my intending X could overturn the balance of reasons in favour of intending X over intending Y, even though I am aware that intending X is *akratic*. This is clearly unacceptable, as it amounts to bootstrapping a given intention into being the most reasonable conclusion on the balance of reasons, purely because we hold this intention. This not only allows even irrational intentions to count as legitimate reasons for future reasoning, but also allows that which is irrational to do before my intention is formed to become rational after the fact. Let me try to illustrate the problem through an analogy: If I (honestly) make a prediction X, then I must have reasons to believe that my prediction is correct. Bootstrapping would occur, in this case, if we allowed this prediction to count as a necessary reason for holding this belief (that my prediction is correct). This would mean that simply making the prediction would count as a reason to believe that the prediction is correct. This means that a prediction which my reasoning tells me is irrational can become rational simply by making the prediction.

The second problem originates with the fact that it is possible for an agent to intend immoral ends. If we combine this with the assumption that “accepting something as a normative reason means accepting its justifying role in practical reasoning,” then, in so far as an immoral intention-for-the-future gives reasons for an immoral intentional act, we are forced to concede that such an action might be justifiably rational. If we assume that intentions

provide reasons<sup>66</sup> for the means to realising those intention's ends, we must conclude that, in this case, it would be reasonable to intend those means. Bratman (2009b: 416) uses the example of ethnic cleansing. Say that I had the intention-for-the-future to perform ethnic cleansing. If my intentions provide reasons for intending the means to those intentions, then I am rationally justified in intending those means (or, by extension, taking intentional actions with my intention-for-the-future as a reason). Thus, unless we are willing to commit to the view that of ethnic cleansing is rationally justified, we have to concede that an intention to X does not generally count as a reason to pursue the necessary means to X.

Both the problem of bootstrapping and the problem of immoral ends impact not only on the relation between intentions and reasons, but also on issues of practical reason and responsibility. I will devote the rest of this section to Bratman's solution to the bootstrapping problem and leave the discussion of the relation between intention and responsibility for Section 5.

Bratman defends this conclusion by arguing that we can and should divorce intention from *necessarily* providing reasons to intend the means to that intention. In other words, an intention-for-the-future is *not necessarily* a *reason* for intentional action. This means that in some cases I might act intentionally, but without any intention-for-the-future providing the reason for my intentional action. In making this argument, Bratman agrees with Broome (2001) that the usual argument for deriving reasons from an intention is flawed. The argument in question runs as follows:

- (1) I have a decisive reason to avoid intending X and not intending the means to X, M
  - (2) I intend X
- Therefore
- (3) I must have a decisive reason to intend M

It is worth noting that intending M in this example can also be taken to mean intentionally performing M. The reformulation would look like this:

---

<sup>66</sup> In the work where he spells out his position concerning this particular aspect of the argument (*Intention, Practical Rationality and Self-Governance*, 2009b) Bratman is careful to refer to "normative" reasons for action, as opposed to reasons understood more generally. There is a purpose to this choice, since what we are discussing are reasons that play a role in determining actions, and are thus normative in the sense of guiding or conduct-controlling. Another way in which they are normative is that failure to abide by them would be a cause for criticisable irrationality, so in a sense one is accountable for failing to act on these reasons, even if this accountability is not moral in nature. Although Bratman's reasons for prefacing "reasons" with the qualifier, "normative," I think this is unnecessary for the purpose of the present argument. The reasons at issue here are normative by definition, so it is unlikely that there would be any ambiguity on this score.

(1) I have a reason to avoid intending X and not intentionally performing the means to X,  
M

(2) I intend X

Therefore

(3) I must have a reason to intentionally perform M

Bratman, following Broome (2001),<sup>67</sup> thinks that both versions of the argument are flawed in that (1) and (2) do not necessitate (3). This is so because even in a situation where I hold (2), and become aware of (1), I can still choose either to revise (2) or commit to (3). This means that simply possessing (1) and (2) does not make (3) necessary, since I can freely forsake (2) instead. This position allows Bratman to retain a commitment to means-end coherence, as represented in (1), while still claiming that intentions do not necessarily count as reasons and thereby avoiding bootstrapping. Nevertheless, while this strategy may seem to defeat the problem of bootstrapping, it is not entirely successful. However, to explain why this is so, we first need a clearer grasp of Bratman's views on practical reason, including the criterion of means-end coherence. This is the topic of Section 5, and I will therefore present my criticisms of Bratman's strategy as part of that discussion. Before I can do that, however, we first need a clearer grasp Bratman's view of the relation between intention and practical reasoning.

#### **4. The relationship between intention and practical reasoning**

As I have mentioned before, the conventional distinction between theoretical reasoning and practical reasoning is that the former is concerned with the reasons for belief, whereas the latter is concerned with the reasons for action. Furthermore, within practical reasoning, the object of such reasoning is taken to be an individual action. I might ask, for example, after my reasons for typing this sentence. Bratman, however, adopts Gauthier's view that the object of practical reasoning is "deliberative procedures" as opposed to "individual actions"

---

<sup>67</sup> Despite their agreeing on this particular topic, the common ground between Bratman and Broome does not extend much beyond this point. Broome's own attempt to explain the intricacies of the relation between intention and reasons relies heavily on the distinction between "normative requirements" and reasons (Broome, 2001: 8). It is Broome's contention that intentions count as normative requirements, but not as reasons, where normative requirements (as the name suggests) means that an agent is normatively compelled toward the intended end, and that this intention also results in a normative requirement toward the means to that end. Yet at no time does the intention count as a *reason* toward either the end or the means. I do not develop an argument against Broome's views in this regard within this thesis. What I will state provisionally, however, is that the addition of the normative requirement seems to me to be a case of adding a concept purely to fill an explanatory gap. If at all possible I think it best to pursue an account of intention that does not need to make this distinction between reasons and normative requirements.

(Bratman, 2013: 656).<sup>68</sup> This fits in nicely with Bratman’s commitment to the idea of plans as the primary means by which an agent controls and directs her conduct. A “deliberative procedure” is a standard of action – that is to say, a conduct-controlling regularity that can be potentially applied to multiple actions. The idea is that if these procedures can be shown to be rational, then their application in particular instances of action would also be rational. By extension, “a specific choice or action is rational if and only if it is adequately supported by rational deliberative procedures” (Bratman, 2013: 658).

With this approach to practical reason in mind, I set out my analysis in the follow way: I begin by analysing the role that Bratman ascribes to means-end coherence. Building on the results of this analysis, I argue against Bratman’s employment of the notion of self-governance as a way to salvage the normative role of means-end coherence, and finally, I analyse Bratman’s three criticisms of cognitivism.

#### *4.1 The role of means-end coherence in Bratman’s account of intention*

Bratman introduces the notion of deliberative procedures as part of his explanation of the relationship between intention and practical reasoning. His main concern in this regard is to work out what is required for an intention to be rational. As we have seen, he considers the most important such requirements to be: intention-belief consistency and means-end coherence. Having already discussed intention-belief consistency in Section 2 above, let us now try to make sense of the notion of means-end coherence. Bratman describes the requirement as follows:

Although plans are typically partial, they still must be appropriately filled in as time goes by. My plans need to be filled in with subplans concerning means, preliminary steps, and relatively specific course of action, subplans at least as extensive as I believe are now required to do what I plan...Failure to fill in my plans as needed in these ways will leave them means-end *incoherent*. (Bratman, 1987: 31)

Although Bratman refers to “plans” here, the same coherence requirement can be taken to apply to intentions individually. One of the most crucial aspects of means-end coherence is

---

<sup>68</sup> There are other elements that he adopts as well, though these will not be discussed in this thesis as they do not influence the relationship between *intention* and practical reasoning. For an example of how Bratman understands Gauthier’s position see his 2013 article *The Interplay of Reason and Intention*.

“*Instrumental Rationality* requirement” (Bratman, 2009a: 13). Bratman describes this requirement as follows:

Suppose I intend end E, believe that a necessary means to E is M, and believe that M requires that I intend M. My attitudes concerning E and M engage a basic requirement of practical rationality, a requirement that, barring a change in my cited beliefs, I either intend M or give up intending E. (ibid.)

Most recently, he describes the instrumental rationality requirement of means-end coherence as follows:

*Means-End Coherence*: The following is always pro tanto<sup>69</sup> irrational: intending E while believing that a necessary means to E is M and that M requires that one now intend M, and yet not now intending M. (Bratman, 2009b: 413)

Thus, means-end coherence can be considered a deliberative procedure that provides a standard against which the rationality of individual plans or actions can be measured. Yet why does Bratman advocate means-end coherence as a requirement of intention if a given intention-for-the-future cannot be a reason for a given intentional action? Since the whole point of introducing the criterion of coherence is to show that having an intended end places us under rational pressure to intend the appropriate means, then, if the intended end does not give us *reason* to intend a given means, what would be the point of insisting on means-end coherence in the first place? However, this is a flawed criticism. Holding an intention-for-the-future is not necessarily a reason for intentional action in so far, and only in so far as, the intention-for-future can be revised. If an agent were to refuse to revise her intention-for-the-future for any reason, she would then be rationally required to adhere to means-ends coherence, or be guilty of criticisable irrationality. In other words, although the requirement of means-end coherence may not mean that intending X is a reason to intentionally perform the means to X, if I decide to X while knowing that to X requires intentionally Y, then I am rationally required to Y. In the example of ethnic cleansing: I do not have a reason to intentionally perform the means to ethnic cleansing simply because I have an intention-for-the-future to do so. I could instead choose to cease intending thus. However, if, for whatever

---

<sup>69</sup> Bratman emphasises the pro tanto nature of this requirement because he does not wish to rule out the possibility that there may be instances where this requirement may be overturned by some more significant requirement. He does not discuss what such a requirement might be.

reason, I refuse to revise my intention-for-the-future, then I am rationally required to conform to the appropriate means.

Yet this is not the end of the story. What about a situation where an individual's intention is such that for some reason it cannot be revised? This is a difficulty raised by Setiya (2007), and one that Bratman tries to address. What if, for example, there is a smoker who is addicted, and so has an intention to smoke that cannot be revised – or that is not “psychologically modifiable?” (Bratman, 2009b: 426).<sup>70</sup> In cases such as this, Bratman (ibid.) thinks it unavoidable that we must agree that:

(1) I have a decisive reason to *avoid* holding the combination: (i) intending to smoke and (ii) not intending to intentionally smoking, where intentionally smoking is a necessary means to fulfilling my intention to smoke.

(2) I intend to smoke, and this intention is not modifiable

Therefore

(3) I must have a reason to intentionally smoke

However, this leaves us with the original problem that it may be possible for an individual to have the intention-for-the-future to commit ethnic cleansing, albeit only in the more restricted case where this intention is psychologically unmodifiable. Since Bratman is unwilling to concede that such an intention might be rational, he needs to change his angle of attack. He does this by asking what we mean when we say that there is a decisive reason to conform to means-end coherence. His answer involves the introduction of the principle of “*Transmission Reasons*” (Bratman, 2009b: 424). This is the principle that, if I have a reason R in favour of X, and the means Y is necessary for X, then R is a reason for X *and* a reason for Y. Quite obviously the previous example adheres to this principle if X is taken to be means-end coherence and Y is taken to be: intentionally smoke. Since in this case my intention-for-the-future is not modifiable, it follows that intentionally smoking is a necessary element in achieving means-end coherence (X), and so I must intend thus. In light of this explanation, Bratman then focuses on the last unaccounted-for variable in *Transmission Reasons*, namely reason R. In this case, R is what gives means-end coherence its normative force. Bratman (2009b: 430) claims that the reason why our intention-for-the-future and intentional action should adhere to the requirement of means-end coherence is our commitment to self-governance. More accurately, he argues that “we have an intrinsic reason to govern our own

---

<sup>70</sup> By which he means: “an intention [that] is not modifiable by the agent because of some such underlying psychological incapacity” (Bratman, 2009b: 423).

lives” (ibid.: 432), and that means-end coherence is a necessary means for us to effectively self-govern. The result of this is a case of *Transmission Reasons* where “an intrinsic reason to govern our own lives” is R, self-governance is X and obedience to means-end coherence is Y. So I adhere to means-end coherence because I have an intrinsic reason to govern my own life, which is also the reason for my commitment to self-governance.

How does this argument aid Bratman in disproving that an intention-for-the-future provides a reason for an intentional action in a case like this? It does this by demonstrating that the requirement for means-end coherence – which is represented by proposition (1) – does not always apply. It only applies in cases where self-governance is possible. That is, the deliberative procedure of means-end coherence is only rational provided that self-governance is assumed. Obviously, if my intentions are not at all open to revision then I am quite definitively incapable of self-governance. This is Bratman’s solution to the problem of a psychologically unmodifiable intention to commit ethnic cleansing counting as a reason for intending the immoral means to such an act.

However, he does recognise one set of cases where he thinks that it is plausible to say that the intention-for-the-future provides a reason for intentional action. The set of cases in question are ones where the agent’s intentions are “volitional necessities” (Bratman, 2009b: 439). These cases are characterised by an agent having an intention that they recognise they cannot revise, but rather than this indicating an inability to self-govern, this very inability to revise the intention is taken to be a crucial aspect of self-governance. An example can be something like our commitment to physical integrity. Imagine an agent who has an intention to preserve her physical integrity. This intention is not revisable, and the agent is aware of this unrevisability and endorses it wholeheartedly as a necessary means to self-governance. In this case, she is capable of self-governance, even though she cannot revise her intention. This brings us to Bratman’s final conclusion on the matter: intention-for-the-future does not *generally* provide a reason for intentional action. Instances of intentions that cannot be modified do not disprove this position, as in such cases self-governance is impossible, and when self-governance is impossible then there is no requirement for means-end coherence. However, in cases where the agent has an intention that is a volitional necessity, holding that intention does provide a reason for intentional action.

## 4.2 Critique of Bratman's notion of self-governance

In terms of the above argument, the reason for why we should adhere to means-end coherence is precisely our commitment to self-governance. However, does the concept of self-governance do the work that Bratman requires of it? Recall that the argument involving self-governance is part of Bratman's overall attempt to prove that means-end coherence is a rational requirement for intention, while at the same time maintaining that, generally speaking, intention-for-the-future does not provide an agent with a reason to act so as to realise the necessary means for achieving that intention-for-the-future. For Bratman's argument to be convincing, the notion of self-governance must be made to do a lot of work. It is presumed to act as a *motivation* – as in: my inclination to self-governance motivates me to achieve the means of self-governance – and as a *reason* for adherence to means-end coherence as a deliberative procedure. Let us first look at the latter aspect. Given the principle of *Transmission Reasons*, the following must hold:

- (1) I have a reason to avoid holding the combination of wanting to be self-governing and not wanting to adhere to the necessary means to be self-governing (which includes means-end coherence)
  - (2) I want to be self-governing
- Therefore
- (3) I must have a reason to adhere to means-end coherence

If this does not hold then I would have no reason to adhere to means-end coherence simply because I want to be self-governing. The deliberative procedure would be defeasible. Obviously, this is the re-application of the same argument that Bratman originally used to show that an intention does not necessarily yield a reason for intending its means. As such, it should be no surprise that it leads to a similar conclusion. Simply put: I can choose not to want to self-govern. But this is problematic. Surely *choosing* not to self-govern *is* to exercise self-governance? Further, one could choose to selectively remove an aspect of oneself from the sphere of self-governance. For example: I can choose to allow myself to become addicted to a powerful narcotic. Or I can choose to give myself over to the control of another in order to protect my physical integrity. Does this work? It is difficult to say. It seems we would need to know more about what it means to call an agent self-governing. At any rate, it is difficult to see how Bratman's notion of volitional necessity would be different from choosing addiction. In that case, the addict would have no reason to adhere to means-end coherence, meaning that the very basis on which we would normally criticise an addict for irrational

behaviour (that is, failure to adhere to the deliberative procedure of means-end coherence, where the end in question is effective self-governance) is no longer available.

A related problem is that intentions that are volitionally necessary would seem, to me at least, to be better interpreted as cases where the agent simply rejects the inclination towards self-governance, due to some compelling reason. That is: I do not revise my intentions (i.e. employ my ability to self-govern) to maintain physical integrity, because I have a reason to maintain such integrity that is more compelling than my inclination to self-govern. If this is the case, then we can dispense with treating intentions that are volitionally necessary as being a discrete set of intentions, unless what renders them discrete is understood as the presence of particularly compelling reasons not to exercise self-government. Alternatively, one can say that Bratman, in positing volitional necessity, is mistaken in claiming that such intentions are not revisable. Rather, in cases of volitionally necessary intentions such as physical integrity, what we are really talking about is the role of a non-rational element in the practical reasoning process. In cases such as these I am either unable to self-govern and am “controlled” by this non-rational element, or I retain self-governance but adhere to this non-rational element nonetheless. In the former case I would argue that intention is not present, and in the latter that the intention is in fact revisable. I discuss this in greater depth in Chapter 5, Sections 3 and 4.

