Intentions in tension: Is there more to intentional action than just belief and desire?

Pekka Tolvanen

Master's Thesis

UNIVERSITY OF HELSINKI

Faculty of Social Sciences

Practical Philosophy

Helsinki, November 11, 2019

Tiedekunta – Fakultet – Faculty Faculty of Social Sciences	Koulutusohjelma – Utbildningsprogram – Degree Programme
Tekijä – Författare – Author Pekka Tolvanen	
Työn nimi – Arbetets titel – Title Intentions in tension: Is there more to intentional action than just belief and desire?	
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Practical Philosophy	

Työn laji – Arbetets art – Level Master's Thesis	Aika – Datum – Month and year November 2019	Sivumäärä – Sidoantal – Number of pages 60

Tiivistelmä – Referat – Abstract

Understanding intending is crucial to the understanding of purposeful human action. In the philosophy of action beliefs and desires are usually taken to be the necessary conditions of intending. The disagreement over how intentions specifically are related to beliefs and desires, is often put in terms of whether intentions are independent mental states or not. Belief-desire accounts of intending don't feature intentions as independent mental states, whereas belief-desire-intention accounts of intending do.

The goal of most accounts of intentional action is to account for three senses of intentionality: intentional action, intention-with-which and intending. Intentional action and intention-with-which are usually taken to be explicable in terms of belief and desire. Thus the focus of this thesis is on intending.

This thesis aims at providing a more comprehensible picture of the kinds of arguments that have been given for and against the reducibility of intentions. It also provides an overview to reductivist belief-desire accounts and nonreductivist belief-desire-intention accounts and a tentative classification of arguments against reduction. Finally a recent Humean reductivist belief-desire account of intending is explored more thoroughly.

Avainsanat – Nyckelord – Keywords philosophy of action, intention, belief, desire
Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors Raul Hakli
Säilytyspaikka – Förvaringställe – Where deposited
Muita tietoja – Övriga uppgifter – Additional information

Contents

1 Introduction

This thesis is about intentions. One might intend to open a door, write a thesis or be an astronaut when one grows up. Intentions and intentional action define a big part of our human existence as through intentional action we have control over our lives instead of merely having things happen to us, reflexively responding to the environment around us or acting inadvertently.

The concept of intention has many practical and theoretical applications: intentions guide our everyday life as we plan our own actions; intentions help us understand other people and coordinate with them; intentions are crucial to many strands of moral philosophy, psychology and law; and intentions are often used in modelling plans and coordination in multiple fields from strategy to artificial intelligence.

The philosophical debate on intentions is home to several interrelated discussions. The one I will be addressing concerns the reducibility of intentions. The discussion is usually divided into two attitudes toward the question: some argue that intentions are distinct, irreducible mental states, while others identify intentions with other mental states, usually combinations of desires and beliefs. One way to put the question is this: is our way of talking about intentions just a shorthand for something else or not?

While a lot has been written about intentions, the arguments given for and against the reducibility of intentions have rarely been in the spotlight, but dispersed amidst the broader discussion. The aim of this thesis is to give a more comprehensible picture of the kinds of arguments that have been given for and against the reducibility of intentions. To accomplish this, I will also introduce some of the more popular reductionist and non-reductionist accounts of intention.

This overview on the discussion of the reductivity of intentions will set the stage for a more thorough treatment of Neil Sinhababu's recent neo-Humean account that aims to reduce the state of intending into belief-desire combinations in a way that has both old and novel elements. Finally I will defend Sinhababu's account against some of the historical arguments and offer my own critique that concerns Sinhababu's use of a counterfactual background condition.

I will begin by introducing the key concepts at hand: in chapter 2 I will introduce the many faces of intention and intentionality; mental states and folk psychology; action, agency and practical reasoning; desires and beliefs — the mental states at the heart of most reductive attempts and say a few words about reduction. While this thesis is not a definitive list of accounts of intending, I believe that the chapters 3 to 5 provide a representative selection of rigorous reductive and nonreductive arguments. In chapter 3 I will go through some of the most important early reductive belief-

desire theorists with substantive reductive accounts. In chapter 4 I will introduce a number of nonreductive accounts and their arguments mostly directed against the early reductionists of chapter 3. Next I will turn to recent neo-Humean attempts at reduction in chapter 5 that attempt to provide answers to the nonreductivist critiques. The discussion will continue in chapter 6 as I examine how the Humean theory of intending fares with the problems posed by nonreductivist critiques and present some of my own observations that mostly have to do with the novel elements in his account of intending. Finally in chapter 7 I will present my conclusions.

2 Intention in the philosophy of action

2.1 Intentions

First of all, what are intentions? According to Alfred Mele, an intention is a psychological state that "initiates and motivationally sustains intentional action; it guides and monitors behavior; it coordinates one's activities, including one's interaction with others; and it can both prompt and suitably terminate practical reasoning" (Mele 1992b, p. 145). This is a rough definition that of course raises many further questions. What is intentional action? What is motivation? What is practical reasoning? This introductory chapter will give brief answers to these questions. While incomplete at least without further elaboration, Mele's definition is still helpful in getting a first grasp of what kinds of properties are usually associated with intentions.

Many themes central to the modern discussion of intentions were first explicated in Elisabeth Anscombe's book *Intention* (1958). Anscombe noted that the word intention is used in at least three ways: to express intending (intending to Φ[1]), intentional actions (Φing intentionally) and the intention-with-which a thing is done (Φing with the intention of doing X). Anscombe's goal was to offer a unified account of intentions that spans all three aspects and her threefold distinction has formed the basis of the discussion on intending (Anscombe 1958, p. 1).

The relationship between the many different uses of intention is still contested and there is no universally accepted unificatory account. On Anscombe's threefold distinction I will mostly steer clear of intentional action and intention-with-which and focus on intending, which I take to be identical to having the mental state of intention — be it reducible or not. As most of the literature is not exclusively focused on the question of reductivity and many writers have had unificatory ambitions, it would be very hard if not outright impossible to completely abstain from mentioning

---

[1]Greek letters beginning from Φ will be used as variables to represent actions.

intentional action and intention-with-which later in this thesis. However, my focus will be on intentions as a mental state as they seem to be most crucial to the question of reductivity, i.e. whether or not intentions are independent mental states.

A few noteworthy distinctions should be made concerning intentions. First, intentions can be simple or complex. Simple intentions are intentions to Φ, whereas complex intentions are intentions to bring about a situation S by doing Φ. Second, the distinction between proximal and distal intentions. A proximal intention is "a propensity to execute a plan for immediate action" and a distal intention is a propensity to execute an intention-embedded plan for action in the nonimmediate future (Mele 1992b, p. 144). This propensity, i.e. tendency, is based on being settled upon a course of action or "upon making a full-blown attempt to A and not just . . . *starting* to A" (Mele 1992b, p. 145).

Another important idea from Elisabeth Anscombe is that actions can be intentional under one description and unintentional under another (Anscombe 1958, p. 29): to use an example from Donald Davidson, turning the light could be described as moving a finger, flicking the switch, turning the light on, illuminating the room and surprising a prowler one didn't know about. The first four descriptions are intentional presuming one believed the switch, the wiring and the light bulb to be functional, but the last one is unintentional (Davidson 1971, p. 16).

Since Donald Davidson's early work (1963) intentions have been seen either as irreducible mental states or as mental states that can be reduced to other mental states. The usual reductive view is that intentions are composed of beliefs and desires. The desire concerns an end, i.e. a goal of the agent and the belief usually concerns the means to that end. Intentions are also often associated with some version of belief about doing (see for example Setiya 2008) — the agent's belief that the intended action is certain, probable or possible — or, at the very minimum, the absence of a belief that the action is impossible.

The reductive and nonreductive accounts usually concern one of two kinds of intending: intending to Φ or a propositional attitude intending that Φ. The difference is that in the former case Φ usually refers to the agents action and in the latter case it can refer to states of affairs. Even though the relationship between the two expressions is much more complicated, for the sake of streamlining the discussion I will assume that they are close to interchangeable and discuss the accounts of intending as they are presented in the literature. I don't believe this should affect the arguments for or against reduction.

2.2   Mental states and folk psychology

The independent existence of intentions in an ontological sense does not hinge only upon whether or not they are distinct from beliefs and desires. When talking about all mental states in the philosophy of action, we are talking about commonsense folk psychological concepts defined with philosophical accuracy. There is ample discussion about the way folk psychological mental states might be said to exist if they exist at all.

Daniel Dennett's (Dennett 1991) moderate stance toward intentional explanations and folk psychological mental states is that the way we talk about beliefs and intentions might not necessarily have to match the way our minds work in order for the patterns of behavior, that we refer to with folk psychological concepts, to be real. I hope that this discussion can accommodate mental states at least in a Dennettian sense, and does not presuppose any specific theory of mental representation.