I now turn to the matter of self-governance as a motivation. The relevant question to ask here is: *Why* do we want to self-govern? Bratman simply calls the drive to self-government “intrinsic” (Bratman, 2009b: 432), and leaves it as that.<sup>71</sup> However, this is not the only way to talk about self-governance. As I have shown, we can potentially think of individual cases where an agent does not self-govern, or chooses to opt out of self-governance with regard to a particular action or set of actions.

Imagine the following fictional situation: I am told that I can choose to hand myself over to some scientist, who will hook up my brain to a series of computers. These computers possess a near omniscient problem-solving ability. If I choose to commit to this scenario, I will no longer be able to govern my own actions; the computers will direct me in an effective

---

<sup>71</sup> Gauthier is committed to a similar view; he takes “the aim that one’s life go as well as possible” (Bratman, 2013: 658) to be the bedrock of any attempt at practical reasoning. The difference between this and Bratman’s self-governance has to do with how much subjectivity it to be allowed. Bratman considers there to be objective markers for self-governance, whereas Gauthier’s “as well as possible” is very subjective.

manner. The most crucial thing to imagine here is that these computers are able to predict my desires such that they will be completely sated. Is agreeing to this existence irrational? Now in real life we might answer yes, but only because there is no assurance that the computers will necessarily succeed. However, in this fictional scenario we can assume the computers to be infallible. Note also that my committing to this scenario is no different from committing to so-called volitionally necessary intentions, and as such Bratman cannot argue that my actions when under the computer's control are unintentional. After all, if I were to have an intention-for-the-future that is volitionally necessary, and fulfilling this intention-for-the-future would predetermine all my actions, then Bratman would still permit these to be considered intentional actions. In my example, I choose to commit to the following intention-for-the-future: "I intend to allow the computers to control my actions," which then dictates all the actions to come. This does not seem to be necessarily irrational. Why would I be motivated to choose against this course? Why *necessarily* choose to self-govern? Most crucially, however, if I can be said to act intentionally in a case where I choose to have computers direct my actions, this would bear scant resemblance to our everyday use of the term "intentional action."

Faced with these complexities, I suggest that it is a mistake to appeal to self-governance as proof of the normative force of means-end coherence, or as a solution to the problems generated by the claim that intention-for-the-future is not a reason for intentional action. Although I concede that a fuller explanation of what is meant by self-governance might yet be able to resolve these complexities, in the absence of such an explanation I suggest that there is a simpler and more convincing alternative. This is the idea that an intention-for-the-future is not *itself* reason, but is *backed* by reasons. On this view, the reason for an intentional action can be found in these background reasons, rather than in the intention-for-the-future itself. Means-end coherence retains its normative force, because these background reasons must adhere to this requirement. The only difference is that it is my reasons for intention, not my intention itself, that must be rational justifiable. I will be arguing for such an alternative in Chapter 5, Section 3. In the rest of this section, I will examine one of the more common approaches that has been put forward as an alternative to Bratman's theory – namely cognitivism – together with Bratman's criticisms of this alternative.

### 4.3 Analysis of Bratman's critique of cognitivism

An alternative means to resolve the issue of how intention relates to the requirements of practical reason in the form of means-end coherence can be found in the work of the so-called cognitivists. A cognitivist account would argue that the requirement of means-end coherence is not in fact an aspect of practical reasoning, but of theoretical reasoning. That is to say, it concerns belief rather than action. Although this approach still recognises means-end coherence as a requirement of intention, it does not commit us to the idea that if I intend X then I must, to be coherent, have a practical reason to intend the means to X. Rather, it says that if I intend to X and I believe that to accomplish X I must intend the means to X, which is M, but do not intend M, then I have incoherent *beliefs*. What we can note from this is that reasons for action play no role here, only reasons for belief. This circumvents the entire issue of whether or not intention-for-the-future provides a reason for intentional action.

Bratman's criticises the cognitivist approach on three grounds: (i) it assumes a link between intention and belief that is too close to be plausible; (ii) the possibility of falsely believing one intends the means to an intention renders the approach untenable and (iii) by rejecting any role for practical reason it fails to account for the volitional quality of intentions. The first and second criticisms are, I think, weak. The third, on the other hand, is very convincing.

The first criticism recognises that, in order for the cognitivist position to hold, intention must be closely tied to belief. The cognitivist maintains that, if I have an intention to X, believe that to fulfil X I must intend the means (M), but do not intend M, then my beliefs are incoherent. Means-end incoherence is thus explained in terms of belief incoherence. This would in turn be a case of incoherence in theoretical reasoning, since it concerns reasons for beliefs, as opposed to reasons for actions. But this relationship between intention and belief can be brought into question, and this is exactly what Bratman does. Firstly, there might be circumstances under which an agent intends X, without believing that she will do X. For example: I am writing my will with the intention of providing for my children. Yet, at the very same time that I am writing my will, because I am nearly broke and close to death, I do not believe that I will succeed in fulfilling my intention. Hence I am intending that which I do not believe.

Now it is important to remember that Bratman doesn't want to prove that belief serves no role in intention. He argues, in fact, that I should not intend that which I believe to be impossible – that is, I must retain intention-belief consistency – if I am to avoid criticisable irrationality. Nevertheless, the intention-belief consistency requirement is considerably less stringent than the relationship between intention and belief that the cognitivist account requires. To quote another example:

Suppose [...] I intend to stop at the bookstore on the way home. Still, I know that I am forgetful; so I am not confident that I will stop – after all, once I get on my bicycle I do have a tendency just to pedal on home. About this case I am inclined to say: I intend to stop, but I do not believe I will stop (though I do not believe I will not stop). (Bratman, 2009a: 21-22)

However, I think that Bratman is mistaken here. It seems to me that this matter could be resolved by simply stating that the agent believes that he will *try* to stop at the bookstore. Note that I am not arguing that he has the *intention to try* and stop at the bookstore; rather, the term intention itself implies the trying, just as it implies the belief. The addition of *trying* means that the agent, in this case, can hold the following belief, “I believe that I will try to stop at the bookstore on my way home,” without further difficulty.

Let us now turn to Bratman's second criticism of cognitivism, namely the possibility of false belief. Bratman argues that there could be a case where I have an intention to X, believe that means M is necessary for X, yet fail to intend M. I am therefore means-end incoherent. However, at the same time, I could have the false belief that I am, in fact, intending M. In this case, my beliefs are not incoherent; they are false. Bratman's contention is that since in this case I am means-end incoherent, but my beliefs are not incoherent (though they are false), the requirement of means-end coherence cannot be reduced to a matter of beliefs (Bratman, 2009b: 428). There are two reasons why this argument does not hold. Firstly, Bratman's critique presumes that there is a state such that an agent can believe that they are intending the means to an intention X, M, without really intending M, and I think this is simply impossible. If M is taken to be the actual means to X, and not just what I believe to be the means to X, then if I believe I am intending M, then M must be what I am intending. Say, for example, that I falsely believe that Y (and not M) is the means to X. I then intend X, and so intend the means to X. However, due to my false belief, I do not intend M, but instead intend

Y. In this case I am intending Y precisely because I believe Y (and not M) to be the means to X. It is impossible not to intend M while believing that I *do* intend M, since in that case I would actually be intending Y, precisely because of my false belief in Y.

Secondly, and building on the previous point, I would argue that any case where means-end incoherence occurs, this will go hand-in-hand with my beliefs being incoherent, as opposed to only false. What is required is clarity on whether the incoherence at issue is internal (subjective) or external (objective). The internal measure of my incoherence (if any) is determined by whether or not I have the subjective ability to be aware of my incoherence. The external measure is determined by whether or not I am incoherent given the actual facts of the matter. It is my contention that Bratman's claim that means-end incoherence can occur separate from belief incoherence is the result of measuring the former by the external measure and the latter by the internal measure.

This is best explained with an example: Say that I wish to flag a car on a roadway. I therefore have the intention-for-the-future of flagging a car. Let us say that on this roadway the necessary means for this intention-for-the-future is to hold out my thumb. Now suppose that, due to a story that my friends have told me, I falsely believe that rudely gesturing at the passing cars is a necessary means to my end. I then intentionally do as I believe is necessary, and so fail to intend the appropriate means. By Bratman's reasoning, I have failed to adhere to means-end coherence, as this would require me to intentionally hold out my thumb (intend the necessary means to my end). He further claims that in this case my belief is false, but not necessarily incoherent, as it could still be coherent with my other existing beliefs. So there is a disjoint between my means-end incoherence and my coherent false belief. However, I think that this example is not as straightforward as all that. While it is obvious that my intended means *did not match* the necessary means to my end, this is not necessarily what means-end reasoning requires of us! To be means-end coherent might only require that I intend what I *believe* to be the means to my end, which in this case I did. In that case, I would only be means-end incoherent when I intend something other than what I *subjectively* believe to be the means to my end. What has occurred here is that Bratman has judged belief according to an internal standard (i.e. is my belief coherent with my existing *subjective* beliefs?) while measuring means-end coherence by an external standard (i.e. are the means I intend coherent with my end, given all the facts of the matter). We can bring the two in line simply by measuring both by the same measure.

We can refine this argument even further. Let us note that until the moment where my intended means actually fails, I do not know with absolute certainty whether or not I am means-end incoherent. This judgement can only be made with absolute certainty from an objective perspective. Now, let us presume that, as I stand at the side of the road gesturing rudely, the belief arises within me that, given my continuing failure, my foregoing belief about how to flag down cars was false. In fact, once I have the facts of the matter, and the beliefs that they entail, my false belief will be incoherent with my other beliefs. Yet I could not know this before I developed new beliefs or acquired new facts, so by the internal measure I *was* coherent. However, an omniscient observer would judge that my beliefs are incoherent with the facts. Thus from an objective standpoint, I was both incoherent in my beliefs and also means-end incoherent. On the other hand, from an internal (subjective) standpoint, both my beliefs *and* my means-end reasoning were coherent up until the point that my attempt failed. So, provided that they are both evaluated from the same standpoint, there cannot be a case where an agent is means-end incoherence, while also having beliefs that are false, but not incoherent.

While I find Bratman's criticism of the link between intention and belief unconvincing, I do think it is a strong argument that reasoning about intention cannot be reduced to theoretical reason. His argument here is straightforward: at the conclusion of theoretical reasoning I will have as an outcome some or other belief, and this belief is the answer to the question "What should I believe?" or perhaps "What is true?" (Bratman, 2009a: 17). By contrast, when I engage my practical reason, although beliefs might play a part, I am asking a different question, namely, "What should I do?" or "How should I act?" The outcome of this latter process will be a deliberative procedure, not some or other belief. These deliberative procedures then serve as the rational requirements which apply to, and guide, intention. Crucially, my practical reasoning could be flawed, or hijacked by non-rational elements, and so my deliberative procedure, and by extension my intention, need not be perfectly rational. What matters is that all intentions are open to rational requirements that stem from practical reasoning, even if this reasoning is only reconstructed in hindsight (recall the discussion on Davidson and practical reasoning in Chapter 3, Section 3). Although an intention may entail a belief (and I argue that it does), an intention cannot be reduced to *any* kind of belief. This is obvious. If intention could be so reduced, then any belief could be an intention. Bratman shows, convincingly, I think, that the motivating force of an intention comes from my taking the beliefs developed by theoretical reason, and using them to satisfy the requirements of

agency. It is worth noting that this view is compatible with the view that takes belief to be a necessary component of intention; just a certain kind of belief.

This rather lengthy discussion renders the following conclusions: Bratman's definition of practical reasoning is different from that of Anscombe and Davidson, and owes much to the work of Gauthier on this topic. The significance of his theory for our understanding of intention is that it shifts the judgement about the rationality of an individual intention away from the intention itself, to the deliberative procedures that function as (normative) requirements for intention. To avoid having to give up on the requirement for one kind of deliberative procedure, namely means-end coherence, while nevertheless rejecting the view that intentions-for-the-future might count as reasons for intentional actions, Bratman introduces the notion of self-governance as the reason for our commitment to means-end coherence. I have argued that this is an unwise move, and that falling back on self-governance causes a series of problems for his theory. As a result, he has not succeeded in explaining the rational relationship between intention-for-the-future and intentional action. Turning to the issue of cognitivism, I have explained and criticised Bratman's arguments against understanding intention as inherently linked to belief, while agreeing with him that practical reason should not be collapsed into theoretical reason in the manner that the cognitivists assert.

## **5. The relationship between intention and (moral) responsibility**

The discussions in Sections 3 and 4 have a bearing on our understanding of the relationship between intention and moral responsibility. As Claire Finkelstein (2005: 583) points out, for an agent to be seen as (morally) responsible for an action, this action must have been performed "*for a reason.*" Given Anscombe's landmark claim that "a person acts intentionally just in case he acts *for a reason,*" by extension, an agent is responsible "if (and only if) he did it intentionally" (Finkelstein, 2005: 584). As I have shown in the two previous sections, Bratman fails to explain the link between intention-for-the-future and intentional action, which makes it difficult to work out how his account would resolve issues of moral responsibility.

### 5.1 Intended consequences and ethical rationalism

Bratman does mention responsibility in two important ways. First, his notion of “motivational potential” allows us to see that moral responsibility does not only relate to the *intended* consequences of an action, but also to the *intentional* consequences. If I have an intention to X, and then I intentionally X, this action is intended. But if I intentionally X, and X is in the motivational potential of Y, then I *intentionally* X, but X is not *intended*. For example: I am intentionally running a marathon with the intention of finishing it. As I do this, one of the consequences is that I am wearing down my shoes. This wearing down of my shoes is not intended; it is not the aim of my intention-for-the-future. However it is intentional, as it is part of the motivational potential of my intention-for-the-future to complete the marathon. The difficulty with seeing the wearing down of my shoes as part of the motivational potential of finishing the marathon is that wearing down my shoes must have been considered in the practical deliberations that resulted in the intention to finish the marathon. This leaves the door open as to whether or not wearing down my shoes is to be thought of as having been done *for a reason*.<sup>72</sup> Bratman certainly does not wish us to conclude that it is my intention to finish the marathon that provides the reason for my wearing down my shoes (this was the whole point of his arguments in Section 4.1). The obvious answer for him is to appeal to means-end coherence, but this appeal brings the issue of self-governance into play. For the reasons given in Section 4.2, I do not see this as a viable strategy. Bratman’s argument thus leaves us unable to say whether or not I am, in fact, *responsible* for wearing down my shoes.

The other point where responsibility comes up in Bratman’s account is more indirect. As mentioned previously, he is concerned that if we allow intention-for-the-future to provide reasons for intentional action (say, through the medium of motivational potential being understood as imparting reasons to act) then we might rationally justify immoral ends. This seems like a very legitimate concern, until one realises that this need only be troubling if we are striving to conform to some notion of ethical rationalism. If the ethical is not identical to the rational, then the possibility of *rationally* justified immoral intentions is not a worry. Rather, we can conclude that it is possible to be fully rational and yet still be immoral. By simply admitting this possibility, the concern about intending immoral means to immoral

---

<sup>72</sup> This lack of clarity results in contrasting interpretations. Finkelstein (2005), for example, concludes that no reason is present, and so these intentional actions are not relevant for moral responsibility in Bratman’s account. In contrast, Chan (2000) argues that it shows that all intentional and intended actions are equally open to moral responsibility.

ends (like in the ethnic cleansing example) can be allayed. However, even if we were to embrace ethical rationalism, it is not obvious that the target of our concern should be the possible rational connection between intention-for-the-future and intentional action. I think a better plan of attack would be to argue that the balance of reasons that results in us identifying the intention-for-the-future in question as rational is in fact flawed. One does not have to prove that a rationally justified intention to undertake ethnic cleansing, for example, cannot and should not transfer its rational justification to the means to this end. One simply needs to show that holding this intention-for-the-future is not rationally justified. If this argument holds up, then ethical rationalism can be maintained without recourse to Bratman's dubious line of reasoning in this regard.<sup>73</sup>

This approach doesn't solve all the problems, however. There is still some difficulty with one of the underlying assumptions of ethical rationalism: that there is some objective measure of what counts as a rational end. Bratman clearly favours this view. Unlike Gauthier, who derives the answer to the question "What is rational to do?" from the notion of "living well," Bratman falls back on the idea of self-governance, with the view that effective self-governance entails some objective criteria that exclude certain types of action. That is to say, to be effectively self-governing would disallow immoral acts. How correct this position is I cannot say, and investigating it falls outside the scope of my arguments here.

## 5.2 *The problem of irrational intentions*

Turning to the problem of irrational intentions, I suggest that Bratman's contention that performing an intention requires a *striving* for rational coherence and consistency (as opposed to their *achievement*, as in Davidson) helps us make sense of cases of *akrasia*. This has been discussed in Section 2. The difficulty is that we generally think that we do not only bear moral responsibility for our intentional actions but also for failed intentional actions and intentional actions that result from irrational intentions (such as in cases of *akrasia*). This indicates that moral responsibility isn't derived from predicted success or rationality, but has to do, rather, with the presence of intentionality.

---

<sup>73</sup> It is worth noting that it is not my intention to argue either for or against the enterprise of ethical rationalism in this thesis.

So much for *akrasia*. But what about pathological behaviour? While it is usual for us to hold individuals morally accountable for various forms of *akrasia*, we do not usually do this in (all) cases of pathological behaviour. Bratman's account does not equip us to answer the question of why, for example, a kleptomaniac's actions are to be understood differently from those of a psychologically "normal" thief. What he can say is that the kleptomaniac lacks self-governance, and indeed pathological behaviour might be understandable in its entirety as a failure of effective self-government. But this seems to imply that these actions would then be both intentional *and* outside moral responsibility. Does this then mean that cases of volitional necessity are also immune to questions of moral responsibility? Surely the answer here is no. In any event, this approach once again relies on the idea of self-governance to do all the work. If for no other reason, this makes the theory untenable.

To summarise: Bratman makes an important contribution to the discussion of the relationship between intention and moral responsibility by stressing the need to distinguish between intended and intentional consequences of intentional action. On the other hand, there are some problems with his assumptions regarding ethical rationalism. However, I have shown that ethical rationalism does not lead to the problem of an immoral intention-for-the-future rendering rational an immoral intentional action, as Bratman thinks it does. As for the problem of irrational intentions, the introduction of the notion of motivational potential allows us to make sense of cases of moral responsibility related to *akrasia*, but not cases related to pathological behaviour, largely because in such case Bratman's account must once again rely on the notion of self-governance to do most of the explanatory work. As I have argued previously, this makes his position brittle to attack.

## **Conclusion**

The conclusions to be drawn from the foregoing analysis of Bratman's account of intention are the following: Bratman follows Davidson's approach in a number of respects, and nowhere more so than in his solution to the unity of the three uses of intention. As is the case with Davidson, Bratman argues that this unity can be achieved through a reduction of the three uses to intention-for-the-future as a mental entity. However, he explains the content of this mental entity quite differently from Davidson. For Bratman, intention-for-the-future is a conduct-controlling pro-attitude, characterised by its relationship to two rational regularities – means-end coherence and intention-belief consistency. Also of importance is the introduction

of the concept of motivational potential. I have shown that this approach is a direct improvement over the theories of both Anscombe and Davidson. Further, I have argued that the introduction of the notion of plans and their role in practical reasoning allows Bratman to explain the relationships between intentions in a more convincing way than either Anscombe or Davidson.

With regard to the epistemic requirements for intention, Bratman makes an important contribution by affecting a shift away from viewing intention as a rational judgement and towards viewing intention as *striving* toward a set of rational regularities, including intention-belief consistency. On the question of the relationship between intention, motivation and causes, Bratman once again follows Davidson. Both thinkers view intention as a pro-attitude, although they differ on what must be added to a pro-attitude for it to count as an intention. For Davidson, this pro-attitude must have the status of an unconditional judgement, while for Bratman it must be conduct-controlling. Although I will argue for a different take on the pro-attitude component of intention in Chapter 5, I think that the general view of intention as a pro-attitude is convincing. Bratman also agrees with Davidson in seeing intention as a cause. However, Bratman is still not able to overcome the problem of causal deviance. Where he does diverge sharply from Davidson is in claiming that intention is certainly not a reason for action. As I have shown, although Bratman's criticisms on this point are convincing, his own attempt to explain the role of reasons in relating intention-for-the-future to intentional action is not. The discussion of the relation between intention and reasons has led, in turn, to an analysis of the role of practical reason in defining intention. Here I have shown that Bratman's argument that we can maintain the commitment to means-end coherence even when intention-for-the-future does not count as a reason for intentional action is fatally flawed. Lastly, we have seen that, although Bratman's account is vulnerable due to its reliance on the problematic notion of self-governance, it does help us to understand how we might ascribe responsibility in some cases of *akrasia* and irrational intentions.

It should be clear from the foregoing analysis of Bratman's theory of intention that, while it advances our understanding of intention in several crucial respects, it also leaves the following difficulties unresolved: (i) Making sense of mutually exclusive intentions; (ii) the rational link between intention-for-the-future and intentional action; (iii) the relationship between intentions and plans; (iv) the relationship between intention and belief; (v) the rational link between intention-for-the-future and intentional action; (vi), the issue of

volitionally necessary intentions, and (vii) responding to the charge of cognitivism. In the next and final chapter I will propose an alternative account of intention that will retain the convincing aspects of Bratman's account of intention, while overcoming the difficulties that I have identified in my analysis of his account.

## CHAPTER 5: AN ALTERNATIVE ACCOUNT OF INTENTION

### Introduction

After having worked through the accounts of intention developed by Anscombe, Davidson and Bratman, it is now possible to set out what I consider to be a convincing account of intention. My particular aim in this chapter is to present an account of intention that (hopefully) fulfils all the requirements stipulated in Chapter 1 and overcomes the difficulties that have been uncovered in my examination of the accounts of Anscombe, Davidson and Bratman. The account of intention I will defend here consists of the following claims: First, the three uses of the concept “intention” can, and ought, to be unified under intention-for-the-future; second, that intention-for-the-future is a complex mental entity consisting of a revisable pro-attitude and a belief that the agent will try to fulfil this pro-attitude; third, the epistemic requirements for something to count as an intention is that the agent must be able to reflexively reconstruct, without external information, the belief component of her intention-for-the-future; fourth, that intention is a motivation, as it contains a pro-attitude component; fifth, that intention is a cause of action; sixth, intention is not a reason, but it is the result of reasons; seventh, and building on the previous claim, intention is not a rational judgement or an outcome of practical reasoning, but it is always related to the deliberative procedures that are such outcomes; eighth, agents are morally responsible for all the intentional consequences of their actions, which includes all intended and foreseen consequences; ninth, and last, in some cases of rationally foreseeable consequences an agent can be morally responsible, while in others they may be corrigibly responsible.

The persuasiveness of this alternative account depends, in large part, on its ability to resolve the problems inherent in the other theories of intention. Of the difficulties that I have discussed, there are fourteen that I think are relevant for my own account. They will appear as numbers (1) to (13) in this introduction and throughout the chapter. As each difficulty is related to the fulfilment of one of the criteria for a convincing account of intention, and as each criterion has a separate section devoted to it, I discuss each difficulty in the appropriate section.

As the reader should have come to expect, Section 1 focuses on the long-standing project of trying to unify the three seemingly irreconcilable uses of intention. We have seen that

Anscombe tries to unify the uses under the banner of intentional action, without success. Early Davidson's attempt at uniting the uses under intention-with-which is equally untenable, while later Davidson and Bratman offer more convincing solutions in seeking to reduce the three uses to intention-for-the-future, understood as a mental entity. Given the intractable problems presented by the alternatives, I follow Davidson and Bratman in seeking to unify the three uses under intention-for-the-future. However, there remain two significant difficulties in this regard. These are: (1) making sense of mutually exclusive intentions and (2) understanding the relationship between intention-for-the-future and intentional action. In this section I argue that these difficulties can be resolved by way of a more accurate definition of the *content* of the mental entity of intention-for-the-future, which in turn allows for a greater role for belief and the notion of *trying*. In terms of my proposed definition, intention-for-the-future is a composite of a pro-attitude and a special sort of belief regarding this pro-attitude. Further, I argue against the possibility, advocated by Searle, that (3) sometimes an agent can perform an intentional action without a preceding intention-for-the-future and only with an "intention in action" (Searle, 1980: 52). Finally, I examine the relationship between plural intentions. Here I show that "plans" play a substantial role in making sense of this relationship and that a proper understanding of this relationship also resolves the difficulty (4) of the relationship between intentions and plans.

Section 2 attempts to answer the question of what the epistemic requirements for intention might be. Since I argue for a position that is closest to intention-as-belief, the primary difficulties to be resolved are the following: (5) explaining belief as a necessary element of intention-for-the-future and intentional action, and (6) resolving the issues around epistemic opacity, specifically in cases of *akrasia* and pathological behaviour. I show that these difficulties can be overcome through the introduction of the following two requirements: (i) the belief component of intention must have the potential to be reflexively reconstructed without the addition of external knowledge and (ii) the pro-attitude component of intention must be revisable.

In Section 3, I try to make sense of the relationship between intentions and motivations, between intentions and causes, and between intentions and reasons. I argue that intentions are to be understood (i) as pro-attitudes, (ii) as effective causes, but (iii) *not* as reasons. Since it is now generally held, along with Davidson and Bratman, that intention, whatever else it is, is a pro-attitude, this aspect of my account is probably the least controversial one. However,

contrary to both Davidson and Bratman, I argue that this pro-attitude must be revisable, which brings me to difficulty (7): Is it possible for there to be volitionally necessary intentions, of the sort Bratman endorsed? If the answer is yes, then my account of intention is incorrect. I endeavour to show that there are no such intentions. As for my claim that intention is an effective cause, this raises the problem of causal deviance (8), which none of the theorists discussed here has been able to resolve. I try to resolve this difficulty by developing a more complete understanding of the relationship between intention-for-the-future and intentional action. Finally, the claim that intention does not have the status of a reason raises the problem (9) that it makes it difficult to answer the Anscombean “Why?” question. In other words, if an intention is not a reason, then what makes explanations of intentional actions *rational* explanations? I will argue that this lack of clarity can be resolved through the novel use of the principle of *Transmission Reasons*.

Section 4, the relationship between intention and practical reasoning, is a part of the thesis that I wish to leave fairly open, purely because a comprehensive treatment of practical reasoning would far exceed the scope of my argument. What I do want to do here is to resolve the two difficulties that I think must be overcome for any account of intention to be convincing. These are: (10) explaining the relation between intention and practical reasoning and (11) overcoming the charge of cognitivism. I argue that the answer to (10) is that intention is necessarily related to the deliberative procedures that result from practical reasoning, but is not equivalent to such a procedure. By doing this, I hope to avoid the problems entailed by Anscombe’s position that intention does not require practical reasoning, Davidson’s position that intention is a species of rational judgement, and Bratman’s position that means-end coherence and intention-belief consistency are the two regularities of practical reasoning that characteristically apply to intention. My response to (11) is that, although belief plays a necessary role in formulating an intention, this does not imply a cognitivist stance. To justify this response, I show that belief has a role in both theoretical and practical reasoning, and that this does not reflect a reduction of the latter to the former.

In Section 5, I examine the relationship between intention and moral responsibility. The two primary difficulties in this regard – which I identified as the two questions that had to be answered in Chapter 1, Section 5 – are: (12) Am I responsible for everything I do intentionally? and (13) Am I responsible for anything I do unintentionally? My answers to these questions are “yes” and “yes.” Defending my answer to the first question requires that I

resolve three important issues: (i) Explaining how the intentional, intended, foreseen and unforeseen consequences of intentional action relate to moral responsibility, (ii) providing a response to Knobe and Burra's claim that the moral evaluation of action occurs before intention is ascribed (and indeed partly determines this ascription), and (iii) explaining how moral responsibility functions in relation to intentional pathological behaviour. Issues (i) and (iii) are both resolved, I believe, by my explanation of what counts as an intention-for-the-future or an intentional action. I respond to issue (ii) by arguing that it confuses the moral evaluation of an action with ascriptions of moral responsibility to an agent. However, I concede that, on this point, my arguments might run counter to our everyday intuitions. After dealing with difficulty (12), I offer up a few provisional suggestions concerning difficulty (13), most importantly the notion of *corrigible responsibility* – which, I argue, is – a type of (non-moral) responsibility that pertain to certain rationally foreseeable, but unforeseen, consequences of intentional actions.

### **1. The unity of the three seemingly irreconcilable uses of intention**

In the account of intention I will present here, I start off by assuming that intention-for-the-future is a mental entity. I further contend that intentions-with-which and intentional actions can be reduced to intention-for-the-future. Clearly, I will be following the trail laid down by Davidson and Bratman. However, as far as the *content* of intention as a mental entity is concerned, I hope to blaze my own trail – one I believe resolves the difficulties which Davidson's and Bratman's accounts could not. My claim is that intention should be understood as a composite consisting of a pro-attitude and a belief, where the belief can be described as: the belief that I will *try* to achieve [what the relevant pro-attitude aims at]. This implies that what separates an intention from other pro-attitudes is not, as it is for Davidson, the presence of a rational judgement, nor, as Bratman asserts, that the pro-attitude be conduct-controlling in so far as it is open to certain rational requirements (or regularities). Rather, what is unique about the pro-attitude involved in intention is that the agent possesses a certain type of belief concerning that pro-attitude. Let us call this type of belief an *intention-unique belief*. My arguments for defining the content of the mental entity of intention in this way, as well as my counter arguments to potential criticisms, are presented in sections 2 to 4 below. In this section, I merely offer my suggestions for resolving the issue of the three seemingly irreconcilable uses of intention.

The relationship between intention-for-the-future and intentional action can be described as follows: I perform X intentionally when I perform X *with the belief that it was part of the fulfilment of* an intention-for-the-future Y. Another way of understanding this relation is to consider Bratman's notion of the motivational potential of an intention-for-the-future. What allows an intentional action to be typified as part of such a motivational potential is that *I performed action X in the belief that it was part of the fulfilment of an intention-for-the-future Y*. This point is best illustrated by an example: I have the intention-for-the-future to run a marathon. I then run the marathon *with the belief that it is in fulfilment of* my intention-for-the-future to do so. As such, this action can be said to fulfil the motivational potential of my intention-for-the-future to run the marathon. Therefore, I ran the marathon intentionally. Note that this approach assumes a tighter link between intention-for-the-future and intentional action than Bratman's alternative, which in turn has consequences for the causal and the rational relation between these two aspects of intention. On my view, the purpose of an intentional action is explained with reference to intention-for-the-future, which imparts intentionality to the relevant action. So, in the statement: "I am typing this sentence with the intention of finishing my thesis," the action described (typing the sentence) is an intentional action, while the intention described (finishing my thesis) is the intention-for-the-future that circumscribes the action.

### *1.1 Defending the reduction to intention-for-the-future*

Let us now examine how this conception of the unity of the three uses fares in resolving the two serious difficulties that have plagued Davidson's and Bratman's attempts to reduce the three uses to intention-for-the-future. These two difficulties can be formulated as: (1) explaining mutually exclusive intentions and (2) explaining the rational relation between intentional action and intention-for-the future.

The best example of (1) is probably still the gamer example provided by Bratman. His own account of intention solves the problem embodied in this example, but at the expense of providing a coherent solution to difficulty (2); (as we have seen, he introduces the arbitrary notion of a "settled objective" to fill the explanatory gap). A convincing account of intention would therefore have to resolve both (1) and (2), without relying on the notion of settled objectives. The fundamental problem that Bratman's example highlights is that, while we would normally consider cases of holding mutually exclusive goals irrational, it is unlikely

that we would call the gamer in the example irrational. If the gamer sets up the games as Bratman describes (see Chapter 4, Section 1), and then attempts to hit T1 and T2, we would not, in everyday life, take such behaviour to be an example of criticisable irrationality. Although I think Bratman's solution to this problem is flawed, I also think that it hints at way in which the problem might be resolved. I suggest that we ought to follow Bratman in jettisoning the "assumption of tight fit" between intention-for-the-future and intentional action, and concluding that in order to intentionally X does not require an intention-for-the-future to X. I further agree that his notion of "motivational potential" is a useful theoretical place holder for the set of criteria that should determine whether a given action is deemed intentional or not. In my view, however, Bratman's definition of the criteria for determining whether or not an action belongs to an intention's motivational potential is flawed (see Chapter 4, Section 1). I therefore propose the alternative criterion already mentioned above, namely that the action is *performed with the belief that it is part of the fulfilment of* [the given intention-for-the-future]. This approach retains a direct link between the intentional action and some *particular* intention. At the same time, it raises a new concern. If we say there is a direct link between intentional action and the particular intention it aims to fulfil, then we might have to concede that the gamer has the intention-for-the-future to hit T1 and the intention-for-the-future to hit T2. In this case, the gamer would be displaying criticisable irrationality, which means that my approach fails. On the other hand, if we follow Bratman and make the link too tenuous, then we run into difficulty (2): explaining the rational relation between intention-for-the-future and intentional action. I believe that the key to plotting a course between these two pitfalls is to give a greater role to uncertainty, by introducing the notion of *trying*.<sup>74</sup> By "trying," I mean that the agent believes that she may or may not succeed in her intention. The value of this is that it recognises the inherent uncertainty entailed by a future-directed commitment. Both Davidson and Bratman assert that an intention to X is *not* a commitment to *try* to X, but a commitment to *do* X. This means that, for Davidson and Bratman, one of the rational requirements applicable to my intention, is that I must intend the means that will be necessary to fulfil my intention. On the other hand, if we allow for the role of trying, the rational requirement applicable to intention is that I must intend the means that will be necessary for *trying* to fulfil my intention. I will now see how this approach fares in resolving the problem of the gamer.

---

<sup>74</sup> For a different approach to mine which also employs the notion of trying, see Thompson (2008: 91–92, 133–146).