Of course if the language of folk psychology were misleading altogether in an eliminativist sense advocated by Paul Churchland (1981), it might pose problems for a serious discussion on intending. I won't be arguing for the necessity of folk psychological concepts here, as it does not fit the scope of this thesis. For the purposes of this thesis I would characterize these kinds of mental states as necessary representations at the very least, even when they do not accurately express the inner workings of the mind. I take the usual non-reductionist position to be that intentions are independent mental states in a similar way that beliefs and desires are and not take a stand how they might actually exist.

2.3   Practical reasoning

Intentions are often formed through practical reasoning. The idea of practical reasoning dates back to Aristotle and it consists of trying to answer questions such as "what to do?", rather than "what is the case?" The question is not ethical either in the sense of "what should I do", but a practical one. After we answer the question "what to do?" we either act intentionally or intend to act later. It is perhaps best illustrated by using an example using syllogistic reasoning, in other words logically deducing from a set of premises to a conclusion. In theoretical reasoning, both the premises and the conclusion are propositions, sentences that can be true or false and the objects of beliefs, like this:

All men are mortal.

Socrates is a man.

–

Socrates is mortal.

Instead, for Aristotle the conclusion of practical reasoning is not a proposition, but action:

> All sweet things ought to be tasted.
> Yonder thing is sweet.
> —
> Action
> (Aristotle 1934, VII.3.1147a25-30)

The same idea is still utilized, even though the concept has become much more refined. The action should be intentional at least under some description as it is the conclusion of practical reasoning. As agency is our capacity to act, those who act are usually called agents. The conclusion of practical reasoning need not be immediate action, but it can also be an intention. This is one way of differentiating propositional value judgements such as beliefs from judgements that lead to action such as intentions. Whereas theoretical reasoning modifies our beliefs, practical reasoning modifies our intentions.

What then are the premises of practical reasoning? Already in Aristotle's thought we can see the familiar set of desires and beliefs as the premises of practical reasoning, the conclusion of which is immediate action: "I want to drink, says appetite; this is drink, says sense or imagination or mind: straigthway I drink" (Aristotle, 701a30). As the premises of practical reasoning rationalize our actions, they can also be called reasons for those actions.

Elizabeth Anscombe famously noted that reasons generally answer the question "why?" in the context of intentional action (Anscombe 1958, p. 9). Reasons were especially relevant to the early reductivity debate, as reasons-explanations have been the goal of many early reductive belief-desire models that equated reasons with beliefs and desires (Mele 1992b, p. 122). However, it is now debated if reasons should be identified with the states of beliefs and desires themselves (e.g. the belief that the sun will shine tomorrow) or their contents (e.g. the sun will shine tomorrow) (Mele 1992b, p. 115).

2.4 Desires and beliefs

Let us turn our gaze toward desires and beliefs. These mental states have had a number of different names over the years. For the most part this seems to be a matter of convention, but the changing terminology should be taken into account. I will try to balance between doing justice to the literature and not confusing the reader by mostly using the terminology of desires and beliefs. When talking about

the views of specific authors I will be using their terminology to avoid errors in conceptual translation.

Beliefs are most often taken to be propositional attitudes, that is, mental states whose intentional content can be expressed by propositional sentences, usually deeming the proposition true or likely from the agent's perspective. They are often distinguished from desires and intentions by their direction of fit: whereas desires and intentions are mental states that the world must fit, beliefs are mental states that must fit the world. Elisabeth Anscombe (1958, p. 56) was one of the first to describe this distinction in detail and it was aptly named by John Searle (1975, p. 346–347). In Anscombe's example direction of fit is illustrated by a detective following a shopping man around town and keeping a record of everything the man buys. The man's shopping list has a word-to-world-fit, which means that the list aims at realization: to succeed the man must change the world to match his list by buying the things in the list. In contrast, the detective's record has a world-to-word direction of fit and the record aims at truth: to succeed, the detective must change their record to match what happens in the world. The same arguably goes for mental states: the shopping list provides the content of desires or intentions, while the detective's record provides the propositional content of beliefs.

In the philosophical discussion the word desire differs somewhat from how it is used in common language. In common language, the word desire often has dramatic connotations. We might describe our feelings as desires when it comes to pleasure or knowledge, but rarely when we're talking about things we don't like or are indifferent towards, even if we eventually want to do them, such as cleaning the toilet or eating tasteless food.

In the philosophical sense a distinction is often made between appetitive desires and volitive desires (Davis 1984, p. 45). The appetitive sense of desire includes connotations of appetite, urge and craving, whereas the volitive sense is more bland kind of wanting where desires are seen as reasons for action in practical reasoning. The philosophical discussion on intention is focused on desires in the volitive sense, where the appetitive connotations are not included.

Besides the focus on the volitional aspect of desire, there is no general consensus on what desires are. In this discussion the concept of desire ranges from a thin dispositional conception like that of Michael Ridge (1998) to a thick conception of desire that emphasizes and utilizes different properties of desire like that of Neil Sinhababu (2017).

The terminology isn't wholly shared either. The mental states that are most often called desires (e.g. Brand 1984, Davis 1984, Harman 1986, McCann 1986, Bratman 1987, Ridge 1998, Setiya 2007, Sinhababu 2017) have also been called pro

attitudes (e.g. Davidson 1963) and wants (e.g. Davidson 1963; Audi 1973). Desire seems to be most widespread term in the recent literature on intending, so for the sake of clarity I will use desire to mark the relevant pro attitude. Pro attitudes are positive mental attitudes directed toward actions under a certain description. These include for example desires, wants, urges, wishes, obligations and intentions.

Most of these are also called desires in the broad volitive sense (Davidson 1963, p. 688), which means desires, wants and pro attitudes have be used almost interchangeably. Whether or not desires in this broad volitive sense also include intentions is of course controversial as it is linked to the the question of reductivity. Some theorists also draw a distinction between desires and aversions, where aversions are usually seen as just negative desires and largely ignored. Like most others, I will be using the term desire to refer to both desires and aversions.

Some accounts depict beliefs and desires dispositionally, notably Michal Ridge's account of intending in chapter 5 that is inspired by Michael Smith's (1995) well known account of Humean motivation. Dispositions also feature prominently in Michael Bratman's (1987) characterization of commitment in chapter 4. Neither Ridge or Bratman explicitly explains what they mean by dispositions. A disposition to Φ doesn't mean that one should have a mental attitude towards Φ, that one should at some point actually Φ nor that one couldn't do the negation of Φ. Rather having a disposition to Φ or being disposed to Φ means roughly that one has a tendency to Φ.

2.5 Reduction

When can a reduction of intentions to beliefs and desires called successful? The crux of the matter is whether or not intentions share the same kind of independence as mental states as beliefs and desires do. Showing that beliefs and desires are involved in intending is not enough to provide a reductivist account, as this idea is prevalent among the nonreductionists. To succeed in reduction, one should be able to give an account of human action substituting the concept of intention with the concepts of belief and desire.

One should keep separate the following questions: if intentions causally follow from beliefs and desires only, and if intentions are beliefs and desires only? In the former case the intentions could plausibly continue independent existence even if the desires and beliefs that gave them birth are abandoned or forgotten. In the latter case the intentions disappear when the beliefs and desires — or possibly just some of them — disappear or decrease sufficiently in strength.

It's also useful to note that there are multiple ways to frame reducibility and irre-

ducibility. Michael Ridge distinguishes between weaker and stronger versions of the irreducibility thesis (Ridge 1998, p. 157). The weak version argues that intentions are not fully constituted by beliefs and desires, whereas the strong version argues that intentions are not even partially constituted by beliefs and desires. Another kind of a distinction could be made between the irreducibility of all intention in one end of the spectrum and the possibility of irreducibility in some fringe case of intention in the other end. In between there is space for many possible positions that could posit irreducibility for some types of intentions only — for example irreducibility in all cases of future intentions. To my knowledge this distinction hasn't been utilized in the literature.

I take the reductionist thesis to be roughly that intentions don't exist at all in human psychology as independent mental states and the minimal nonreductionist position to be the reverse of this, that there are at least marginal cases of independent intentions. The actual nonreductionist stance is usually somewhat stronger: that independent intentions are commonplace in human action and integral to us behaving as we do.

It seems that reductive accounts need to bridge at least two important gaps to succeed. First, the relationship between beliefs and desires about some action Φ and the actual action Φ needs to match the relationship between the intention to Φ and the action Φ. Second, we will come to see later that beliefs and desires need to be intimately connected to reductively explain phenomena usually associated with intentions. Searching for an ever tighter connection will reach its peak in the recent Humean reductive account of Neil Sinhababu (2017).