With the incorporation of the notion of trying we can say that the gamer had the intention-for-the-future to hit one of the targets. This means that the agent has a pro-attitude to hit one of the targets, as well as the belief that she will try to perform this pro-attitude. Does the agent have an intention-for-the-future to hit T1 (or T2)? I argue yes, the agent does. One of the rational requirements applicable to my intention to hit one of the targets is that I should intend those means that would be necessary for trying to successfully fulfil my end. When the gamer decides to set out the two targets simultaneously, she does not know which one she will hit. Even as she is playing, she cannot be certain as to which target she will strike, if any. It could actually increase her chances of fulfilling her overall intention-for-the-future of striking one of the targets if she were to try to strike both simultaneously. In this case, it could be argued that the gamer is following the *most rational* path in this example, as she is maximising the chances of fulfilling her end. At the very least, by aiming to strike both targets she does not reduce the likelihood of fulfilling either intention. However, it would be irrational to act as the gamer in the example acts if, playing the two games simultaneously makes it impossible to hit either one of the targets (or even if it makes it less likely that the gamer will succeed in hitting one by *trying* to hit the other). And in such a case we would, I think, consider her behaviour to be a case of criticisable irrationality.

Could one not claim, however, that I must still, in the process of meeting the most rational means to the relevant end, hold an irrational combination of intentions? After all, even after denying any direct link between intention-for-the-future and intentional action, Bratman still needed to add the notion of settled objectives in order to make sense of this example. I contend that this is unnecessary. It would only be irrational to hold an intention-for-the-future to hit T1 while also holding one to hit T2, if by intention we meant that the agent will succeed, or is subjectively certain of success. If, on the other hand, we treat seriously the idea that intention entails *trying*, then it is no longer a matter of irrationality. There is no irrationality in *trying* to do two things of which only one can succeed, provided that the agent does not know which one will succeed. In this case the agent is simply “hedging her bets” so to speak. This does not mean that there are not cases where holding two intentions might be irrational. If, for example, *trying* to do X would make doing Y impossible, then I could not rationally intend X and Y. In the example of the gamer, if she were to succeed in hitting T1, and in the aftermath still had the intention of hitting T2, then she would be guilty of criticisable irrationality. However, this will not occur if we correctly allow for the role of intention-with-which. My intention-for-the-future to hit T1 (or T2) is subordinate to my

intention-for-the-future to hit “one of the targets.” That is, I hold the former intentions as means toward the latter intention. This is then obviously a case of intention-with-which. If the latter is achieved, then the reason for holding the former is gone. This means that if I intend to hit T1, with the intention of hitting one of the targets, then this intention to hit T1 only makes sense as long as I have not hit one of the targets. The moment I do strike a target, the intention-with-which is either fulfilled (if I struck that target) or unnecessary (if I struck the other target). In either case, I will not continue to hold the intention after the point of fulfilment, as the reason for holding the intention is no longer there. This solution retains a direct link between the relevant intention-for-the-future and the relevant intentional action and avoids having to introduce the notion of settled objectives.

Another possible solution – one that does away with the idea that hitting T1 and hitting T2 are intentions-for-the-future – goes as follows: I have a single intention-for-the-future, to hit one or other of the targets. I have two intentions-with-which, each related to either T1 or T2 respectively. The intention-with-which related to T1 would be, “I am trying to hit T1 with the intention of hitting one of the targets.” Obviously the action of trying to hit T1 would then be an intentional action, and my intention-for-the-future is still “hitting one or other of the targets.” This approach resolves the issue by avoiding us having to say that I have the intention-for-the-future of hitting T1 (or T2) specifically, and so avoiding the risk of irrationality. It also avoids having to introduce the notion of settled objectives. I think either of these explanations is acceptable, and that adjudicating between them would require asking the gamer what her intentions(-for-the-future) actually are.

The argument that trying is a necessary aspect of intention is not without its critics. Davidson, for instance, considers this argument in favour of trying as part of a larger attempt to include belief as a necessary element of intention. In his words:

The thesis that intending implies believing is sometimes defended by claiming that expressions of intention are generally incomplete or elliptical. Thus the man writing his will should be described as intending to try to secure the welfare of his children, not as intending to secure it, and the man with the carbon paper is merely intending to try to produce his copies. The phrases sound wrong: we should be much more apt to say he is trying, and intends to do it. But where the action is entirely in the future, we do sometimes allow that we intend to try, and we see this as more accurate than the bald

statement of intention when the outcome is sufficiently in doubt. Nevertheless, I do not think the claim of ellipsis can be used to defend the general thesis. (Davidson, 2001: 92)

Davidson then goes on to argue that the reason that claims of ellipsis cannot be used to defend the general thesis of intention as belief is because not all statements of intention can be accurately shown to have an elliptical form. I think that Davidson goes wrong here, precisely because in trying to make his point, he moves away from a discussion of *trying*, and instead takes it for granted that formulating an intention-for-the-future as an intention to try is equivalent to formulating an intention-for-the-future with *conditionals*. By an intention-for-the-future with conditionals, he means something like, “I intend to go to the party if the police do not arrest me (as I suspect they may).” Further, he argues that if all intentions would be formulated most accurately by clarifying *all* conceivable conditionals, then we would be left with the “nearly empty, ‘I intend to do it if nothing prevents me, if I don’t change my mind, if nothing untoward happens.’ This tells us almost nothing about what the agent believes about the future, or what he will in fact do” (Davidson, 2001: 94). For Davidson, this is evidence that an accurate description of intention-for-the-future does not require an explication of all possible conditionals, and that the explication of a conditional should be limited to cases where the agent considers the conditional as meaningful to the pursuit of the intention, e.g. “I intend to leave the party if the music is too loud” (ibid.: 95).

However, it is not at all clear to me that the notion of intending to try requires the explication of certain – or all – conditionals. Surely, it *is* more accurate, and not “near empty,” to assert that “I intend to go to the party” does not mean I *will go*, or I *am certain that I will go*, but that I *will try to go*. Indeed, this formulation reflects the agent’s commitment to achieving the relevant intention, whereas the intention-with-conditionals formulation does not. This allows us to maintain the role of uncertainty in the agent’s ability to predict future action (which is only accurate to reality), while not requiring what will always be an infinite list of conditionals (which would render the statement empty), and retaining the role of commitment to action.<sup>75</sup> In the end, the only argument Davidson has against *trying* is that when I say “I

---

<sup>75</sup> I think there is also good evidence in ordinary language usage for this point. When I fail to perform an intention, it is usual to say, “I tried” as a way of expressing that I did not revise my intention, but was thwarted in what I tried to accomplish. I think it is fairly rare for a person when asked, “Why didn’t you do it?”, to respond by listing conditionals and then declaring that since these conditionals occurred the commitment to act

intend to try” the “phrases sound wrong” (Davidson, 2001: 92). He takes this as evidence that “intending to try” should only be limited to those cases where the outcome is sufficiently in doubt and argues that to add “trying” to other cases is, in fact, adding a term that would render the intention-for-the-future in question meaningless.

I agree with Davidson that “I intend to try” sounds wrong. However, in my view, this is due to a different reason than the one proposed by Davidson. A word can feel “wrong” in a sentence for many reasons, but one of the most common is if that term makes the statement tautological. I suggest that this is the case with “trying” and “intention.” For example, it certainly sounds wrong to say “I am an unmarried bachelor.” This is not evidence that the two terms (“unmarried” and “bachelor”) in some way defeat each other. Rather, it is evidence that one of the terms already incorporates the other in its meaning. Given my contention that the intention to X should be understood as the belief that I will try to X, the full version of the statement “I have an intention to try to go to the party,” then becomes, “I have the pro-attitude, and the belief that I will try to fulfil this pro-attitude, to try to go to the party.” To clarify the source of the wrongness here, let us bracket out the pro-attitude component: The resulting statement reads: “I have the belief that I will *try to try to* go to the party.” The second “*try to*” is tautological. It is this tautology, I think, that lies behind Davidson’s feeling that the statement “sounds wrong.”

But, it could be asked of my solution, what about cases where we do express “try” in our statements of intention? Surely then these should also sound wrong, and the fact that they do not means that my explanation is flawed. I disagree. There are cases in ordinary language usage where the use of a tautology serves the purpose of emphasising an aspect of the agent’s meaning. For example: I explain my plans to break into a house and steal some valuables to a novice accomplice. I am dubious about my accomplice’s abilities of stealth and his moral readiness to break the law, so I express part of my plan as follows: “I am going to *stealthily* sneak into the house and then *illegally* steal the jewellery off the dresser.” Now this expression is clearly tautological, but the tautologies are purposeful. They serve to emphasise aspects of some of the terms employed – in this case, “sneak” and “steal.” In the same way, we sometimes use “trying” when seeking to emphasise the uncertainty inherent in our intention. This uncertainty is *always* present, as is the awareness of trying, but we usually

---

ceased to apply. Rather, the listed conditionals may serve to indicate why the agent was *thwarted* (the world did not co-operate sufficiently).

only draw attention to it in the tautological way I have just discussed in cases of, sufficient doubt. So, it is only when I seek to emphasise to my listener the uncertainty normally implicit in my intention that I will express this intention in the form “I intend to try.”

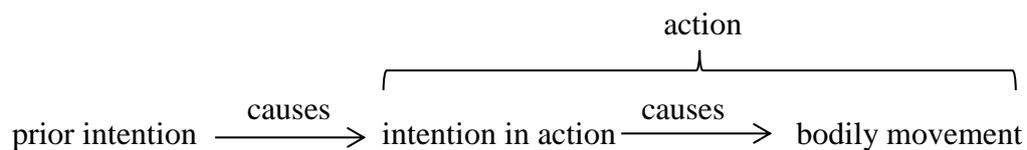
We might now ask: what is the pay-off of this approach to intention? I have highlighted the role of *trying* in order to resolve difficulty (1) – making sense of mutually exclusive intentions – without rendering an answer to difficulty (2) – making sense of the relationship between intention-for-the-future and intentional action – impossible. So, does my solution to (1) help us resolve difficulty (2)? Recall that this difficulty arises when rational between intention-for-the-future and intentional action becomes unclear, such as occurs in Bratman’s account. By rational link is meant the way that the reasons for a given intention-for-the-future are transferred to an intentional action. I believe that this difficulty is resolved by requiring an action to be directly related to the fulfilment of a particular intention-for-the-future – even if the intention-for-the-future is not aimed at the action itself – in order for such action to qualify as being part of that intention-for-the-future’s motivational potential and therefore as intentional. This solution does, however, raise the question of whether I am assenting to the view that the intention-for-the-future is a *reason* for that intentional action. I am *not* assenting to this, and my reasons for not doing so, and my arguments for maintaining the rational link between intention-for-the-future and intentional action without doing so, will be presented in Section 3.

### *1.2 Critique of Searle’s argument for intention in action*

This brings us to difficulty (3): Is it the case that all intentional action has an associated intention-for-the-future? To my mind, the most convincing argument in favour of the view that intentional action can occur independently of an intention-for-the-future is that made by Searle. He presents this argument in *The Intentionality of Intention and Action* (1980), in which he tries to sort out: “one of the messiest tangles of puzzles in contemporary philosophy,” to wit: “the problem of the relation of action and intention (Searle, 1980: 48). Searle’s solution to this messy tangle is to argue that intention-for-the-future can be divided into two types, prior intention and intention in action. Prior intentions are not always present in cases of intentional action. Intentions in action, on the other hand, are. It is worth noting that Searle’s arguments concerning intention and action take place under the umbrella of his attempt to formulate an account of Intentionality. Unsurprisingly, he sees intention-for-the-

future and intentional action as instances of intentional states, and so tries to show that they adhere to the rules that he has proposed for such states. As it is not my aim in this thesis to develop a comprehensive account of Intentionality, I will not involve myself with this aspect of Searle’s argument. I will concentrate instead on some of the individual points that Searle makes regarding intention and action. Contra Searle, I hope to show that these points can be used to justify my claim that any intentional action must have a preceding intention-for-the-future.

In terms of Searle’s definition of prior intention and intention in action, only prior intention is really an instance of intention-for-the-future as I have defined it.<sup>76</sup> Intention in action, on the other hand, is a *component part* of an intentional action – specifically: it is the cause of a bodily movement. Prior intentions, when present, are the causes of the intentional action *as a whole*. Searle provides the following diagram:



(Searle, 1983: 94)

I think there are two objections to Searle’s understanding of intention in action: (i) It cannot make sense of the Anscombean “Why?” question and (ii) his own argument for how the process of practical reasoning results in intentional action seems to allow for the presence of an intention-for-the-future in any event, removing one of the reasons to assume that there is such a thing as intention in action.

The first objection is that, if intention in action is the cause of bodily movement, and a component of the action in question, it cannot be a *reason* for the action itself. Hence, Searle cannot tell us why an explanation of an intentional action should be considered a *rational* explanation as opposed to a *causal* one (which is what the Anscombean “Why?” question requires). It might be possible to give a rational explanation of one’s bodily movement by appealing to intention in action (although I doubt this), but the action as whole cannot be rationally explained, as there is no intention or intention-like state causing the action in its

---

<sup>76</sup> From this point onwards I will use intention-for-the-future only for what Searle calls “prior intentions.” It does not include intention in action.

entirety. Presumably the action (which entails the intention in action and the bodily movement) will be caused by a desire, or by something else. In any event this will be a causal explanation, not a rational one.

I now turn to the second objection regarding the role of practical reasoning. Searle explains the role of practical reasoning in intention as follows: consider a case where I am dangling from a cliff. I am holding in my hands a rope that is attached to a set of supplies of substantial weight. Given the weight, I may not be able to save myself. In this case, Searle argues that “[t]he standard way the sequence of Intentional states would work is as follows:

I want (I rid myself of weight and danger)

I believe (the best way to rid myself of weight and danger is to loosen my hold)

And by practical reason this leads to a secondary desire:

I want (I loosen my hold).

And this leads, either *with or without a prior intention* [my italics], to an intention in action: the climber says to himself “Now!” And the content of his intention in action is:

I am loosening my hold.

That is:

This intention in action causes my hand to loosen its hold on the rope” (Searle, 1980: 68).

My objection to this is aimed at the notion of a “secondary desire” that leads to the intention in action. It is bizarre to claim that the practical reasoning process can generate a secondary desire in the absence of any prior intention! Further, what are we to understand under the statement, “this lead, either with or without a prior intention, to an intention in action?” Surely this “lead...to” deserves some explanation, otherwise we might take Searle to imply that secondary desire always results in an intention in action. Is it not possible that, instead of a secondary desire, I can simply state that an intention-for-the-future resulted from the reasoning process? After all, if Searle agrees that such a reasoning process has taken place,

why is that not sufficient for the presence of an intention-for-the-future? This point is best explained with the help of another example: I am walking down the road and see a stone lying on the pavement. I step toward it and intentionally kick it down the road. I do not have any greater goal or aim than the kicking itself. In this case, can I really be said to have an intention-for-the-future to go along with the intentional action? Anscombe, like Searle, answers no. In her view cases such as this demonstrate that intentional action is the fundamental use of intention. If Anscombe and Searle are correct – that is, if it is not the case that every intentional action has a related intention-for-the-future, so that the intentional action is *performed with the belief that it will be partly in fulfilment of* this intention-for-the-future – then my attempted unity of the three uses is incorrect. Unsurprisingly, I argue that there *is* an intention-for-the-future present in the case of the stone-kicker. In my account, an intention-for-the-future is a composite of a pro-attitude and an *intention-unique belief*. In the case of the stone-kicker there is clearly a relevant pro-attitude (perhaps analogous to the secondary desire of Searle), and I would argue that the belief is equally present. If somebody stopped the kicker mid-kick and asked him what he *believed* he was doing, I do not doubt that he would answer “I believe I am going (to try) to kick that stone.” Since both the components of an intention-for-the-future are present, intention-for-the-future can be said to be present, and my account is correct. What Searle might call an intention in action is, I think, really a description of the role played by intention-with-which. When I kick the stone, my intention-with-which can be considered to be a component of the intentional action, in the sense that only if the intention-with-which is present can the action be considered to be in the appropriate relation with the intention-for-the-future to count as intentional. I could say, “I am kicking the stone with the intention to kick the stone.” My intention to kick the stone consists of both a pro-attitude and an *intention-unique belief*, and so can be considered an intention-for-the-future. This intention-for-the-future is the cause of my kicking the stone. Intention-with-which (and intention in action) is nothing but the recognition of this relationship.

Consider an even more difficult example: I am walking down the street and somebody throws a ball at my head and I catch it on reflex. We would most likely say that my action was intentional, yet I had no beliefs regarding the ball prior to my catching it, and no awareness of any pro-attitude to catch it. In this case, my account of intention will probably be more controversial, for I would argue that, in fact, catching the ball is *not* an example of intentional action, precisely because of the lack of the above-mentioned elements. To defend my

position, consider the following: I have been informed that if I catch a ball thrown at me, a crazed philosopher will kill a member of my family. At some point I am walking down a street and somebody throws a ball at my head. I catch it on reflex. Have I just intentionally caught the ball? If so, have I intentionally doomed a member of my family? In this case I think that any agent would declare that she did not *intend* to catch the ball, she just did it *without thinking (about it)*. What this indicates to me, is that there is a necessary link between forethought and intentional action, and that forethought about intentional action will always entail an intention-for-the-future. Indeed, even in Searle's account the agent needs a secondary desire in order to perform an intentional action, and according to his explanation such a desire requires a belief. On my reading of his arguments, this indicates that Searle would also deny intentional status to this action.

### *1.3 Explaining the role of plans in cases of plural intentions*

The last topic I discuss in this section is that of plural intentions and the relation between them. In my view, Bratman is correct in thinking that intentions usually, though not always, form part of larger plans. These plans will usually incorporate multiple intentions, and serve the purpose of cross-temporal organisation for the agent. However, while Bratman claims that it is the relationship with the overarching plan that will determine whether or not a given action is part of the motivational potential of an intention-for-the-future, I do not think that such a plan is strictly necessary for the action to count as part of an intention's motivational potential. My account therefore resolves difficulty (4) – explaining the relationship between intention and plans – in a different way from Bratman.

My claim is that plans necessarily imply potential intentions, whereas intentions do not necessarily imply plans (potential or actual). That is, although all plans require intentions for them to succeed, not all intentions require plans in order for them to count *as* intentions. To explain why I say this, I first need to clarify what I mean by plans. Plans are part of the outcome of an agent's practical reasoning. Although there might be cases where an intention is only tenuously linked to a process of practical reasoning (a possibility I discuss later on), it is obvious that this cannot be true of a plan. I cannot honestly claim to have a plan, unless it is the outcome of some kind of reasoning process. Bratman, as we have seen, views plans as intentions writ large, and holds that plans, like intentions, must adhere to two norms of practical reasoning, namely means-end coherence and intention-belief consistency. Plans, like

intentions, that fail to meet these criteria are irrational. However, I think that this is oversimplifying matters. I would say that plans, as outcomes of practical reasoning, are required to adhere to the norms of practical rationality *whatever these might be*. This will doubtless include means-end coherence and intention-belief consistency, but there may well be many more.<sup>77</sup> More importantly, an irrational plan is by definition a plan that would likely fail, provided there are no unforeseen circumstances, whereas a plan that would likely succeed, provided there are no unforeseen circumstances, will usually be rational. This does not mean that all irrational plans fail, or that all rational plans succeed, it only means that *provided there are no unforeseen circumstances* an irrational plan is one that is likely to fail, and a rational plan is one that is likely to succeed. This follows from the function of practical reasoning as specified by Bratman (following Gauthier): the function of practical reasoning is to test the rationality of deliberative procedures, where this test is whether or not the given procedure enables the agent to “fare well,” or fulfil the motivation to self-govern. Ignoring for the moment the subjective nature of “faring well,” what this indicates is that at least one way that the rationality of a procedure is tested is by its probability of success in fulfilling the goal it aims at.<sup>78</sup> If we think that plans can be evaluated in these terms – as I do – then, given that no unforeseen circumstances plays a significant role, an irrational plan would be one that is likely to fail, whereas a rational plan would be one that is likely to succeed. Further, I think that in a case where an agent holds an irrational combination of intentions, it is the agent’s plan that is irrational, not her intentions (though, of course, the intentions could also be irrational). For example, I have a plan to turn my head left and right simultaneously. This means that I have two intentions: (i) “Turn my head left” and (ii) “Turn my head right.” Unlike in the gamer example, these really are mutually exclusive intentions, as even trying to fulfil the one will defeat my attempt to try to fulfil the other. Hence, this is an irrational combination of intentions. Yet neither intention is irrational in-and-of-itself. It is the plan that is irrational.

An intention, by contrast, is not held to the same standard. It makes sense to say that an individual has an irrational intention, and is likely to succeed in fulfilling it. To use a hackneyed example: I know that on the balance of reasons it is most rational for me to quit smoking, despite this I hold the irrational intention to smoke, and I succeed in fulfilling my intention. The irrationality of my intention to smoke does not relate to the chances of me

---

<sup>77</sup> Bratman (1987: 32) does admit this possibility himself.

<sup>78</sup> This is not that far removed from Anscombe’s understanding of practical reasoning.

successfully fulfilling the pro-attitude to smoke. This indicates that the rational requirements that hold for an intention in-and-of-itself are different to those that hold for plans.

So, how do intentions fit into plans? Intention is the link between a plan and the actions required for it to succeed. By itself, a plan does not compel us to act. I can have a plan that I never intend to act upon, and this would not be irrational. Indeed, this is a common part of life. We all make plans all the time that we never move to bring about, and this is perfectly acceptable. On the other hand, holding an intention commits us to action or revision of the intention. So whenever I move toward the execution of a plan, the potential intentions that it incorporates will be actualised, and hence compel me to act or risk criticisable irrationality. The role of the plan, therefore, is partly to organise my intentions in such a way as to allow for the fulfilment of my aim (whatever it might be). This is the reason why a plan, to be rational, must avoid containing incompatible combinations of intentions.

My arguments concerning the relations between intentions are limited to those intentions that are part of a plan. There are doubtless ways in which intentions that are not part of a plan relate to each other, however, I do not have any arguments on how such relations should be understood. Although this is certainly an area for future engagement, I leave the topic here.

To conclude: I have argued here that the unity of the three seemingly irreconcilable uses of intention can be found in intention-for-the-future, understood as a mental entity. This mental entity should be understood as a composite of a pro-attitude X, and an *intention-unique belief* that the agent will *try to achieve* X. My claim is that this approach, combined with my reformulation of what is meant by the motivational potential of an intention-for-the-future, explains the nature of the unity of the three uses. To prove this, I have shown that my account can make sense of the case of the gamer example – as an example of difficulty (1) – without recourse to the notion of settled objectives. Significantly, I have also shown that my solution to (1) also allows us to solve difficulty (2), as it explains the rational link between intention-for-the-future and intentional action. Furthermore, I have demonstrated that there is a necessary relationship between intentional action and intention-for-the-future – thereby resolving difficulty (3) – and countered the arguments advanced by Searle, as well as examples commonly used to prove the opposite. Finally, with regard to difficulty (4), involving the relationship between intentions, I have shown that, although plans are still an important element in understanding such relations, these do not take the form that Bratman

supposes. Instead, plans and intentions have different (though overlapping) sets of rational requirements and intention serves as the action-enabling element of plans.

## 2. Epistemic requirements

In this section I propose what I consider to be the most salient epistemic requirements for intention. Having done so, I then show how this approach either avoids or resolves the difficulties faced by the foregoing accounts as far as the epistemic requirements are concerned. The two most important features of my argument in this regard are the notions of *intention-unique belief* and the formulation, “*with the belief that*,” both of which feature in my definition of intentional action.<sup>79</sup> This argument is designed to resolve the following difficulties that previous accounts of intention could not address: (5) explaining belief as a necessary element of intention-for-the-future and intentional action, (6) resolving the issues around epistemic opacity, specifically in cases of *akrasia* and pathological behaviour.

### 2.1 Defending belief as a necessary component of intention

With regards to difficulty (5), my position is closer to that of the cognitivists, although I do not necessarily count myself among their number. Specifically, I see merit in the notion of intention-as-belief (see Chapter 2, Section 2), according to which belief is an intrinsic component of intention-for-the-future. My commitment to seeing intentional action as “performed with a belief” is a related matter I also deal with when considering difficulty (5). This view requires me to answer three possible objections: (i) that there might be cases where I intend X but do not believe I will X; (ii) that it is possible for me to believe I intend X, when I in fact do not intend X and (iii) that my reasons for intending to do something are not the same as the reasons for believing that I will do it.<sup>80</sup>

My response to objection (i) once again centres on the role of *trying*. Although I agree that it is not the case that whenever I intend to kick a ball, for example, I must believe I *will* kick the ball, I dispute that this shows that belief is not a necessary component of intention. When I have what I call an *intention-unique belief*, this belief is not a belief that I *will* X, but a belief

---

<sup>79</sup> For reference: I argue the *intention-unique belief* is one of the two components of an intention-for-the-future, and my description of intentional action is: “that which is done *with the belief that it is* in fulfilment of an intention-for-the-future.”

<sup>80</sup> Criticisms (i) and (ii) come from Bratman, and I have discussed them in Chapter 4, Section 4.3. Criticism (iii) is developed by Davidson (2001: 95).

that I will *try* to X. This approach allows us to make sense of examples, such as the one of the man writing his will (see Chapter 5, Section 1.1), by saying that the man might not have the belief that he will *succeed* in ensuring his families well-being, but he does have the belief that he will *try* to do so.

As for the second objection, I refer the reader to the lengthy discussion, in Chapter 4, Section 4.3, of the possibility of believing that I intend X while I do not actually intend X. I take the arguments I presented there to show that this objection does not hold.

However, that still leaves the third objection to the view that belief is a necessary component of intention-for-the-future. This is the claim that the reason for having an intention differ from the reasons for holding a belief. However, this claim only works if we assume that intention is *exclusively* a belief – something which I do not assume. Furthermore, I think it is precisely in cases of intention that the reasons for our belief and the reasons for our intention *are* the same.

In the account of intention I am proposing here, belief is a *component* of intention; it is not necessarily the whole of it. This means that my account does not require that the two sets of reasons be identical, since the reasons for a component of a whole need not be identical to the reasons for that whole. Nevertheless, I do think that an agent's intention and *intention-unique belief* share the same reasons. Consider how Davidson expands his argument:

Here is why I intend to reef the main: I see a squall coming, I want to prevent the boat from capsizing, and I believe that reefing the main will prevent the boat from capsizing. I would put my reasons for intending to reef the main this way: a squall is coming, it would be a shame to capsize the boat, and reefing the main will prevent the boat from capsizing. But these reasons for intending to reef the main in themselves give me no reason to believe I will reef the main. (Davidson, 2001: 95)

Davidson makes this argument after having dismissed *trying* as an element of intention. As I have argued earlier, his dismissal of *trying* is a mistake. If we were to allow *trying* to play its role, then Davidson's objection does not hold. Although the reasons for the intention to "reef the main," in the example above, may not be sufficient to lead an agent to "believe I will reef the main," it is certainly sufficient for her to "believe I will *try* to reef the main." Indeed, it

seems to me impossible for an agent to have the set of reasons Davidson lists, and not to hold the appropriate *intention-unique belief*. In conclusion, I submit that there is no compelling reason not to consider belief a necessary component of intention-for-the-future.

Let us now consider difficulty (5), namely whether or not all intentional action is necessarily carried out *with a belief*. I claim that it is. However, the most obvious counter to such a claim would be examples of actions that are clearly intentional, but where the agent does not believe that they are undertaking the actions as part of the fulfilment of an intention-for-the-future. Consider the following example from Davidson: I want to make ten copies of a document using carbon paper. I write on the top page, pressing down so that my writing may be imprinted on the page underneath. I do not know whether this will imprint through to the tenth page, though I can believe that it will or that it will not. What I do know is what I intend to do, what I do not is what I intentionally do (Davidson, 2001: 92). In this case, I may believe that the action will fulfil my intention-for-the-future, or I may not, but in either case the action is intentional, hence intentional action does not *require* such belief. If this is the case, then I cannot claim that belief is required for an action to belong to the motivational potential of an intention-for-the-future. But I have claimed that belief can have this role, so how to defend my account? Davidson's objection again assumes that intention entails a certainty of fulfilment, rather than a certainty that the agent will *try* to fulfil her intention. If we allow for trying, then we can say that, when I am making the imprint, I *always* believe that I am fulfilling part of *trying* to achieve the pro-attitude, which is to imprint all the way through the stack of papers. I cannot think of an example of what can be called intentional action that does not meet this requirement. To present this as a challenge: is it possible to act intentionally, without the belief that by doing so you are trying to fulfil a pro-attitude? I do not think it is. Yet again, the notion of *trying* allows us to circumvent a number of unnecessary difficulties thrown up by the accounts of Davidson, Bratman et al.

## 2.2 *The role of epistemic opacity and understanding irrational intentions*

Lastly, I consider difficulty (6), which involves the problem of irrational intentions, of which *akrasia* and pathological behaviour are prime examples. As I mentioned in Chapter 4, Section 2, I think that Bratman's resolution to the problem of *akrasia* is solid. Provided that we see the rational requirements of an intention as regulating ideals to be striven toward, rather than seeing intention as a rational judgement in itself (as Davidson does), then the solution is quite

straightforward. A case of *akratic* intention is a case of an irrational intention, which means a case of intending against the balance of reasons.

Pathological behaviour is more troublesome. Although such behaviour might well be an instance of irrational behaviour, there is an additional element at play that complicates matters. This element is epistemic opacity. One of the signature qualities of certain types of pathological behaviour is that the apparent intentions upon which an agent acts, or aspects of the reasoning processes that lead him or her to a given apparent intention, are not known to them. This raises the possibility that an agent might have an intention that they do not believe themselves to have (see Bratman, 2009a: 30). For example: if we consider Lady A from Galgut's article (discussed in Chapter 3, Section 5), one might say that Lady A leapt from her carriage because she had the intention to break her leg, but she was unaware of this and falsely believed that she jumped with the intention to save herself, because she was sure the carriage was about to crash. If this description is correct, it would mean that, in some cases, even my proposed recourse to "believing to try" would not be sufficient to save the role of belief. However, I think the description is mistaken: the behaviour is not intentional at all.

In order to support this claim, I must first say something about epistemic opacity. My argument is not that, in all cases where an agent is not consciously aware of the process leading to her intention, or lacks such awareness of the intention itself, the agent cannot have an intention. There are numerous situations where an agent performs an intentional act of which that they are not *consciously* aware at the time. This is obviously an instance of epistemic opacity, but one that I think differs from the ignorance about one's own intentions that characterises the subset of pathological behaviour mentioned above. If, for example, I am running a marathon, and as I run I wear out the soles of my shoes, but I do not *consciously* consider this fact as I run, then this action is epistemically opaque to me in a certain way. However, what separates this kind of opacity from the kind present in the subset of cases of pathological behaviour is that I could, without any external increase in knowledge, reflexively reconstruct my actions so as to be aware of my belief that I am wearing out my shoes. My belief regarding the intentional action is opaque only in terms of *conscious awareness*, and *no external knowledge must be added* for me to overcome this opacity. If after the marathon anybody asked me, what I believed about my actions to the fullest extent, then part of my answer would be that I believed I was wearing down my shoes – or better, that I believed I was wearing down my shoes as a part of the fulfilment of my intention-for-

the-future to run the marathon.<sup>81</sup> This stands in contrast to a case like that of Lady A. For Lady A, not only is her belief regarding the motivation to break her leg not conscious, but she also cannot form such a belief without *external knowledge being added*. She cannot reflexively reconstruct her reasoning process in order to identify the relevant pro-attitude she actually had – which would allow her to form the appropriate belief regarding this pro-attitude – without outside help, unlike in the case of the marathon example. Her belief regarding her motivation in leaping from the carriage is opaque to her, to a degree, and in a way, that my wearing out my shoes is not. Thus, the belief at stake in the latter example can be reflexively reconstructed without the addition of external knowledge, whereas with the former this is not possible. This leads me to add a requirement to *intention-unique belief*: this kind of belief must be of the sort that it can be reflexively reconstructed with only internal knowledge (without added external knowledge).

Is this sufficient to explain the example of Lady A? I have been carefully avoiding calling Lady A's motivation to break her leg an intention, but can I defend this position? My point about reflexive reconstruction would only be sufficient evidence that Lady A's motivation is not an intention if it was already accepted that belief is a necessary part of intention, but since this is exactly what is in dispute, this solution will not suffice. It seems to me that, if asked what her intention was in leaping, Lady A would provide us with what she *falsely* believes to be her intention. That is, she falsely believes that she believed that the carriage would have overturned and that she should leap to save herself. However, the actual reason for her leaping was that she wanted to break her leg to punish herself. What we have here is an example of a belief without a pro-attitude, and a pro-attitude without a belief. When she jumps from the carriage, she has a belief that this is to save herself, and this belief is false. The belief is false because Lady A did not in fact have the relevant pro-attitude. Lady A does eventually uncover the actual pro-attitude, but it requires additional, external, knowledge for her to do so. As such, she does not reflexively identify her belief in the appropriate way for it to be an *intention-unique belief*. At no point does she jointly possess both the relevant pro-attitude and an *intention-unique belief*, and as my account requires both to be present for intention-for-the-future, I argue that no intention-for-the-future is present.

---

<sup>81</sup> To illustrate this with another example: I am cutting wood. As I do so the blade of my axe is blunting. Am I intentionally blunting the axe? My account says "yes," because the agent can reflexively reconstruct the belief without external information that blunting the axe is a part of trying to fulfil the pro-attitude of cutting the wood.

How are we then to make sense of the fact that we would be inclined to call Lady A's leap an intentional action, given that, in terms of my argument, no intention-for-the-future has actually been formed? I am tempted to say that in these cases the *ascription* of intentionality to the action is in fact mistaken. Let me provide a science fiction example: I see a person buying an apple. I would reasonably assume that this is an intentional action. I then discover that what I took to be a person was in fact an android, with a difference engine brain and a pre-programmed series of routines (including buying apples). Once I become aware of this, I will realise that my ascription of intentionality was mistaken. An android, as long as it cannot do other than what it is programmed to do (cannot *choose* its actions), cannot have intentions ascribed to it by others or by itself.<sup>82</sup> This leads me to add a requirement to the pro-attitude component of intention-for-the-future, namely: the pro-attitude must be open to revision.<sup>83</sup> This gives us the final formulation of intention-for-the-future: a belief that I will try to achieve [the relevant pro-attitude, which is open for revision] which can be reflexively reconstructed with only internal knowledge. So could Lady A have chosen to act differently? I think not. In order for her actions to be open to revision (which perhaps is a component of self-governance in Bratman's sense), the pro-attitudes that motivate them cannot be epistemically opaque to her in the way that her pro-attitude to punish herself is. That is, she could not have chosen differently without the input of external knowledge. For this reason, I think it justified to say that Lady A's action in leaping was not intentional.