In the next three chapters I will be looking at some influential models of intention suggested in the literature and some of the most common arguments for and against the reduction of intentions. Broadly speaking most nonreductive theorists give two kinds of arguments: first, they emphasize certain properties of intentions and posit that reductive models cannot explain them; second, they attempt to give counterexamples to certain belief-desire accounts. The counterexamples usually aim to show that sometimes we can intend without having the relevant predominant desires and beliefs, in other words, that it's possible to intend while fulfilling the conditions given by the reductive account. The most common argument from the reductive side is to give an exhaustive belief-desire account of intention as an attempt to to answer the challenges put forward by the nonreductivists.

3 Early belief-desire theorists

The systematic defense of reductive belief-desire theory began from Donald Davidson's attempt to defend reasons-based rationalizing explanations as a species of causal explanation against a number of scholars, including Elisabeth Anscombe (1958). For early Davidson (1963), beliefs and desires are the primary reasons of action that also cause the action. Rationalizing, in other words stating the reasons of action, is thus causally explaining action. Therefore, as stating beliefs and desires is sufficient, intention as an independent mental state is not needed for explaining action. Later Davidson settles for a nonreductive account after analyzing future-directed intending (Davidson 1978).

Donald Davidson's essays on action and intentionality have since inspired both reductive and nonreductive accounts. While Davidson later qualified his reductivist claim about intending and later developments proved reasons-based reductive accounts wanting, reductive accounts of intending inspired by Davidson enjoyed wide popularity for quite some time from the 1960s to the 1980s. Formulations of intending as sets of beliefs and desires were presented by Robert Audi (1973) and Wayne E. Davis (1984) and in this chapter I will introduce their views. These reductive belief-desire accounts of intending later sparked the first substantive discussions on whether or not a reduction of intending is warranted.

3.1 Donald Davidson: Three attempts at reduction

Donald Davidson's *Actions, Reasons and Causes* (1963) is the cornerstone of reductive belief-desire models of intention. Davidson's long-term goal was to provide a unificatory account of the three Anscombean facets of intentionality: intentional action, intention-with-which and intending. In *Actions, Reasons and Causes* Davidson attempts to show that at least one facet of intention can be explained causally through its reasons. For Davidson, reasons are combinations of pro attitudes toward actions and beliefs regarding actions. Explaining intentional action by its reasons is a rationalization in which the reasons rationalize the action (Davidson 1963, p. 685).

He discusses in length different ways in which intentional action is described in everyday interactions while answering several why-questions to show how rationalizing action works. At the same time he weaves together a theory of intentional actions as actions that are done because of reasons, the "because" implying causality. On top of connecting reasons with intentional action, Davidson also manages to connect reasons with intention-with-which, another Anscombean facet of intentionality.

A major idea in the essay, although already presented by Anscombe (1958, p. 11), is that an action can be described in several ways that refer to separate events. Ac-

cording to Davidson, an action can be intentional under one description and unintentional under another description. To cite Davidson's example, if I intentionally flip a switch, turn on the light, illuminate a room and unintentionally (unknowingly) alert a prowler, I haven't done four different actions, but rather one action that can be described in four ways, some of which are intentional (Davidson 1963, p. 686–687). This kind of an action individuation is a contested issue (for a contrasting account cf. Goldman 1976).

Adding Davidson's account of action descriptions outlined above to his theory of reasons as pro attitudes and beliefs, Davidson ends up giving a necessary condition for primary reasons of intentional actions:

> C1. R is a primary reason why an agent performed the action A under the description d only if R consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that A, under the description d, has that property.
> (Davidson 1963, p. 687)

And a causal augmentation to C1:

> C2. A primary reason for an action is its cause.
> (Davidson 1963, p. 693)

In other words, the agent's primary reason for an action consists of a pro attitude towards actions with some property and the belief that the action at hand has that property. Furthermore, the belief and the desire cause the action. To know them is to know "an intention with which the action was done" (Davidson 1963, p. 689). Davidson doesn't discuss in detail if there could be other factors relevant to explaining intentional action in addition to primary reasons. He does nevertheless state his belief that a modified C2 condition could make both C1 and C2 sufficient as well as necessary conditions for reasons that explain actions (Davidson 1963, p. 693). At this point Davidson also believed that intention-with-which was the central concept in the analysis of intention that will translate to intentional action and intending (Davidson 1980, xiii). This was Davidson's attempt at reduction.

For reductive explanation to be possible Davidson (1963) also seems to presuppose knowledge of the causal connection between the primary reason (belief-desire pair) and the action. He criticizes A. I. Melden for disregarding causal explanations and consequently the connection between the reasons and the action (Davidson 1963, p. 692–693). Davidson argues against Melden that merely the fact that the agent has reasons for action isn't enough to explain action unless the reasons can be shown to be the agent's reasons for the particular action that the agent did (Davidson 1963,

p. 692). Even if the agent has reasons for an action under a certain description, the action can be unintentional if the agent acts for some other reasons. For example a driver raising their arm might have reasons to signal, but if they raised their arm to stretch or the action was caused by a causal anomaly, there would be no intentional signaling (Davidson 1963, p. 693). This is why the connection between the agent's reasons for action and the causal reasons of action is so important for Davidson's account.

In a later essay Davidson develops his idea to also spell out the relation between wanting and intentional action:

> P1. If an agent wants to do x more than he wants to do y and he believes himself free to do either x or y, then he will intentionally do x if he does either x or y intentionally.
>
> P2. If an agent judges that it would be better to do x than to do y, then he wants to do x more than he wants to do y.
>
> (Davidson 1970, p. 23)

In the second chapter I introduced Elisabeth Anscombe's distinction between the three guises of intention: intending (intending to Φ), intentional action (Φing intentionally) and the intention-with-which a thing is done (Φing with the intention of doing X). So far all of Davidson's analysis has been trying to show that two out of three Anscombean guises of intention — intentional action and intention-with-which — are analyzable in terms of beliefs and desires. Both intentional action and intention-with-which differ from the third Anscombean use of intention, intending, in that in the first two the action related to the intention has always already happened or is currently underway. In contrast, intending can relate to not only actions that are currently underway, but also actions that are merely planned. Davidson later found intending "without practical reasoning, action or consequence", which he calls "pure intending", to be more difficult to analyze as the action it refers to hasn't taken place and indeed may never take place (Davidson 1978, 83).

However, according to Davidson (1978, p. 88) including pure intending in the account is crucial because all intentional action that takes time must include something like pure intending. Because of the close connection between intending, intentional action and intention-with-which, the inability to describe one aspect of intending would cast a shadow of doubt on the account describing the other two.

Davidson (1978) acknowledges a further problem for his early account. In the case of intentional action or intention-with-which, the agent's (not necessarily conscious) judgements can be tied to concrete actions (Davidson 1978, p. 96–97). In the case of pure intending this is not the case: a pure intention "cannot single out a

particular action in an intelligible sense, since it is directed to the future" (Davidson 1978, p. 99).

Davidson eventually presents a nonreductive positive account of pure intentions as all-out judgements supplemented with "the rest of what I believe about the immediate future", especially with the belief that nothing will come up to make the prospective action all-out undesirable or impossible (Davidson 1978, p. 99). Coming to have an intention is to arrive at such an all-out judgement while holding aforementioned beliefs about the future (Davidson 1978, p. 99–100). The beliefs as background conditions are crucial as there is no reference action for pure intending as we already saw. Intentions are value judgements of the same genus as desires, but intentions are distinguished by their all-out, unconditional form (Davidson 1978, p. 102).

3.2   Robert Audi: A clear belief-desire formulation

Inspired by Davidson's early essay, Robert Audi attempts to reduce intending in terms of wanting and believing (1973). He offers a formulation for both complex (x intends to bring about Ø by doing A, where x is an agent, Ø is some state of affairs and A is some action) and simple (x intends to do A) intentions, but for our purposes the latter formulation is more illustrative ((x) and (A) are universal quantifiers: for all x and for all A) :

(x) (A) [x intends to do A if and only if

(1') x believes that he will (or that he probably will) do A; &

(2') x wants, and has not temporarily forgotten that he wants, to do A; &

(3') either x has no equally strong or stronger incompatible want (or set of incompatible wants whose combined strength is at least as great), or, if x does have such a want or set of wants, he has temporarily forgotten that he wants the object(s) in question, or does not believe he wants the object(s), or has temporarily forgotten his belief that he cannot both realize the object(s) and do A.]