This does not mean that all instances of pathological behaviour are necessarily unintentional, or that no intention-for-the-future is ever present in such cases. What is crucial is the kind of epistemic opacity at stake, and its bearing on the possibility of revising the pro-attitude component of my intention. Thus, if Lady A came to fully understand her pathology, our description of her actions would change. If she were to duplicate her earlier leap, but this time be in full awareness of her actual pro-attitude, and by extension with the correct *intention-unique* belief then we would be justified in saying she has an intention-for-the-future,

---

<sup>82</sup> I think there is also room here to consider what this implies for animals and plants. Is it possible for flora or fauna to have an intention? This is a fascinating question, but outside the scope of this thesis. However, I cannot help but think that the answer would revolve around whether or not the animals or plants *could have* behaved differently from how they did behave. Alternatively this could be a discussion about choice, perhaps relating to Anscombe's insight mentioned earlier that intention is intimately linked to choice and self-ascription. For his part Searle certainly believes that at least animals can be considered to perform intentional actions (Searle, 1980: 64).

<sup>83</sup> This relates to Bratman's idea of volitionally necessary intentions, as by requiring the pro-attitude component to be revisable, I am rejecting the possibility of such volitional necessity of the whole. This will be discussed in Section 3.

provided that she could have chosen against this pro-attitude. Now that her pro-attitude is epistemically transparent, she should be able to revise her actions. If her actions are not revisable – if the pro-attitude to inflict self-punishment is so powerful that no revision is *possible*, even after the pro-attitude is known – then no intention-for-the-future is present. In such a case, I do not believe that Lady A would ascribe intention to herself, and we should not ascribe it to her. This conclusion will have significant consequences for the discussion of responsibility in Section 5.

Overall, it is my conclusion that difficulty (5) – the arguments for why belief should not be understood as a necessary component of intention – do not hold. For this reason, I take my account to be justified in treating this relationship as necessary. In response to difficulty (6), I have adopted a view of *akratic* intentions that derives from Bratman’s account, namely that these are instances of irrational intention. With regards to pathological behaviour, I have argued that it is a requirement for an *intention-unique belief* that it must be possible to reconstruct the relevant belief upon reflection without the input of external knowledge, and that a requirement of the pro-attitude component of intention is that it must be revisable. Armed with these requirements, I have argued that the cases of pathological behaviour that are meant to support Bratman’s claim that an agent can hold an intention without knowing it, are not, in fact, instances of intention at all. Having shown this, I have also argued that not all pathological behaviour falls into this set, and that in some (perhaps most) cases of pathological behaviour an intention-for-the-future is indeed present, and by extension such behaviour can be deemed an instance of intentional action. By resolving all three the difficulties relating to the epistemic requirements for intention, I have shown my account to be more convincing than the three alternatives I have discussed.

### **3. The relationship between motivations, causes, reasons and intention**

This section focuses on three distinct relationships: the relationship between intention and motivation, the relationship between intention and causes and the relationship between intention and reasons. My discussion of the first of the three relationships centers on why the pro-attitude component of intention should be understood as being revisable. To support my argument that it should be, I argue against (7) the possibility of volitionally necessary intentions. Secondly, I consider the relationship between intention and causes and try to solve (8) the problem of causal deviance. Lastly, I try to resolve difficulty (9) – whether or not

intention-for-the-future should be considered as a reason for intentional action – through the novel application of the principle of *Transmission Reasons*. This final point will, unsurprisingly, run over into the next section concerning practical reasoning.

### *3.1 Intention as a revisable pro-attitude*

I claim that the pro-attitude of component of an intention-for-the-future must be revisable. For this to hold, there must be a reasoning process at work, or at least the potential for such a process, otherwise revision would not be possible. What is more, if the pro-attitude component of an intention must be revisable, then the intention as a whole must be revisable. This position is directly at odds with Bratman's notion of volitionally necessary intentions – that is, intentions that I cannot will to be other than they are and that are therefore unrevisable – which I raised as difficulty (7). However, consider again the example of the android purchasing the apple. I have argued that we would not consider this an intentional action precisely because the android is incapable of revising its pro-attitudes. It cannot choose. This can be expressed generically as: when we see a person who *cannot help but act* in a certain way (X), and then ask them “Did you perform X intentionally?” or “Did you intend to X?” the usually answer will be “No, I could not act otherwise” or “No, I could not stop myself.” It is doubtful that in such cases the agent would ever provide a rational explanation when asked the Anscombean “Why?” question. Obvious examples of this are all involuntary bodily movements. This is not to say that I would necessarily be distressed, or in disagreement, with such actions. I am thankful when my body involuntarily retracts from a fire, and might add that, had I been able to, I would have performed such a retraction intentionally. However this does not change the fact that this action was involuntary. If I am right about this, then “volitionally necessary intentions” are either a contradiction in terms, or “volitionally necessary” does not actually entail an absolute inability to revise. I am not sure how one would think of volitionally necessary, yet revisable, intentions.

Perhaps one might say that an action governed by such an intention is an action I would choose against the balance of reasons (or alternatively, against my own ability to self-govern). For example, I might have a pro-attitude to protect my physical integrity (which is not inherently rational or irrational) that is so motivating that, even though the balance of reasons advises against it, I still choose to act in accordance with this motivation. How far this approach can take us, I am not sure. The important point for my purposes is that the idea

of volitionally necessary intentions are being unrevisable is certainly untenable. I leave the topic of volitionally necessary intentions here and turn to the issue of intentions, causes and reasons.

### 3.2 *Defending intention as a cause of intentional action*

One of distinct advantages of my approach to understanding intention as a pro-attitude is that it allows for a more convincing solution to questions regarding intention's relationship to causes and to reasons. I concur with Davidson and Bratman that intentions should be considered the causes of intentional actions, and that this common-sense assumption, which is in line with our everyday intuitions, should be maintained until such time as there is compelling evidence that this position is untenable. For this reason, I will not be arguing for *why* intention should be seen as a cause, but rather I will assume this, and look to defeat any argument that seeks to disprove it. As concerns the thesis, this means that the claim that intentions are causes of intentional action can only be considered convincing to the extent that it is able to resolve (9) the major problem of causal deviance, as this problem directly threatens the idea of intentions as the causes of intentional actions. I think the solution to this problem lies in the appropriate understanding of the link between intention-for-the-future and intentional action. This link is best expressed as follows: "I perform X intentionally when I perform X with the belief that it was part of the fulfilment of an intention-for-the-future Y."

Now consider the everyday example of causal deviance I have discussed previously: "I have the intention-for-the-future to signal my confederate by blinking. Due to the stress of having this intention-for-the-future; I blink as a nervous tic." So, in this case, I blink *because* of my intention-for-the-future, but I do not blink intentionally. However, if intentional action is defined as an action undertaken *because of* an intention-for-the-future, then why is the blinking in this case not intentional? In my view, such blinking would only count as intentional if the action were undertaken concurrently with a *belief* by the agent that the action was being taken *in order to fulfil* the intention-for-the-future. This means that blinking due to nervousness would not count as intentional, as I would not be acting on the belief that I am blinking to achieve the fulfilment of the intention-for-the-future. In fact, when asked why I blinked in this case, I would probably provide a wholly causal rather than a rational explanation for the action ("I blinked because I was nervous and it made me twitch"). If this solution to the problem of causal deviance holds, as I believe it does, then there is no present

reason not to assume that intention is a cause of intentional action. As such, I will be maintaining this assumption.

### *3.3 Intention-for-the-future and intentional action as sharing reasons*

Next I tackle the relationship between intentions and reasons. As the two previous chapters illustrated, this is almost certainly one of the two major problems facing contemporary accounts of intention (the other being the cognitivism debate). To briefly recall difficulty (9): if intention-for-the-future is accepted as being a reason for intentional action, we are faced with two serious problems develop: (i) unacceptable bootstrapping in practical reasoning and (ii) immoral ends can be considered to be justified reasons for the immoral means to these ends. On the other hand, if we reject any rational link between intention-for-the-future and intentional actions, we are faced with two different problems: (i) an inability to properly answer the Anscombean “Why?” question and (ii) an inability to explain why there is rational compelled on an agent to appropriately match her intentional actions to her intention-for-the-future. Given these two sets of problems, my account aims to reject the idea that intention-for-the-future provides a reason for intentional action, while explaining the rational link between intention-for-the-future and intentional actions in a way that avoids the difficulties normally associated with this view. I submit that the solution lies in a new application of the principle of *Transmission Reasons*, which Bratman adopts from Setiya (see Chapter 4, Section 4.1).

*Transmission Reasons* is the principle that: If I have a reason R in favour of X, and the means M is necessary for X, then R is a reason for X, and a reason for M. Bratman employs this principle in order to argue that, if X is taken to be “self-governance,” and M is taken to be “Means-End coherence,” then we are rationally compelled to conform to M, provided that we possess “an intrinsic commitment to self-governance” as R (Bratman, 2009b: 432-434). This line of reasoning is meant to explain why we are rationally compelled to match intentional actions to intention-for-the-future, without having to claim that the latter is a reason for the former in-and-of-itself. I have argued in the preceding chapter that this argument fails due to its overreliance on the notion of self-governance, and I will not repeat it here. My aim in this section is to show that we can solve the problem of the relationship between intentional action and intention-for-the-future by applying the principle of *Transmission Reasons* in the appropriate context.

That is: allow my intention-for-the-future to be X, my intentional action to be M, and R to be the outcome of the reasoning process that conditions my intention. On this view, the reasons for my intentional action (for intending the means) are those reasons that have resulted from the process of practical reasoning. And these self-same reasons are also the reasons for my intention-for-the-future. In other words, the rational link between intention-for-the-future and intentional action is that they *share this set of reasons*. My reason for performing intentional action M is not my intention-for-the-future X, but rather the set of reasons that has led me to my intention-for-the-future in the first place. This approach avoids bootstrapping as simply having the intention-for-the-future does not *add* any reasons to the balance of reasons. The problem of immoral ends is also avoided, since, even if I did have an immoral intention-for-the-future, this wouldn't count as a reason for an intentional action. However, there is still the possibility that the underlying reasons themselves might be immoral. This will be discussed further in Section 5, as it is directly related to questions of moral responsibility.

An account of intention that does not take intention-for-the-future as a reason for an intentional action allows an agent, when asked the Anscombean "Why?" question, to provide a rational explanation for her intentional actions, or for her intention-for-the-future, by providing the reasons that underlie them both. As for the issue of matching means and ends we can now say that it involves the rational process that gives rise to the agent's reasons for her intention-for-the-future (end) *and* intentional action (means). I take this process to be practical reasoning, and the reasons that are the outcome of this process as having to conform to the rational procedures that characterise this process, one of which I take to be means-end coherence. I continue this discussion in the next section, as it concerns intention's relationship to practical reasoning.

In light of the above, I propose the following understanding of causal and rational explanations of intention. As I have already stated, any behaviour that can be given a rational explanation will also have a causal explanation, but not all behaviour with a causal explanation will have a rational explanation. When I act with an intention, I act for reasons, but my intention is not itself a reason. When I act intentionally, there are reasons for my actions, but these are not my intention-for-the-future. My intention, by itself, is the cause of my actions, but it is not the reason for them. If I answer the Anscombean "Why?" question by simply providing my intention with nothing is added, this *would* be a causal explanation of my actions. The rational explanation only comes into play when I explain the *reasons for the*

*intention* (intention-for-the-future or intentional action), and hence – for they are the same – *the reasons for the action*. So, for example, if someone asked me, “Why did you kick the stone down the street?” and I answered, “Because I intended to do it,” this would be a causal explanation of my actions. It would only become a rational explanation if I expanded my answer along the lines of: “I intended to because I wanted to hit that mark on the wall across the road.” In other word: an explanation *by* intention is a causal explanation, but an explanation *of* intention is a rational explanation.

Taken overall, the discussion in this section has rendered the following insights: To begin with, the motivational component of intention is best understood as a revisable pro-attitude. To support this view, I have shown that what Bratman calls volitionally necessary intentions (7) are either not actually intentions, or are not immune to revision in the way he claims. In the second place, intention is to be understood as an effective cause, and difficulty (8) – the problem of causal deviance – is defeated once we understand the relationship between intention-for-the-future and intentional action in the following way: an action is intentional if, and only if, it is performed in the belief that it is part of the fulfilment of the relevant intention-for-the-future – that is, the intention-for-the-future to which the belief pertains. Finally, with regards to intentions and reasons, I have argued that intention-for-the-future does not count as a reason for intentional action, and that the proper way to understand the relationship between the two uses, so as to avoid the main difficulty that has beset previous attempts – namely difficulty (9) – is to employ the principle of *Transmission Reasons*. This principle allows us to see that the rational link between intention-for-the-future and intentional action lies in their *shared* reasons resulting from practical reasoning. Understanding this final component takes the discussion into the next section.

#### **4. The relationship between intention and practical reasoning**

As I detailed in the opening paragraphs of this chapter, I will defend the claim that intention is necessarily related to practical reason through the deliberative procedures that are the outcomes of practical reasoning. In the course of defending this claim, I will specify the nature of the relationship between intention and practical reason and show how this account of intention avoids the charge of cognitivism. In doing so, I will be addressing difficulties (10) and (11) respectively.

#### 4.1 Intention related to practical reason through deliberative procedures

While question (10) cannot be fully addressed without a full and proper understanding of practical reason, an exhaustive account of practical reason falls outside the scope of my thesis. What I want to do here is to outline what I think to be the most convincing claims about the relation between intention and practical reason, while making as few assumptions as possible about the workings of practical reason itself. I begin with what I take to be two generally uncontested assumptions. The first is that intention is *always* related to a practical reasoning process, although sometimes this process is not epistemically transparent to the conscious mind of the agent, and can only be reflexively reconstructed.<sup>84</sup> This is not a controversial view, and follows both Davidson's and Bratman's arguments.<sup>85</sup> The second assumption is that intention is specifically related to the outcomes of the reasoning process. That is, there is a movement from motivation to intention via the process of practical reasoning. This builds up to the third, more controversial assumption, and one that I will defend at some length, namely that intention, while necessarily *related* to practical reason, is not itself a conclusion formed at the end of a practical reasoning process. Rather, it is guided by the deliberative procedures that are such conclusions. This guidance takes place because those deliberative procedures that apply to intention represent rational requirements that the intention must strive to fulfil. For this reason, I claim that intention stands in a mediated relationship to practical reasoning.

This last point is clearly opposed to Davidson's account, which treats intention in its entirety as a rational judgement that is itself the conclusion to a practical reasoning process. My position is closer to that of Bratman, in so far as he thinks that practical reasoning is directly concerned with our *deliberative procedures* rather than with our individual intentions (or actions). In embracing this position, Bratman posits a mediated relation between intention and the reasoning process, with intention being related directly to our deliberative procedures, rather than to the practical reasoning process itself. While I agree that the relationship between intention and practical reasoning is mediated by deliberative procedures, I want to avoid having to identify the full range of deliberative procedures at stake here – e.g. means-end coherence and intention-belief consistency – simply because specifying all such

---

<sup>84</sup> Importantly, as I have said in Section 2, this reconstruction must be able to take place without the addition of external information.

<sup>85</sup> But not Anscombe. In Chapter 2, Section 4, I discussed, and argued against, her assertion that intention is possible even without the presence of practical reasoning.

procedures would require a much more comprehensive account of practical reasoning than I am able to provide here. This is undoubtedly a part of my account of intention that should be taken as very provisional, and I think there remains much work to be done on this topic. However, in order to continue with my argument, I will take it that *any* process of practical reasoning yields one or more deliberative procedures that will then apply to intentions.

As should have become clear from my argument up to now, I propose that intention should be understood as comprising of two components: a revisable pro-attitude and an intention-unique belief. An intention-unique belief is the belief that the agent will try to fulfil the pro-attitude in question. Where there is both a relevant pro-attitude and an intention-unique belief, the agent can be said to have an intention-for-the-future. This intention-for-the-future is always open to rational scrutiny, and modification in light of this scrutiny, based on the deliberative procedures that are applicable to intention. These procedures are, of course, the results of the agent's practical reasoning. For this reason, I claim that the relationship between intention and practical reasoning is necessary, but mediated by deliberative procedures. It is worth noting that the intention, though open to rational evaluation, need not be in line with the guidance of the deliberative procedures. I could still choose against the balance of reasons. This would, however, always be an irrational intention.<sup>86</sup> Our intention is, then, not a deliberative procedure itself, but (mostly, or at least sometimes) they are shaped by the demands of such procedures. That is, we are *rationally* compelled toward rational intentions.