(Audi 1973, p. 395–396)

Audi's account is a clear-cut reductive belief-desire account as it features only beliefs and desires. The intending agent has certain beliefs about their doing A (1) and wants (2) and on top of that certain incompatible wants are excluded (3). The format of the account is especially interesting. It is one of the earliest formulations

of intending that list the relevant conditions one by one in terms of desires, beliefs and background conditions.

What is missing from the account is the link between wants and beliefs. Following Audi's account, one could arguably intend to A even if one's belief that one will A is not based on one's want to A, but rather something that is not up to the agent. Audi gives an example in which x wants to insult y, believes that y will be insulted by x's involuntary grimace and has no incompatible wants (Audi 1973, p. 399). According to Audi's formulation of intending x intends to insult y. Audi seems to believe that this is not detrimental to the account, but rather a positive feature.

However, this attitude is not shared by all reductive theorists. Wayne E. Davis gives the following counterexample that seems to fit Audi's formulation as well, but it seems that the agent does not intend to do anything: "John was pushed out of a helicopter over a trout pond. He expects to scare away the fish when he hits the water. He wants to scare them, because his arch-enemy is fishing. But he could hardly intend to frighten them, and his doing so will not be intentional" (Davis 1984, p. 48).

The significance of this account is that in that most later belief-desire models of intending are formulated in a similar manner containing the desire, beliefs and background conditions. Audi's account was also one of the few well explicated reductive accounts left after Davidson abandoned his reductive project in the late 70s (Davidson 1978). In chapter 4 on nonreductive belief-desire-intention accounts of intending we will see how Alfred Mele (1988; 1992b) argued specifically against Audi's reductive account.

3.3   Wayne E. Davis: The connection condition

Wayne A. Davis (1984) offers another reductive belief-desire analysis of intending. Davis follows Donald Davidson in arguing that to be sufficient for intention, joint beliefs and desires need to be connected in the appropriate way. To accomplish this, Davis introduces a connection condition that supposedly marks the difference between intending and other pro attitudes that resemble intending (Davis 1984, p. 48).

Davis defines intending as follows, where S is the agent and p is any sentence in the appropriate grammatical form, for example "S will do A" (Davis 1984, p. 43):

> S intends that p [if and only if] S believes that p because he desires that
> p and believes his desire will motivate him to act in such a way that p.
> (Davis 1984, p. 51)

To understand the component parts of Davis's account and to better compare it with other belief-desire accounts we can clarify his formulation by saying that S intends that p if and only if (1) S believes that p, (2) S desires that p, (3) S believes his desire will motivate him to act in such a way that p and (4) 2 and 3 both explain (are the reason why) and are grounds for (are the reason for which) 1.

What Davis calls the connection condition is in this formulation captured by (3) and (4). The reason (4) is included is that it is how Davis specifies the word "because" in his original definition (Davis 1984, p. 51). The connection condition is one way to express "the appropriate way" of connecting desires and beliefs that late Davidson called for (Davidson 1978, p. 87). Davis's original reason for introducing the connection condition was to exclude intending things that the agent believes to be out of their power. To achieve this the agent must believe that the object of the intention is up to the agent and "[s]omething is up to [the agent] ...if it depends on his desires in a certain way" (Davis 1984, p. 51). In other words the agent must believe that the action depends on his desires in a certain way. This is expressed mostly by (3). The connection condition is also meant to replace Robert Audi's condition that S have no stronger competing desires (Audi 1973, p. 395–396), since those kinds of desires would not motivate S to act in such a way that p and thus be a reason for the belief that p (Davis 1984, p. 50).

4 Nonreductive accounts and reductive cognitivists

The strongest opposition to reductive belief-desire accounts of intention detailed in the last chapter was formed in the 1980s by theorists such as Michael Bratman, Alfred Mele, Hugh McCann and Myles Brand. The latest non-reductivist accounts I'm presenting here come from Gilbert Harman and Kieran Setiya. Their common challenge is to both give a critique of the existing belief-desire accounts and provide a positive non-reductive belief-desire-intention account that features intention as an independent mental state. I will also present a arguments from J. David Velleman who argues for his own reductive cognitivist view of intention by mostly arguing against the standard belief-desire reduction that is focused on desire's ability to motivate.

4.1 Myles Brand: Naturalistically inspired arguments

Myles Brand (1984) has argued against reductive belief-desire accounts, especially Robert Audi's (1973) account. His naturalistic claim is that philosophical intuitions

about folk psychology provide a first approximation, but it should be supplemented by scientific psychology (Brand 1984, p. 147).

Brand also has more specific arguments arguments against reduction. He argues that sometimes agents can have intentions without never having had the relevant desires (Brand 1984, p. 122–123). Examples of this kind of intentional action without desire are acting out of duty, coercion, psychological or physiological compulsion or necessity of circumstance. In some of these cases the agent can even hold a desire not to do the intended act, but still act intentionally (Brand 1984, p. 123). Brand further argues that it would be a mistake to reply that the agent's desire to do the intended act is identical with her desire to do her duty and that thus the agent simply desires to do her duty, since these two desires are distinct from one another (Brand 1984, p. 122–123).

However, it's not clear why the reductive theorist would need to identify these desires with one another, if the agent's desire to do the intended act can be instrumental upon the desire to do their duty. In that case the agent's desire to do the intended act would spring from the desire to do their duty.

Brand also argues that the several differences between desires and intentions indicate that desires are not a species of intentions or vice versa. He claims that the differences are also present where intending is claimed to be "a complex attitude consisting of desiring plus believing" (Brand 1984, p. 126).

First, Brand argues that while desire and belief are graded attitudes, intention is not (Brand 1984, p. 125). The strength of desires and beliefs can vary temporally over time, whereas the strength of intentions isn't graded or scalable at all — we only intend or don't intend — and that this also applies to belief-desire complexes as conjoining two graded attitudes can only yield a third graded attitude.

Second, desires often conflict with one another, but intentions don't. One can desire to have pie and desire not to have the same pie, but it would be inconceivable to intend to have pie and intend to not have the same pie (Brand 1984, p. 125–126). Brand further argues that adding beliefs in the mix would make the resulting attitude irrational but not inconcievable, as sometimes "persons have contradictory beliefs about what they will do" (Brand 1984, p. 127).

It could be argued that the concept of conflicting intentions is not inconcievable, but merely irrational. While intending to Φ and not Φ would seriously lack coherence, arguably so would believing that I will Φ and that I will not Φ. In addition to contradictory beliefs, we can also hold contradictory intentions. For example, one might hold the belief that one will be at place P at noon and hold the belief that one will be at place Q at noon. Similarly, one might intend to be at place P at noon and intend to be at place Q at noon. Perhaps what seems inconceivable to Brand

is the agent being aware of the conflict, but he doesn't quite say why conflicts of intentions should be less conceivable to us than conflicts of beliefs. Otherwise this seems very similar to Hugh McCann's (1986) critique that we will address along with possible answers in the next section.

Third, desires and intentions also differ in that it is possible to desire the goal, but not desire the means, whereas if we intend the goal, then we also intend the means to it (Brand 1984, p. 126). Of course it might be possible to intend to do something without intending any means at all if we don't know what the means are. In those cases we still tend to believe that there is or that there can be a means that we will intend to act upon in the future. Brand further claims that an added belief won't make an undesired means any more desired (Brand 1984, p. 126).

Fourth, Brand argues that intending is always self-directed, while desire is not and neither are beliefs about actions. In the sentence "Richard desires that Pat vote" there is only one action that determines if the sentence is true so to speak: Pat's voting. In the sentence "Richard intends that Pat vote" there are two actions that determine if the sentence is true, one performed by Richard and one performed by Pat (Brand 1984, p. 126). Adding beliefs is of no use as "beliefs about future actions are not restricted to one's own actions" and thus the resulting attitude will not always be self-directed (Brand 1984, p. 127).

It seems that adding specific kinds of beliefs similar to Audi's (1973) account would change the dynamics considerably. For Richard to intend that Pat vote, Richard should believe not just that Pat will vote, but also that Pat's voting will be up to something done by Richard. These beliefs refer to two actions by two agents and the latter belief exhibits self-directedness as they refer to the agent's actions. These kinds of examples might render some plausibility to the argument that belief-desire complexes could have the potential to ameliorate the difference in self-directedness.

Wayne E. Davis's (1984) account seems to contain a similar kind of self-referentiality as the intending agent must hold a belief about their own future action. It could also be argued that later accounts of intention that feature belief-desire complexes more complex than mere conjoined beliefs and desires, such as that the recent Humean accounts (Ridge 1998; Sinhababu 2017), might not be as vulnerable to the difference argument as presented by Brand, as they do not simply equate intentions with desires or beliefs.

Brand's overall critique of belief-desire reductions is tied with his own project of a naturalized account of action, in other words an account of action informed by the special sciences, in this case psychology. Brand criticizes the lax use of folk psychological concepts for the benefit of clear-cut philosophical accounts. To this

end, Brand utilizes Wilfrid Sellars's critique of Donald Davidson's (1963) use of pro attitude as an omnibus term for relatively long-term dispositions whose function as the proximate or immediate causes of action is not explained (Sellars 1973, p. 190). Brand's own critique is directed toward beliefs as the cognitive component of intending. He argues that this view oversimplifies the content of intentions and that scientific psychology can offer a more detailed view of cognitive attitudes in intending (Brand 1984, p. 147).

Brand's positive account of intentional action is that the proximate cause of action is an immediate intention (Brand 1984, p. 35), which has both cognitive and conative features that should be expressed more ambitiously than with mere beliefs and desires. In intentional action the cognitive component would be the guidance and monitoring of ongoing activity on the other hand and the prior representation of complex activities, i.e. plans, on the other (Brand 1984, p. 47). A complete account of the conative elements should involve a scientifically informed account of emotions and human motivation systems (Brand 1984, p. 237), but at the time of writing Brand merely suggests that the account could be put in terms of computational theory, while remaining essentially "blackboxed", that is to say without knowledge about the specific mechanisms in the human mind (Brand 1984, p. 268).

4.2 Hugh McCann: Against intention as the strongest desire

Hugh McCann has argued that the main problem with reductivist accounts is that they don't capture "the sense of resolve and purpose usually associated with intention" (McCann 1986, p. 193). A weak conception of desire by itself isn't enough to entail intention, since for example we don't intend everything we desire. A later popular solution (see for example Davidson 1970) from the belief-desire reductivists has been to refer to the agent's "strongest or 'definitive' desire" or "what the agent judges to be best overall" (McCann 1986, p. 193). According to McCann this won't do because of two reasons:

First, people say sometimes that they intend to do what they want and sometimes that they intend to do what they judge best. If both judgements and desires can lead to intentions and what we judge best and what we most desire don't always match, then knowing what the agent desires most and judges to be best won't allow us to infer what the agent intends unless the belief-desire theorists can explain when intentions are formed from the strongest desire and when from the best judgement (McCann 1986, p. 193).

I am not aware of a direct answer to this problem, but I believe that one possible

reductionist response to these kinds of critiques could be to follow Alfred Mele (1987, p. 38–39) and make a distinction between our evaluative judgements and motivational judgements. The reductive theorist could then argue contrary to Mele (1987) that while our evaluative judgements can differ from our strongest desires, our motivational judgements cannot. Correspondingly our motivational judgement can generate intentions, but our evaluative judgements cannot. The reductive argument goes like this: if we come to the evaluative judgement that it would be best for us to stop drinking and we sincerely utter the words "I intend to quit", our sense of intending will still be self-deceit without actual intending unless we also most desire to quit. We constantly have desires that are in line with our best judgement and this might create the false impression that judgements sometimes result in intentions, but whenever they are not in line we follow what we desire most. After all, we often judge best for us to stop smoking, work out, go to bed early and wake up refreshed, but still won't.

This argument would be similar to Davidson's distinction between *prima facie* judgements and all-out judgements that result from practical reasoning (Davidson 1978, p. 98) that we've already discussed and in line with Neil Sinhababu's Humean account of intending that we will discuss later. Mele's account differs from this in that he believes that evaluative judgements can result in intentions and when they are out of line with motivational judgement, the result is akratic action: doing something that we intend not to do. We will return to this later in the section concerning Alfred Mele.

Second, McCann argues that while there is nothing wrong with having conflicting incompatible desires, such as the desire to fly to the Maldives and the desire not to fly in order to protect the environment[2], in contrast having conflicting intentions seems to violate some kind of a rationality requirement (McCann 1986, p. 193–194). This means that intentions cannot be simply desires, but something more is needed to impose similar rationality requirements on reductionist accounts. McCann argues that a strong belief condition about doing (a belief that one will Φ) could approximate the necessary commitment and solve the problem of conflicting intentions, but eventually ends up dismissing it as too strong (McCann 1986, p. 194–195).

McCann suggests that while it makes sense to impose some much weaker belief conditions for intending, such as not having the belief that they will never do Φ, this doesn't help the reductivist cause as a weak belief condition wouldn't solve the problem (McCann 1986, p. 194–196). However, I believe that just a slightly stronger belief condition, such as Robert Audi's "x believes that he will (or that he probably will) do A" (Audi 1973, p. 395–396) that we saw in the last chapter, might also solve

---

[2]The example is mine.

the problem as it seems irrational for an agent to believe that they will probably act on two conflicting intentions.

Neil Sinhababu (2013) has later tried to avoid committing himself to strong beliefs about doing as a necessary condition of intending, but still utilizes them to answer the puzzle of conflicting intentions as presented by Michael Bratman (1987, p. 16–17). He argues that strong beliefs about doing are "a contingent psychological fact": they are not necessary for intending, but still abundant enough to provide our everyday intentions consistency (Sinhababu 2013, p. 684, 691). Sinhababu seems to accept that intentions exclude conflicting intentions as a rationality requirement (Sinhababu 2013, p. 691), but Sinhababu doesn't explain what might happen in the rare cases when we don't have a belief about doing and end up forming inconsistent intentions. We will return to Sinhababu's conception of belief about doing in a later chapter in his response to Kieran Setiya.

Reductionists who dislike the idea of a strong belief about doing requirement for intentions have often utilized the concept of strongest desire. While conflicting strongest desires can exist if they are of the same strength, conflicting desires that have the same strength are often explicitly excluded from these accounts of intending (see for example Audi 1973; Sinhababu 2013). After all, if we genuinely cannot decide of two options, it might be plausible that we won't choose either unless we're further motivated to choose between the two. All in all, it seems that reductionists have several ways to answer to McCanns second critique, but answers to the first one are more controversial as we will see when we return to the problem with Alfred Mele's arguments.

4.3 Michael Bratman: Planning and commitment

Perhaps the most influental nonreductive belief-desire-intention theorist, Michael Bratman (1987) sees intentions as plan states that are central to extending our agency through time. On Bratman's account intentions are distinctive folk-psychological mental states on par with beliefs and desires.

Bratman notes that as human beings we are intellectually resource bound rather than "frictionless deliberators" (Bratman 1987, p. 28). Because of this scarcity we constantly settle for a course of action rather than keep deliberating on our plans every second to match our current best knowledge. If we did consider our future actions over and over from the time we conceive the idea to the time of execution, our actions would certainly be more coherent, but we would have drastically less time to come up with new ideas or actually do anything besides deliberate on our existing ideas.

Given these limitations and the need to navigate through life, Bratman argues that intentions provide us "a way to allow prior deliberation to shape later conduct." (Bratman 1987, p. 29) This is why Bratman thinks that planning is such a crucial part of human existence, that future-directed intentions are so central to intending (Bratman 1987, p. 4–5) and that commitment is characteristic of intention. Future-directed intentions are intentions that are directed towards the future, whereas intentions that relate to current action are often called intentions-in-action. They differ from Donald Davidson's pure intending (Davidson 1978, 83) in that future intentions can be results of practical reasoning and they may have consequences.

I will next introduce two well-known original arguments from Bratman against reductionist belief-desire accounts. The other has to do with the characteristics of commitment in future-directed intention and the other is related to the famous philosophical example of Buridan's ass.

4.3.1 The initial challenge

Bratman posits that intentions arise through different sorts of psychological processes that fall under practical reasoning. According to Bratman, one such mechanism is the demand for means-ends coherent plans. If we have committed ourselves to an end, we must at some point find the means to that end. This means that having a partial, incomplete plan puts normative pressure toward filling it out as it becomes necessary, thus making us elaborate our old plans and in the process build new subplans (Bratman 1987, p. 31–32).

This demand for means-ends coherent plans can also be described in terms of intentions, i.e. plan states, as a framework against which practical reasoning weighs our beliefs and desires. The framework of intentions presents us with new problems concerning our future endeavours. At the same time it constrains their admissible solutions to those that are not in conflict with our plans (Bratman 1987, p. 35).

According to Bratman, these notions form the essential characteristics of future-directed intentions, which I have illustrated hierarchically in Chart 1. These characteristics consist of two deeply conjoined dimensions of commitment: volitional commitment and reasoning-centered commitment. The first dimension of commitment, volitional commitment, simply means that future-directed intentions are pro attitudes that motivate action in a conduct-controlling way. In other words they don't merely influence on the agent's reasoning and decision-making as desires and beliefs do, but rather intentions control the agent's conduct.