Does it follow from the fact that intention is related in this way to practical reasoning that the intention must itself be rational? I do not think it does. Firstly, irrationality can result from a flawed practical reasoning process, which yields a flawed deliberative procedure. Because of this flawed procedure, the agent might mistakenly believe that her intention is rational, when in fact it is not. Alternatively, the reasoning process might be valid and the deliberative procedures rational, but the agent might still act, by choice or not, in opposition to the guidance of these procedures, such as in a case of *akrasia*. This is a far more contentious possibility. It requires the assumption that an agent can choose against the balance of reasons,

---

<sup>86</sup>On an interesting note, if we take the choice by the agent to intend against the balance of reasons to be an intentional action (and I see no reason why it should not be), this would merely mean that this choice is an irrational intention itself, as it is open to the same rational scrutiny as the chosen intention. If the balance of reasons favours not intending X, then the reasons for choosing X cannot be rational, presuming that the reasoning process is not flawed in some way.

that is, against her practical reason. Whether or not this is possible is a fascinating question. For now, I will simply state that it does not seem *impossible* that an agent can knowingly choose against her rational judgement (in cases such as these, it would be common to say “against her better judgement”).

I have already argued that intention-for-the-future should not be thought of as giving a reason for intending its means, as this assumption would import illegitimate bootstrapping into an account of intention (see Chapter 5, Section 3.3) But what about pre-existing intentions: how do they interact with the practical reasoning process? It might be tempting to think that a previous intention to X somehow entails a special reason for X, which is then introduced into the current reasoning process, but this would be a mistake. Once again, this would lead to a bootstrapping problem, for it would allow an agent holding an irrational intention to introduce this intention into her practical reasoning as a reason, possibly shifting the balance of reasons. This problem is avoided, however, if we do not treat intention itself as a special reason within the reasoning process, but rather as being conditioned by reasons. It is these conditioning reasons that enter into the reasoning process, and not the intention itself.

Although there is still a great deal to be said about the way which intention relates to various kinds of deliberative procedures, I think that the argument I have presented here does give a convincing account of the basics of this relationship, and in doing so resolves difficulty (10). Perhaps more importantly, it does so without requiring intention to be either a rational judgement or a reason in of itself, both of which, as we have seen, are unsustainable assumptions.

#### *4.2 Defending intention-as-belief without cognitivism*

When it comes to the question of how my account deals with the charge of cognitivism, I should say, first, that I want to retain the separation between theoretical and practical reasoning, while still allowing belief to play a necessary role as a component of intention. On the face of it, this isn't an easy position to defend, since theoretical reasoning is usually taken to be reasoning about the truth of our beliefs, whereas practical reasoning concerns reasoning about actions. It might appear, therefore, that my attempt to maintain both the separation of the two reasoning processes *and* a role for belief is a case of wanting to have my cake and eat

it. However, in what follows, I will show that my account of intention is not open to the charge of cognitivism, as I do not reduce means-end incoherence to an incoherence of beliefs.

Recall that the turn to cognitivism begins with the attempt to salvage means-end coherence, after the rejection of the idea that intention-for-the-future counts as a reason for the intentional action, which, in turn, constitutes its necessary means. In order to achieve this, the cognitivists argue that means-end coherence (and by extension, means-end incoherence) is in fact a case of the (in)coherence of beliefs. The arguments they make to advance this position center on whether or not instances of means-end incoherence are, or are not, reducible to incoherencies in belief.

However, explaining means-end incoherence as, at base, a matter of the incoherence of beliefs, reduces a requirement usually associated with practical reasoning (means-end coherence) to one that is required by theoretical reasoning (the coherence of my beliefs). There are two problems with such a cognitivist approach: (i) it requires belief to be associated with intention to an untenable degree, and (ii) it fails to account for the fact that practical reasoning results in a commitment to act, or at least guidance toward action, whereas theoretical reasoning does not.<sup>87</sup>

I have already provided counter-arguments to problem (i) – that intention and belief cannot be as closely associated as my account of intention requires (see Section 2.1). As such, I take it for granted that the role of beliefs in intention is not in dispute. I have also shown why we ought to reject the idea that intention-for-the-future counts as a reason for intentional action. However, I agree that problem (ii) – that a cognitivist account of intention fails to recognise practical reason's unique role in directing an agent's actions – is a legitimate concern. This leaves me in a position where, to avoid problem (ii), I must find a way to explain means-end incoherence that does not reduce it to theoretical reasoning. In order to do this, let us start by examining the following example: I have the intention-for-the-future to make a salad. One of the necessary means to fulfil this intention-for-the-future is that I must buy some carrots. Yet, when I go to the store, I intentionally purchase a steak instead, despite maintaining my intention-for-the-future to make a salad. In doing this, I am means-end incoherent. I did not match my intention-for-the-future (my end) with the appropriate intentional action (the necessary means), and the result is that my behaviour is irrational. As we have seen in the

---

<sup>87</sup> I have discussed this in Chapter 4, Section 4.2.

previous chapter, Bratman argues that if incoherencies of this sort were explained as being the result of incoherent *beliefs*, then we would have reduced practical reasoning about intention to theoretical reasoning, precisely because theoretical reasoning is aimed at beliefs, and practical reasoning is aimed at action. (Bratman, 2009a: 21-22; 2009b: 426-428)

However, I do not think that we are compelled to consider means-end incoherence to be a matter of belief incoherence. Rather, I suggest that means-end incoherence occurs when the practical reasoning process that provides the rational grounding for deliberative procedures about action – such as the requirement of means-end coherence – is flawed/fails. Crucial to my argument here is the role of *Transmission Reasons*. An agent has a substantive reason to adhere to means-end coherence because this requirement is a deliberative procedure that applies to intention. It provides me with a reason to act a certain way, if I wish to fulfil my aim (the relevant intention) in the most rational way (according to my practical reasoning). As such, reasons are shared between my intention-for-the-future and my intentional action, via *Transmission Reasons*; the latter, in other words, function as reasons for both uses of intention. When an agent fails to adhere to means-end coherence, then this is either a case where the agent reasoned poorly, and so misidentified the necessary means to a given end, or it is a case where the agent, in full awareness of the balance of reasons, chooses to intend the incorrect means to an end. In both cases the agent is guilty of criticisable irrationality. Yet in neither case does belief need to play a role in explaining this incoherence. I do not want to claim that this argument is a foolproof defence against the charge of cognitivism. However, I do take it to be a provisional foundation upon which a more complete defence may be built.

To conclude, this section sought to answer two questions: (i) What is the role of intentions in practical reasoning? and (ii) Where does my account stand in the cognitivist debate? In answer to the first question, which I take to be difficulty (10), I have argued that, although intention does have a necessary relationship with practical reasoning, it should not be thought of as a rational judgement or as a special reason, either toward its own means, or for the purposes of further practical reasoning. Rather, the relation between intention and practical reasoning is mediated by the deliberative procedures identified by the practical reasoning process and then applied to intention. As regards question (ii), which I have identified as difficulty (11), I have tried to show that my account of intention avoids cognitivism – that is, reducing aspects of practical reasoning to theoretical reasoning in so far as it is able to explain means-end incoherence without reducing it to a form of belief incoherence. I suggest

that such an explanation can be developed through the application, once again, of the principle of *Transmission Reasons*. This being said, my answers to both these difficulties should be considered provisional, as I noted at the beginning of this section. I think that a more complete understanding of practical reasoning (and perhaps, in the case of the cognitivism discussion, theoretical reasoning) will be required before arriving at any truly conclusive answer to these difficulties.

## **5. The relationship between intention and (moral) responsibility**

In Chapter 1, I stated that the final criterion for a convincing account of intention is whether such an account allows us to answer two questions about intention and moral responsibility, each of which represents a difficulty that a convincing account of intention must be able to overcome: (12) Am I responsible for all the consequences of intentional action? and (13) Am I responsible for anything I do unintentionally? I argue that the answer is a definite “yes” in both cases. In order to justify the “yes” to question (12), I have to do three things: explain how the intended, foreseen and unforeseen consequences of intentional action relate to moral responsibility, provide a retort to Knobe and Burra’s claim that the moral evaluation of action occurs *before* intention is ascribed (and indeed partly determines this ascription) and finally explain moral responsibility in cases of intentional pathological action. After addressing these three issues, I offer a provisional answer to question (13). This answer is that an agent is not morally responsible for any rationally unforeseeable consequence of intentional action, but is morally responsible for some rationally foreseeable, but unforeseen, consequences. Further, I suggest that there are some cases of rationally foreseeable consequences where it is not moral responsibility that applies, but rather what I term *corrigible responsibility*.

### *5.1 Moral responsibility and the consequences of intentional action*

Starting with the question: Am I morally responsible for all the consequences of intentional action?, we know that Anscombe answers with a “no,” Davidson’s answer is ambiguous and Bratman does not provide any answer at all, although he does admit that intention is linked to responsibility somehow (Bratman, 1984: 402-403). Broadly speaking, I concur with Anscombe that we are responsible for all the intended and intentional consequences (Anscombe uses these interchangeably), but not the unforeseeable consequences, of our intentional action. I differ, however, in what I take to be included under the banner of

intentional consequences. Contrary to Anscombe, who distinguishes intentional/intended consequences on the one hand, and foreseen consequences on the other, it is my claim that intentional consequences encompass both the intended as well as the foreseen consequences of intentional action.

In Chapter 2, Section 5, I showed that Anscombe identifies three categories of consequences of intentional actions, namely intended, foreseen and unforeseen consequences. Thanks to the arguments of Knobe and Burra, I further showed that we needed distinguish between *intended* and *intentional* consequences. Whereas *intended consequences* refer to those consequences that are the aim of our intention-for-the-future, *intentional consequences* have a wider ambit. For example, when I am intentionally chopping firewood, the intended consequences include only the fact that the firewood is being chopped, as this is the aim of the intention-for-the-future that caused the intentional action of chopping the firewood. On the other hand, the intentional consequences include the chopping of the firewood as well as the fact that I am blunting the axe through use. All intended consequences are intentional consequences, but not all intentional consequences are intended consequences.

In my view, intentional consequences include both intended and foreseen consequences, and an agent is morally responsible for both subsets. On the other hand, I think that unforeseen consequences never count as intentional, and are not necessarily subject to moral responsibility. My reason for suggesting that both intended and foreseen consequences should be considered subsets of intentional consequences stems from my definition of an intentional action. In my account of intention, an action is intentional if, and only if, it is performed with the belief that it was part of the fulfilment of an intention-for-the-future. On this definition, any aspect of my action that I intended or foresaw would be subject to this belief. Since the intended consequences are what my intention-for-the-future aims at, I must believe that they are part of the fulfilment of my intention-for-the-future. The case of foreseen consequences is a bit more complicated. Given the role of belief in my account of intention, I would suggest that a fair definition of “foreseen” would be: “believed to be likely.” Admittedly, the term “likely” is rather imprecise; however, I think it captures the meaning of “belief” in this context quite well. We can say that, when an agent holds the belief that the action they are performing is part of the fulfilment of an intention-for-the-future – this is, that performing the action is part of trying to realise the relevant pro-attitude – any consequence that she believes likely would be included in her belief about what she is trying to achieve.

For example: I intentionally bomb a building. My intention-for-the-future is to kill a dangerous terrorist. However, I foresee (believe it to be likely) that if I bomb the building, there will be civilian casualties. Is the consequence of my action – that there will be civilian casualties – an intentional consequence? Bombing the building was part of the fulfilment of my intention-for-the-future, and I chose to pursue this intention-for-the-future. This means that my decision to bomb the building was related to my practical reasoning. This practical reasoning must have accounted for all the consequences that I foresaw, and rationally compelled me toward the action with the combination of consequences that was adjudicated most rational, given my pre-existing beliefs. So, *all* the foreseen consequences of my action are in fact “part of the fulfilment of my intention to try to achieve my pro-attitude.” This being the case, those foreseen consequences would count as intentional given my definition of what counts as intentional action.

Another way of understanding this point is to recall Finkelstein’s (2005: 584) claim that for an action to be open to questions of moral responsibility it must be done *with a reason*. So, the question is whether foreseen consequences necessarily occur for a reason. I argue the answer is yes. This follows from the way that *Transmission Reasons* imparts reasons to both intention-for-the-future and intentional action. As these reasons stem from the rational judgements that result from the practical reasoning process, they include all that is accounted for by this process, which includes foreseen consequences. This means that the foreseen consequences occur for as much reason as the intended ones, since the entire set of consequences share the same set of reasons. Given these arguments, it is reasonable to say that the complete set of intentional consequences entails both intended and foreseen consequences. Furthermore, since intended and foreseen consequences are both subsumed by the category of intentional consequences, an agent is morally responsible for all the intended, as well as all the foreseen consequences of her intentional actions. However, for my claim to hold, I must be able to counter the argument developed by Knobe and Burra that it is, in fact, the ascription of moral responsibility that conditions the intentional status of an action, and not the other way around. I will briefly counter this criticism, before turning to the question of intentional pathological behaviour and, finally, the question of whether we can be held morally responsible for what we do unintentionally.

## 5.2 Countering Knobe and Burra's argument

Knobe and Burra (2006) argue that it is a mistake to assume that the ascription of moral responsibility requires knowledge about the intentional status of an action. They argue that their study of everyday ascriptions of responsibility, intention and intentionality, proves that ascriptions of intention actually depend on a preceding moral valuation. If this claim is correct, it would mean that moral responsibility does not depend on the intentional status of an action, but rather the reverse: that the intentional status of an action depends on its moral status.

However, there is actually quite a simple solution to this apparent difficulty. There is a difference between determining the moral value of an action, and ascribing moral responsibility to an agent for that action. There is also a difference between the *self-ascription* of intention and the *ascription* of intention. The first distinction allows us to see that an observer will usually first adjudicate a given agent's action (in a given context) to be moral or immoral, based on whatever ethical scheme they employ. This does not, in of itself, indicate that the agent who performed the action should be considered morally responsible for that action. It is *this* secondary question concerning the ascription of moral responsibility to an agent, which requires an investigation of the action's intentional status. What is more, Knobe and Burra's study does not acknowledge that we are often mistaken in ascribing intention to others. This means that we cannot accept the results of their experimental study uncritically. After all, given the amount of conceptual confusion involving intention, it is quite possible that the results of Knobe and Burra's study only reproduce this confusion, rather than providing for its resolution. It is for this reason that Anscombe (1963: 9) claims that the final authority for the accuracy of ascriptions of intention lies with the agent. If we accept this – as I think we should – then we can conclude that a more accurate picture of intention might be found by studying *self-ascriptions* of intention. Indeed, given my definition of intention-for-the-future as entailing an *intention-unique belief*, it is impossible for an agent to have an intention-for-the-future without also having a self-ascription of intention.<sup>88</sup> Since intentional action requires an intention-for-the-future, any intentional action entails a self-ascription of intention. To my mind, Knobe and Burra's study can provide us

---

<sup>88</sup> It is possible, however, that there might be cases of where an agent self-ascribes an intention, but does not in fact have an intention. This could occur if the agent mistakenly believed that she had a pro-attitude X, when in fact she did not have such a pro-attitude. See my discussion of the case of Lady A in Chapter 5, Section 2.2.

insight only into peoples' intuitions regarding the ascription of intention to others. It may be that such ascriptions of intention are preceded by moral evaluations, but the moral evaluations obviously cannot precede the intentional action itself, and so cannot precede the self-ascription of intention, which must already be present.<sup>89</sup>

### *5.3 Intentional pathological behaviour and moral responsibility*

A further difficulty to address here is moral responsibility in cases of pathological behaviour. We have seen that Davidson's account of intention, in particular, is unable to resolve this question of moral responsibility. This is largely due to the fact that he understands intention to be an unconditional judgement. I suggest that an account of intention that does not depend on intention being an unconditional judgement is better placed to distinguish between cases of pathological behaviour that are subject to moral judgement, and those that are not. I would argue that, in cases where the pro-attitude involved is not open to revision, or the practical reasoning process related to the intention is so epistemically opaque that it cannot be reflexively reconstrued without the addition of external knowledge, the action is not subject to moral responsibility, precisely because the action cannot be considered intentional. On the other hand, any pathological behaviour that didn't violate one of those two requirements would count as intentional, and would thus be subject to moral responsibility.

This can be most clearly shown with two examples. In the case of a pathological liar, who is genuinely incapable of restraining the compulsion to lie, the act of lying would *not* meet the criteria for intentional behaviour, precisely because the pro-attitude is not revisable. Similarly, if we consider the case of Lady A, from Galgut's arguments, it is quite obvious that in this case the reasoning process that led Lady A to leap from her carriage was epistemically opaque to her in such a way that she could not reconstruct it without external knowledge. However, this does not rule out the possibility that some (or indeed most) pathological behaviour does meet the requirements I have set for intentional action, and any action found to be so, would be an action for which the agent is morally responsible. All of this hinges on the idea that, when I say that the pathological liar in this example is incapable of not lying, this is true. If this individual could actually choose not to lie (for whatever reason), then this would no longer be a case of unintentional action.

---

<sup>89</sup> For an interesting discussion of the development of the ability to self-ascribe intention, and the relationship this has to important concepts such as obligation and self-control, see Olson's 2007 *Self-Ascription of Intention: Responsibility, Obligation and Self-Control*.