Chart 1: Characteristics of commitment in future-directed intention (two dimensions)

1. Volitional dimension

   (a) Pro attitude, motivates action in a conduct-controlling way

2. Reasoning-centered dimension (two facts or three dispositions)

   (a) Fact of being settled (stability/inertia/nonreconsideration)

      i. Disposition to retain the intention

   (b) Fact of having a role in further reasoning

      i. Disposition to reason from prior to further intentions

      ii. Disposition to constrain (exclusion)

(Bratman 1987, p. 16–17)

The second dimension of commitment, reasoning-centered commitment, can be divided to two facts or three dispositions (Bratman 1987, p. 16–17). The first fact is that intentions have a sort of inertia that implies that the issue has been settled. It can be also described as the disposition to retain the intention without reconsideration. The second fact is that intentions have a role in further reasoning, since we tend to reason from "intended end to intended means or preliminary steps" and "from general to more specific intentions" while holding our intentions consistent with other intentions and our beliefs. This fact of further reasoning can be described in terms of two dispositions: the disposition to reason from prior intentions to further intentions and the disposition to constrain further intentions in light of prior intentions.

These aspects of future-directed intentions also form the backbone of Bratman's initial challenge for belief-desire models (Bratman 1987, p. 16–17). He argues that these phenomena that are crucial to intention cannot be adequately expressed strictly within a reductive belief-desire framework. Neil Sinhababu (2013; 2017) has recently provided a thorough answer to this challenge and I will return to his replies in a later chapter.

The initial challenge in terms of the reasoning-centered dimension of commitment is that despite having relevant beliefs and desires the agent might still not see the issue as settled. Instead he might be disposed to deliberate the issue further, thus violating the fact of being settled and the disposition to retain the intention. Rephrased in terms of volitional commitment the problem is that beliefs and desires

alone do not have the conduct-controlling aspect characteristic of intentions. If the issue is not settled, the end is not settled and there will be no appropriate disposition to reason about the means (Bratman 1987, p. 18–19). Furthermore, belief-desire models tend to misdescribe the disposition to constrain future intentions and beliefs, as the agent can't have conflicting predominant beliefs by definition, thus excluding irrational intentions altogether (Bratman 1987, p. 19).

Bratman also states that his challenge is not fundamentally as much about the formal incapability of belief-desire models to represent this kind of commitment as it is about the theoretical fecundity of the models: "I think we gain more insight into the kinds of agents we are by putting aside such attempts at reduction and taking seriously the idea that intentions are distinctive states of mind" (Bratman 1987, p. 20).

4.3.2 Buridan cases

Another famous critique against belief-desire accounts that's associated mostly with Bratman has to do with Buridan cases. In the famous example of Buridan's ass, associated with the determinist philosophy of Jean Buridan (1300–1358), a donkey perishes in front of two equidistant bales of hay, as they are equally tempting. Bratman uses the example to showcase a crucial difference in how practical rationality works in the reductionist and nonreductionist views (Bratman 1987, p. 22–23).

In Bratman's version we're heading to San Fransisco and must choose between route 101 and route 280. The belief-desire reasons for taking route 101 rather than route 280 are equal in weight. To get to route 101 the agent should turn right at Page Mill Road. This seems to pose two kinds of problems for reductionist accounts. First, following strictly Donald Davidson's account, the agent cannot form the unconditional judgement, i.e. intend, to take either route as neither route is more desirable and the forming of unconditional judgements is necessarily comparative (Bratman 1999, 219-220).

The second problem is more detailed and relies on our common-sense conceptions of intending. After the belief-desire agent forms a belief-desire intention to choose one of the routes heading to San Fransisco, the reasons remain of equal weight. At Page Mill Road they might reconsider their earlier decision by involving all the relevant belief-desire reasons. Even if they managed to shift the balance of their motivation, at page Mill Road the agent can again compare their reasons. Bratman argues that this is not how we usually think of practical reasoning and intending: rather we don't tend to reconsider our intentions and they guide our actions without having to be subjected to comparison with the belief-desire reasons — rather prior intentions "have direct relevance to the rationality of further intention and action"

(Bratman 1987, p. 23). The nonreductionist can say that after we arbitrarily form the intention — subsequently independent of the belief-desire reasons of equal weight — to take e.g. route 101, turning to route 280 would simply be irrational for the agent because of the prior intention alone (Bratman 1987, p. 23).

Bratman has since famously utilized his planning approach to cover resource-bounded rational agents in the field of artificial intelligence (Bratman et al. 1988), a plan-based model of self-governance (Bratman 2009) and a philosophical account of collective action (Bratman 1999, 2014). It is an open question whether or not presenting any or all of these applications with belief-desire agents would yield similar results, but in terms of fruitfulness the account does speak for itself.

4.4 Alfred Mele: Motivation and akrasia

Alfred Mele has presented another influential nonreductive account of intention and notable critiques against reductive belief-desire accounts. For Mele, intention is a characteristically executive mental state that involves "being settled upon" its object (Mele 1992b, p. 160). According to Mele, the executive characteristic common to all intending and intentional action is what remains even when we substract the following three traits that are usually associated with being settled upon: firmness (how firmly we are settled upon), duration (how long the state lasts) and etiology (how it came about). When these are removed from the state of being settled upon, only the executive characteristic remains and that is what Mele calls "thin settledness" (Mele 1992b, p. 167). If thin settledness "is not revoked or revised, it will initiate appropriate action under favorable conditions" (Mele 1992b, p. 163). This kind of executive settledness is something that belief-desire intentions lack.

He argues against reductive belief-desire accounts by attempting to reveal a disconnect between one's judgement on the one hand and motivation typical to belief-desire accounts on the other with several examples, some of which have to do with irrationality and specifically the old philosophical problem of akratic action. Next we will take a look at his arguments in detail, beginning with the disconnect between motivation and intending.

4.4.1 The disconnect between motivation and intending

Most reductive belief-desire theorists like Davidson (1963), Audi (1973) and Davis (1984) identified intentions with the strongest desire or predominant motivation together with the accompanying beliefs. Predominant motivation means that the agent's "desire to A is motivationally stronger than his competing desires" (Mele 1988, p. 155–156). Much like Michael Bratman (1987, p. 18–19), Mele argues against

identifying intention with preponderant motivation by presenting examples that focus on whether or not being preponderantly motivated and the issue being settled, an important characteristic of intending, go hand in hand. The first example seems to challenge reductive belief-desire accounts in a way similar to the one presented by Hugh McCann (1986) that we discussed earlier.

In the first example Alan is predominantly motivated to humiliate Carl later today at a party, i.e. he wants to humiliate Carl "more than he wants to do anything that he takes to be incompatible with his doing this", but believes this to be morally wrong and thus decides to resist his desire to humiliate Carl (Mele 1988, p. 155–156). Having memorized his earlier attempts to resist similar desires and calculating a low success rate of 25%, Alan believes that he will probably humiliate Carl. Mele's intuition is that Carl actually doesn't intend to humiliate Carl as he has explicitly decided against it, but on belief-desire analyses of intention he does intend to humiliate Carl as he is predominantly motivated to do that. If the intuition holds up, the belief-desire analyses must be mistaken (Mele 1988, p. 156). I will return to Michael Ridge's (1998) belief-desire response later in chapter 6.

Mele's second example aims to show that it is possible to be preponderantly motivated without having settled the issue (Mele 1992b, p. 142). A mother is more motivated to miss rather than attend her son's wedding due to her agoraphobia, i.e. fear of open places. She also believes that she will most likely miss the wedding because of her phobia. However, she has sought professional help and will continue her attempts to weaken her phobia as her motivation to attend is still strong, albeit presently weaker than her motivation to miss the wedding. According to Mele, her determination to overcome her phobia shows that she is not settled upon missing the wedding. The example supposedly shows that if she is preponderantly motivated to miss the wedding, but not settled upon missing the wedding, there must be something wrong with identifying intending with being preponderantly motivated.

It seems to me that the example could be argued against by denying the plausibility of some of the premises. One possibility is this: if the mother truly were preponderantly motivated to miss the wedding, all things considered, why would she act against her preponderant motivation by weakening her fear of crowds? The example guides us to think in terms of choosing between motivation to attend and stronger motivation to miss the wedding. But if we think that her agoraphobia results in an aversion that is stronger than her desire to attend at the moment of deliberation, we might be unwittingly disregarding more elaborate scenarios that factor in her attempts to lessen the impact of her phobia between the time of deliberation and the time of wedding, thus "shifting the balance" (Mele 1988, p. 241) of her motivation. In other words, while missing the wedding might be a preferable

scenario to attending and suffering from agoraphobia, attending the wedding and not suffering from agoraphobia might be preferable to both. In that case the mother might have preponderant motivation to attend the wedding rather than miss it. It is possible that she has a *prima facie* desire in the Davidsonian sense to miss the wedding that is stronger than her *prima facie* desire to attend the wedding if she had to decide right now, but she can also have an even stronger desire to treat her phobia and attend.

So there seems to be two ways of thinking about the example. The nonreductionist interpretation might be that while the mother can be more motivated to miss the wedding than attend the wedding, she can still decide to shift the balance of her motivation. The reductionist interpretation might be that if her desire to miss the wedding is stronger than her desire to attend the wedding, but she still doesn't act accordingly and intend to miss the wedding, we must be talking about *prima facie* desires that include her fear as it is at the moment of deliberation, but not the possibility of shifting the balance of her motivation by alleviating her agoraphobia. What we should be comparing with each other are the desire to miss the wedding all things considered and the desire to attend the wedding all things considered. If we do, we should see that her desire to attend is stronger and this explains why she is taking steps to go to the wedding. The reason why she doesn't settle the issue and intend to miss the wedding is that she doesn't want to. It's not exactly clear which one is the correct interpretation, but at least the belief-desire view seems to have a strong counterargument.

Mele's third example, or rather set of examples, posit that it is possible for someone to have settled an issue while not being preponderantly motivated, thus again revealing a disconnect between preponderant motivation and being settled upon a course of action, a crucial element of intending (Mele 1992b, p. 142–143). The examples are based on having decided on a course of action, but being impeded by fear.

Mele argues that at the time someone becomes settled upon petting a snake, he can be preponderantly motivated not to do so, for he could be psychologically unable to pet the snake because of his surprisingly strong fears (Mele 1992b, p. 142–143). Similarly, a boy might intend to go through a ceremony of blood brotherhood by cutting his own hand and be surprised when fear motivates him to not do it. "The process that resulted in his intending to cut his hand did not also result in a preponderance of motivation to cut it" (Mele 1992b, p. 165). Finally, a man might intend to jump from a burning house, but be shocked to find out that he cannot bring himself to do that (Mele 1992b, p. 165).

Let's look at the fear in question as something that the agent holds like a reverse

desire, i.e. an aversion. As such it is a part of the agent's motivational calculus, even if it does not feature in the agent's conscious evaluations. So far the example seems plausible. On the other hand, the agent might be disposed to be very afraid of snakes, blood or jumping from a height, even if they are not afraid at the time of evaluation and, crucially, doesn't believe that they will be. Here of course the agent is mistaken. As the agent is not afraid or believe that he will be afraid, their fear cannot feature as a conscious or an unconscious element in the agent's motivational calculus. Rather it seems that the fear should only enter the motivational calculus when the agent is surprised by their fear and as a result their motivational balance changes.

Mele argues that while it's possible that the motivational balance will shift as I argue above, it's also possible that the motivational reasons are stronger at the time of evaluation. However, the three examples above seem to rely on that the strongest motivational reason, fear, is neither present nor acknowledged by the agent. If the agent is shocked and surprised by the fear at the time of initiating action, it seems odd how it could be a motivational reason for the agent at the time of evaluation.

4.4.2 Akratic action

A related line of argument concerns the age-old philosophical question of akratic action (Mele 1992b, p. 122; Mele 1987, p. 97). Akratic action, also known as weakness of will, is acting against one's best judgement: believing that there is a better course of action available, but not doing it.

The problem of akratic action for causal theories of action is this: if causal theories of action state that the strongest reasons for action are the strongest causes of action and if it's nevertheless possible to intentionally act against one's strongest reasons, the causal reductive account is in trouble (Davidson 1980, p. xvi). Thus it seems that either we must abandon either some part of the causal theory of action or weakness of will (Davidson 1970, p. 23).

Davidson's answer is that akratic action exists, but the agent's causally strongest reasons do not always need to provide the rationally strongest grounds for acting (Davidson 1970, p. 41). Rather these kinds of actions are genuinely akratic in that the agent has no reason for doing something they have judged as inferior to the alternative (Davidson 1970, p. 42). While akratic action is preceded by practical reasoning with all the necessary premises to its conclusion, which is non-akratic action, the reasoning is faulty or missing altogether: "Every judgement is made in the light of all the reasons ... [b]ut 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." (Davidson

1970, p. 41–42)

Mele finds this answer unsatisfactory, as it doesn't go deep enough into why agents behave akratically. Mele interprets Davidson as saying that the akratic agent acts according to their motivation that's based on a set of reasons r, regardless of their judgement that's based on a more comprehensive set reasons r' that includes r. However, Davidson's account doesn't explain why the agent is motivated by reasons r rather than reasons r' (Mele 1992b, p. 122). We can only say that the agent acts irrationally, but there's no further explanation to the agent's irrational akratic action.

Mele's description of akratic action implies that the disconnect between judgement and motivation is more profound than just fringe cases of irrational action. When people make decisions, they evaluate their options. Decisions made on the basis of such assessments can lead to the agent having an intention, even when the assessment is not in line with the agent's motivation. For example, a smoker might assess their situation, conclude that they should stop smoking, decide to quit and form an intention to quit while still being, perhaps subconsciously, more motivated to smoke than to quit. Unless the motivational balance shifts, the smoker will akratically continue to smoke (Mele 1992b, p. 123). In other words, evaluational assessments as the basis of decisions provide an alternative route to having intentions in addition to desire-driven processes. Akratic action can result when decisions are based mostly on the agent's evaluative assessment and when that assessment is not in line with the agent's preponderant motivation.

Mele concludes that while counterexamples like akratic action don't necessarily mean that traditional belief-desire models should be discarded altogether, they should rather be reworked to present more modest, non-reductive versions (Mele 1987, p. 108; Mele 1992b, p. 123).

4.5 Gilbert Harman: Practical reasoning

Gilbert Harman's account of intentions emphasizes their role as the conclusions of practical reasoning that can lead to action. This practical reasoning abides by the principles of (I) conservatism, minimizing changes in intentions; (II) coherence, encouraging coherence between intentions and beliefs; and (III) desire satisfaction, encouraging changes in intentions that seem to realize our goals (Harman 1986, p. 77, 94). Practical reasoning is distinct from theoretical reasoning, which directly results in beliefs, although it can also affect intentions indirectly by changing our beliefs (Harman 1999, p. 46).

For Harman, intentions have several characteristics: (1) they are primitive mental

states, i.e. not to be analyzed away in terms of beliefs and desires, (2) they are self-referential and (3) they involve an initial plan and the belief that one will do what one intends, which in turn implies the possibility to do something intentionally without intending it (Harman 1999, p. 46–48).

The self-referential nature of intentions (2) means that all intentions to Φ contain an intention to intend to Φ that is inseparable from the intention itself. According to Harman, this does not lead to infinite regress, as the intentions are not distinct from one another (Harman 1999, p. 54; Harman 1986, p. 85–86).

The intending agent must also have an initial plan (3) that concerns the means to what one intends: "One's plan is that one's intention will lead in such and such a way to one's doing A" (Harman 1999, p. 57) Harman thinks that plans also tie up with the self-referentiality of intentions in the following way: if plans say what one intends to do and the content of a plan is that "one's intention will lead in such and such away to one's doing A", therefore when having a positive intention "one intends that one's intention will lead in such and such a way to one's doing A" (Harman 1999, p. 57). This plan does not have to be elaborate or complete, but at the minimum there should be a sense of how one's intention will lead one to plan (Harman 1999, p. 57; Harman 1986, p. 84).

According to Harman it is possible to do intentionally something one doesn't intend and even vice versa: "cases in which one intends to do something and does it, but not intentionally" (Harman 1999, p. 57). The former is possible in cases that include a foreseen consequence, like leaving a tightly packed parking spot, trying one's best not to dent anyone's car, but still denting someone's car. One didn't aim to or intend to dent the car, but still one acted intentionally in denting the car (Harman 1986, p. 89–90). A similar example can be found from Bratman (1987, p. 123). The latter can happen when, for example, Mabel intends to kill Ted at his house, but accidentally runs him over backing out of her driveway. The killing of that person was unintentional, even if when Mabel had the intention to kill the person and she did (Harman 1999, p. 57–58).