#### 5.4 Are we morally responsible for anything we do unintentionally?

The final point I discuss in this concluding section of the chapter is whether we can be held morally responsible for anything we do unintentionally. This question leads us beyond the scope of an account of intention. However, I think it important not to give the impression that I am arguing that an agent bears no moral responsibility for her or his unintentional actions. My (admittedly tentative and provisional) solution to the problem of unintentional action is, perhaps, the most novel aspect of my thesis. I suggest that an agent is *never morally responsible* for unintentional actions that she could not rationally foresee. However, if an agent could rationally foresee a consequence, then this consequence is subject to moral responsibility, but not in the same way that we would hold someone responsible for intentional consequences. How and why moral responsibility applies to some cases of rationally foreseeable action, are questions which I will not attempt to answer in this thesis. Rather, I wish to argue that for some rationally foreseeable (but unforeseen) consequences, it is not moral responsibility that applies, but what I will call *corrigible responsibility*.<sup>90</sup>

For some rationally foreseeable consequences it would be inappropriate, even if the consequence causes harm (be it to humans, animals or property), to consider the agent responsible for having acted immorally. Yet in such cases we often do hold agents responsible for the unintentional, but rationally foreseeable, consequences of their actions. I suggest that what is at stake here is not *moral* responsibility – in the sense of holding the agent to have acted immorally – but some other kind of responsibility. When, for example, someone sets out a target to throw a ball at, but due to emotional distress at the recent death of a loved one fails to foresee that in throwing at this target it is likely that she will hit a person sitting in their living room behind the target, I think it would be very unusual to say that the agent acted *immorally*.<sup>91</sup> However, despite the inapplicability of moral responsibility, it would be even more unusual not to hold the agent responsible in some way for striking the person with the ball. It is my claim that the responsibility at stake in cases such as this is corrigible responsibility. The difference between the cases of rationally foreseeable, but

---

<sup>90</sup> This position is quite close to that of Anscombe. In Chapter 2, Section 5, I discussed her argument that we can ascribe responsibility to an agent either because her action was intentional, or if the action was unintentional but the consequences foreseen, based on the question of “what you do it for.” However, my own account diverges strongly from that of Anscombe, in so far as I consider foreseen consequences to belong to the category of intentional consequences. The result of this is that my account has no need for a notion such as “what you do it for.”

<sup>91</sup> The agent acted foolishly, perhaps, and certainly carelessly, but not immorally.

unforeseen, consequences of intentional action where moral responsibility applies, and cases where corrigible responsibility applies, is the answer to the question, “Why did you not foresee the consequence?” If the answer to this question is that the agent failed to foresee the consequence for a reason that we consider morally inappropriate (such as overweening arrogance, or contempt of others), then moral responsibility applies. However, in cases where the lack of foresight results from reasons that are not morally blameworthy (such as being disorientated by panic), then corrigible responsibility would apply. Corrigibility, then, can be accurately attributed to an action the consequence of which is harmful and could be considered *rationaly foreseeable* for the agent in question, but where it would be unusual to call the agent’s reasons for failing to predict the consequence immoral. The ascription of this kind of responsibility is aimed at correcting behaviour, rather than ascribing blame. It is, in effect, about preventing the repetition of inadequate rational foresight by the agent, when such a lack of foresight has led to harm. Any punishment meted out should serve only to ensure that the agent will not fail in future to foresee the relevant, harmful, rationally foreseeable consequences of her or his intentional action. The easiest way to understand corrigible responsibility is probably by means of another example. Consider the following: I am an amateur mountaineer. During a climb I become stuck in a snowdrift. I have a signalling flare, and fire it in order to signal for help. Unforeseen by me due to panic (but rationally foreseeable nonetheless), this flare causes an avalanche that destroys a town. In this case, the avalanche was an unforeseen (hence unintentional) consequence of my intentional action in firing the flare. Furthermore, I do not think that this is a case of a foreseeable consequence where I would be *morally responsible* for this consequence. However, the action could (perhaps should) be deemed as corrigible. This means that I could be held responsible, and *some* punishment would be appropriate. The purpose of this punishment, however, would be to ensure that the harm caused by my action is not repeated, rather than to punish me for immorality. It is possible that the best way of achieving this aim might be to demand that I pay a fine, or do community work, or any variety of options. The point of the punishment, though, would be to discourage the corrigible behaviour in question – in myself and in others.

In the same vein, it may well be the case that many instances of pathological behaviour – behaviour I would deem unintentional – would very likely be open to corrigible, rather than moral, responsibility. This follows quite naturally if we assume – as I do – that pathology is not usually a morally inappropriate reason for failing to predict the consequence of one’s actions. The case of the pathological liar is a clear example of this. The consequence of this

individual's actions could have been rationally foreseen by the agent. However, it seems to me that this would not be a case where the agent would be considered immoral for acting as she did. For this reason, if the agent's actions resulted in harm, she should be deemed corrigible rather than immoral. Recall that corrigibility as I use it here does not require that the agent be able to alter their behaviour at the time of acting. Rather, it means that future instances of such behaviour can be prevented. Thus, cases of pathological behaviour are not immune to corrigible responsibility, even if they are examples of failures of effective self-government. What corrigible responsibility would require, in cases such as the pathological liar, is that the agent be held accountable in such a way as to prevent future instances of the harmful action. This will, presumably, mean that the place to start is the treatment of the agent's pathology.

In terms of my proposed account of intention, (12) agents are morally responsible for all the intentional consequences of their actions – which includes intended and foreseen consequences – and (13) sometimes corrigibly responsible for unintentional (though foreseeable) consequences of their actions. My argument in this regard has been aided by the specific definitions of intention-for-the-future and intentional action and their mutual relationship that I have developed earlier on in this chapter. I have further argued in favour of differentiating the moral evaluation of an action from the ascription of moral responsibility to an agent, and have argued that everyday intuitions about the ascriptions of intention might (at least sometimes) be a poor guide to understanding the concept of intention. Finally, I have argued that, while we do not bear moral responsibility for consequences that we could not rationally foresee, rationally foreseeable consequences can still be subject to moral *or* corrigible responsibility. It is my assertion that harmful consequences that are rationally foreseeable by the agent, but not open to moral responsibility (for whatever reasons), qualifies as behaviour that is corrigible, meaning that there is justification in such cases to punish an individual for the unforeseen (and so unintentional) consequences of their intentional actions. However, this punishment would not entail an ascription of immorality, and would serve the purpose of rectifying behaviour in order to avoid it in future. I think this is fertile ground for continued discussion.

## Concluding Remarks

It has been my aim in this thesis to analyse the three most influential theories of intention with a view to determining their cogency. To this end, I have identified five criteria that any account of intention must fulfil in order to be considered convincing. Having first assessed Anscombe's, Davidson's and Bratman's theories of intention in light of these criteria, I then tried to develop an account of intention that would fulfil all the specified criteria, without falling prey to the various difficulties that have plagued previous attempts to do so. According to this alternative account, the three seemingly irreconcilable uses of intention can be united in intention-for-the-future. Intention-for-the-future should be understood, in turn, as a composite mental entity consisting of a certain kind of pro-attitude and a certain kind of belief, and this entity is an effective cause. I have further provided an explanation for the link between intention-for-the-future and intentional action, which I believe overcomes the hurdle of causal deviance, as well as overcoming the danger of bootstrapping in practical reasoning. To achieve this, I have shown that, while an explanation of an intention is a rational explanation, intention itself is not a reason. Moreover, intention is not the outcome of a process of practical reasoning, but rather is guided by the deliberative procedures – among them, but not necessarily confined to, means-end coherence and intention-belief consistency – that are themselves the outcome of such reasoning. Despite my insistence on belief as a necessary part of intention, I have attempted to situate my account outside the general scope of cognitivism, by showing that means-end incoherence regarding an intention does *not* have to be understood as belief-incoherence, and therefore does not entail the reduction of practical reasoning to theoretical reasoning. Finally, I have tried to show that there is a necessary relationship between intention and moral responsibility. An agent is morally responsible for everything she does intentionally, where this includes both intended and foreseen consequences. Furthermore, I discussed what I believe is the possibility for a novel way of understanding how we hold agents to account for some actions that are unintentional, but where moral responsibility does not apply, namely: the concept of corrigible responsibility.

Although it is possible that there are more requirements for a convincing account of intention beyond the ones I have addressed in this thesis, I have tried to show that meeting these five requirements is at least a necessary, if not sufficient, condition for developing a credible theory of intention. If nothing else, meeting these five requirements, and overcoming the difficulties that doing so entails, allows us to clear up some of the conceptual darkness that

Anscombe identified more than half-a-century ago. This conceptual clarification has far-reaching consequences, not only for how we think about intention in general, but also for action theory, the relationship between causes and reasons, and for our understanding of the role of practical reasoning. Most importantly, perhaps, a clearer understanding of the relationship between intention and (moral) responsibility has potentially significant implications for how we think our society ought to operate. It is with a view towards making a contribution – however preliminary – towards solving some of these difficult questions that I have tried to map out a convincing account of intention in the course of this thesis. That, at any rate, has been my *intention*.

## BIBLIOGRAPHY

Adams, F. 2006. Intentions Confer Intentionality upon Actions: A Reply to Knobe and Burra. *Journal of Cognition & Culture*. 6: 255-268.

Anscombe, G.E.M. 1957. Intention. *Proceedings of the Aristotelian Society*. 57: 321-332.

Anscombe, G.E.M. 1963. *Intention* (second edition). United States of America: Harvard University Press.

Anscombe, G.E.M. 1961. "War and Murder," in Stein, W. (ed.) 1961. *Nuclear Weapons: a Catholic Response*. London: Merlin: 43–62.

Anscombe, G.E.M. 1974. "Practical Inference" in Geach, M. and Gormally, L. (eds.) 2005. *Human Life, Action and Ethics: Essays by G.E.M. Anscombe*. Imprint Academic.

Anscombe, G.E.M. 1982. Medalist's Address: Action, Intention and 'Double Effect'. *Proceedings of the American Catholic Philosophical Association*. 56: 12-25.

Bayne, S.R. 2010. Elizabeth Anscombe's Intention. U.S.A.: Booksurge Publishing.

Bratman, M. 1984. Two Faces of Intention. *The Philosophical Review*. 93(3): 375-405.

Bratman, M. 1985. "Davidson's Theory of Intention" in *Faces of Intention*. 1999. Cambridge: Cambridge University Press.

Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, M. 2009a. "Intention, belief, and instrumental rationality" in Sobel, D. and Wall, S (eds.) 2009. *Reasons for Action*. Cambridge: Cambridge University Press.

Bratman, M. 2009b. Intention, Practical Rationality, and Self-Governance. *Ethics*. 119(3): 411-443.

- Bratman, M. 2013. The Interplay of Intention and Reason. *Ethics*. 123(4): 657-672.
- Brentano, F. [1873] 2009. *Psychology from an Empirical Standpoint*. London: Routledge and Kegan Paul.
- Broome, J. 2001. "Are intentions reasons? And how should we cope with incommensurable values?" in Morris, C and Ripstein, A (eds.) 2001. *Practical Rationality and Preference: Essays for David Gauthier*. Cambridge: Cambridge University Press.
- Chan, D.K. 2000. Intention and Responsibility in Double Effect Cases. *Ethical Theory and Moral Practice*. 3(4): 405-434.
- Davidson, D. 2001. *Essays on Actions and Events*. Oxford: Clarendon Press.
- Donnellan, K. and Morgenbesser, S. 1963. Knowing What I am Doing. *The Journal of Philosophy*. 60(14): 401-409.
- Driver, J. "Gertrude Elizabeth Margaret Anscombe" in Zalta, E.N. (ed.) 2011. *The Stanford Encyclopedia of Philosophy*. <<http://plato.stanford.edu/archives/win2011/entries/anscombe/>>.
- Finkelstein, C. 2005. Responsibility for Unintended Consequences. *Ohio State Journal of Criminal Law*. 2: 579-599.
- Galgut, E. 2011. Acting on Phantasy and Acting on Desire. *South African Journal of Philosophy*. 30(2): 132-142.
- Gomes, G. 2002. The interpretation of Libet's results on the timing of conscious events: A commentary. *Consciousness and Cognition*. 11(2): 221-230.
- Gormally, L. "G E M Anscombe (1919-2001)" in Coulter, M.L., Myers, R.S. and Varacalli, J.A. (eds.) 2012. *Encyclopedia of Catholic Social Thought, Social Science and Social Policy*. Lanham: Scarecrow Press.
- Grice, H.P. 1971. Intention and Uncertainty. *Proceedings of the British Academy*. 5: 263–279.

- Haggard, P. and Libet, B. 2001. Conscious Intention and Brain Activity. *Journal of Consciousness Studies*. 8(11): 47–63.
- Harman, G. 1976. “Practical Reasoning” in Mele, A. (ed.) 1997. *The Philosophy of Action*. Oxford: Oxford University Press.
- Harman, G. 1986. “Willing and Intending” in Philosophical Grounds of Rationality, Grandy, R.E. and Warner R. (eds.) 1986. *Philosophical Grounds of Rationality*. Oxford: Oxford University Press.
- Harman, G. 2006. Intending, Intention, Intent, Intentional Action, and Acting Intentionally: Comments on Knobe and Burra. *Journal of Cognition and Culture*. 6: 269-276.
- Husserl, E. [1900/1901] 2001. *Logical Investigations*. London: Routledge and Kegan Paul.
- Jay, W. R. 2009. “Practical Reason” in Zalta, E.N. (ed.) 2009. The Stanford Encyclopedia of Philosophy. < <http://plato.stanford.edu/archives/sum2009/entries/practical-reason/>>.
- Kant, I. [1788] 1996. *Critique of Practical Reason*. New York: Prometheus Books.
- Knobe, J. 2004. Intention, Intentional Action and Moral Considerations. *Analysis*. 64(2): 181-187.
- Knobe, J. and Burra, A. 2006. The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study. *Journal of Cognition and Culture*. 6(1-2): 113-132.
- Libet, B., Gleason, C.A., Wright, E.W. and Pearl, D.K. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain*. 102: 623–42.
- Mackie, J. L. 1974. “Comment: Practical vs. Theoretical Reason” in Korner, S. (ed.) 1974. *Practical Reason*. Great Britain: The Camelot Press.

Mail&Guardian. 2013. "Oscar Pistorius to face premeditated murder charges in March." <<http://mg.co.za/article/2013-08-19-oscar-pistorius-trial-set-for-march-3>> (Retrieved: 29 October 2013).

McCann, H. J. 1991. Settled objectives and rational constraints. *American Philosophical Quarterly*. 28: 24-36.

McCann, H. J. 2005. Intentional action and intending: Recent empirical studies. *Philosophical Psychology*. 18(6): 737-748.

Moran, D. 2013. Intentionality: Some Lessons from the History of the Problem from Brentano to the Present. *International Journal of Philosophical Studies*. 21(3): 317-358.

Olson, D.R. 2007. Self-Ascription of Intention: Responsibility, Obligation and Self-Control. *Synthese*. 159(2): 297-314.

Russell, B. 1912. On the Notion of Cause. *Proceedings of the Aristotelian Society*. 7: 1-26.

SABC. 2013. "Pinetown truck driver faces 22 murder charges." <<http://www.sabc.co.za/news/a/acc97b8041086672a14fa1434f2981a1/Pinetown-truck-driver-faces-22-murder-charges-20130909>> (Retrieved: 29 October 2013).

Searle, J. 1980. The Intentionality of Intention and Action. *Cognitive Science*. 1(4): 47-70.

Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Setiya, K. 2007. Cognitivism about Instrumental Reason. *Ethics*. 117: 649–673.

Setiya, K. 2010. "Knowledge of Intention" in Ford, A., Hornsby, J. and Stoutland, F. (eds.) 2011. *Essays on Anscombe's Intention*. Harvard University Press.

Setiya, K. 2011a. "Intention" in Zalta, E.N. (ed.) 2011. *The Stanford Encyclopedia of Philosophy*. <<http://plato.stanford.edu/archives/spr2011/entries/intention/>>.

The New York Times. 2013. “‘Ugly Thoughts’ Defense Fails as Officer Is Convicted in Cannibal Plot.” <<http://www.nytimes.com/2013/03/13/nyregion/gilberto-valle-is-found-guilty-in-cannibal-case.html>> (Retrieved: 29 October 2013).

Thompson, D. L. 1986. Intentionality and Causality in John Searle. *Canadian Journal of Philosophy*. 16(1): 83-97.

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

Velleman, J.D. 1989. *Practical Reflection*. Princeton: Princeton University Press.

Velleman, J.D. 1991. Intention, Plans, and Practical Reason by Michael E. Bratman. *The Philosophical Review*. 10(2): 277-288.

Wilson, G. and Shpall, S. “Action” in Zalta, E.N. (ed.) 2012. *The Stanford Encyclopedia of Philosophy*. <<http://plato.stanford.edu/archives/sum2012/entries/action/>>.

Wittgenstein, L. 1961. *Tractatus Logico-Philosophicus*. Great Britain: Routledge Classics.