Harman also uses the previous example to illustrate how intentions can be distinguished from desires (Harman 1999, p. 58). Unlike intentions, desires can be separated from the desire that the desire will lead in a certain way to the desired event. In the example the desire to kill Ted can be decoupled from the desire to kill Ted at his house, but the intention to kill Ted cannot be decoupled from the intention to kill Ted at his house, supposing that that was Mabel's intention. Mabel got what she wanted, but not what she intended, even though she both wanted and intended to kill Ted. Harman doesn't discuss the implications any further, but this argument doesn't seem to fare as well against accounts that identify intentions with

belief-desire pairs at least when they include a belief concerning the way the desire is to be fulfilled (e.g. Ridge 1998; Sinhababu 2013, 2017).

Harman's most explicit worry towards reductive belief-desire accounts is that defining intentions in terms of beliefs and desires might be analyzing away intentions rather than taking them "seriously as psychological states on a par with beliefs" (Harman 1999, p. 55). Wayne A. Davis has denied this in a short response to Harman. Davis argues that there is no reason to think that even a reductive analysis of intentions should make intentions any less interesting for most philosophers (Davis 1984, p. 50–51).

4.6 Kieran Setiya: The necessity of belief about doing

Kieran Setiya's (2008) account is based around belief as the necessary condition of intending and the relationship between action and practical knowledge. Practical knowledge can mean either Anscombean knowledge without observation (Anscombe 1958, p. 13) about what we are intentionally doing or knowledge about how to do something we are motivated to do. Setiya argues forcefully for the possibility of practical knowledge and the intimate connection of the two meanings of practical knowledge mentioned above, but for the purposes of this thesis I will focus on the implications his view has on belief and its relationship with intentions.

Setiya follows Elisabeth Anscombe (1958) in arguing against inferentialism about belief about doing. The inferentialist stance in short is that our beliefs about our intentional actions, beliefs about doing, can be inferred from such things as bodily movements or prior knowledge of the will.

Setiya begins from Stuart Hampshire's notion that "doing something . . . intentionally . . . entails knowing what one is doing" (Hampshire 1959, p. 102). Setiya states that belief, rather than knowledge, about intentional action cannot be inferred from intentions either, because intentions don't precede belief about doing, but they rather constitute the relevant beliefs (Setiya 2008, p. 396). Setiya concludes that belief about doing must be a necessary condition of intentional action:

> If A is doing Φ intentionally, A believes that he is doing it or is more confident of this than he would otherwise be, or else he is doing Φ by doing other things for which that condition holds.
> (Setiya 2009, p. 129)

In other words: basic intentional actions, "ones that are not done by doing anything else intentionally" (Setiya 2008, p. 390), entail the belief that one is doing said actions. Setiya's objection to the inferentialist view, stemming from the notion

that belief about doing is a necessary truth, is this: if belief about doing was a mere inference as the inferentialists claim, it wouldn't be a necessary truth as one could fail to make such an inference, "fail to put two and two together" (Setiya 2008, p. 394).

Setiya's positive account of intentions is also connected with the notion of belief about doing. He describes intentions as desire-like beliefs: "[o]ne intends to be doing Φ just in case one has the desire-like belief that one is hereby doing it" (Setiya 2007, p. 48). Similarly, in the case of future intentions "one intends to Φ, prospectively, just in case one has the desire-like belief that one is hereby going to Φ" (Setiya 2007, p. 49). However, Setiya's position is still not reductivist (Setiya 2008, p. 395).

According to Setiya the most important features of his account are that it shows how intending plays a motivating role in action, that it involves the belief that one is or will be acting and that it is self-referential (Setiya 2007, p. 48–49).

It should be noted that the doctrine of belief about doing as a necessary truth for intentional action is not a divisive issue between reductive belief-desire accounts and nonreductive belief-desire-intention accounts. For example Robert Audi's reductive account of intending includes belief about doing (Audi 1973, p. 395–396). Furthermore, it has been criticized not only by reductive theorists like Donald Davidson (1963) and Neil Sinhababu (2013; 2017), but also by nonreductive theorists like Michael Bratman (1987) and Alfred Mele (1992b). The issue can only be framed as a critique of those accounts that don't already include belief about doing.

There are many who are not willing to commit themselves to the idea of belief about doing. We will look at a counterexample to Setiya's brand of belief about doing when we examine Sinhababu's response to Setiya in chapter 6.

4.7   J. David Velleman: A cognitivist account

Those who argue beliefs are intentions or form a significant part of intentions, are usually labelled cognitivists. One of them, J. David Velleman has put forward two objections against "the standard story" of causal reductive belief-desire accounts. While Velleman's own account is reductive, I have included him in this chapter as he provides two counterarguments against the standard reductive view. His first objection is that they fail to include the agent in their proper role (Velleman 1992b, p. 461-463). It seems as if we are reduced to puppets without free will, which makes us feel uneasy about the accuracy of those accounts. Velleman thus claims that causal reductive accounts and human agency have so far been incompatible.

Velleman provides his own reductive solution to problems with the disconnect between motivation and intending, such as Mele's wedding example I've introduced

earlier. His solution is that the agent can make the weaker motive prevail by "throwing his weight" to support it and thus make the combined motivation stronger than the competition (Velleman 1992b, p. 480). This is part of what agency is for Velleman, "the principal component of agent-causation is the agent himself" (Velleman 1992b, p. 475).

However, the causal reductivist might argue that the mental states in question form our agency and this is why the agent is not missing any more than a forest is missing from the trees. The agent and their mental states are not "rival candidates in the search for causes" (Goldman 1976, p. 81), but rather they describe same causal process form two different points of view. Velleman has argued that this point is still contestable (Velleman 1992b, p. 463).

A similar response has been put forward by Michael Smith: that the role of the agent is constituted by the "rich pattern of connections between an agent's desires and beliefs and bodily movements" (Smith 2012, p. 399). This is how the agent can intervene between mental states and bodily movements. Smith agrees with Velleman that the causal reductive accounts are not suitable tools for examining the agent's exercise of their agency, but this doesn't diminish the significance of these accounts (Smith 2012, p. 399–400).

Velleman's second objection is that the standard noncognitivist belief-desire accounts that are focused on the role of desire assert that the agent must see everything they act on intentionally as valuable, so they must act "under the guise of good" (Velleman 1992a, p. 3). The failure of the standard view "is in presupposing that whenever a proposition is regarded as true or to be made true, its truth thereby comes to constitute a success — and its falsity, a failure — for which either the attitude or the world must bear responsibility" (Velleman 1992a, p. 13).

According to Velleman the traditional view is based on an idea inspired by John Searle's (1975) notion of direction of fit that belief aims at the truth and desire correspondingly aims at the good is mistaken because desire doesn't have correctness as its constitutive aim (Velleman 1992a, p. 15-17). Rather, "desire has the same subjective justificatory force as fantasy — that is, none at all" (Velleman 1992a, p. 15-17). This controversial point is firmly grounded on the notion of desire and as such might apply to many nonreductionist accounts as well. As I'm not going to discuss here which conception of desire is the most plausible, I'm regretfully not going to engage the argument much further.

We've seen in this chapter multiple arguments against traditional belief-desire accounts. Most of them were directed against the concept of predominant desire or predominant motivation. The nonreductionist view has since gained prominence, but the belief-desire view is still tempting for many. We will next take a look at two

very different theories of intention that both focus on the concept of desire.

5 Humean belief-desire accounts

A relatively novel challenger for nonreductive belief-desire-intention accounts is a neo-Humean revival of reductive belief-desire accounts of intention. These reductive accounts of intention give priority to desires following the way David Hume thought of passions as the basis of intentional action. There has been a long tradition in western philosophy of seeing desire and reason as opposing forces, with reason being the force we should listen to in order to control our desires. Challenging this conception, Hume famously thought of reason as bound to be a "slave of the passions" (Hume, T II.3.3 415). The idea was that if only passions can give rise to intentional action, then we cannot intentionally resist our passions with pure reason. To overcome one passion, another passion needs to be involved. Those passions can relate to our use of reason, but nevertheless they are the only source of motivation.

There is a long history of Humean thought in the philosophy of action, but I will focus on two recent Humean accounts of intention by Michael Ridge (1998) and Neil Sinhababu (2013, 2017). The accounts seem to draw inspiration from different aspects of the Humean tradition as Ridge's dispositional conception of desire is rather thin and abstract, whereas Sinhababu utilizes different aspects of desire in his reductive account.

5.1 Michael Ridge: A Humean dispositional account

The first Humean response to Michael Bratman and other proponents of nonreductive belief-desire-intention accounts was Michael Ridge's *Humean Intentions* (1998). In it, Ridge develops two lines of arguments: one based on his own account of intentions and a proposed strategy to deal with future critics of reductionist accounts.

Ridge's own reductive conception of intentions is based on his dispositional characterization of desires: "A desires to Φ if and only if *either* (a) for any Ψ, if A believes that Ψ is a necessary means to Φ then A is disposed to Ψ, *or* (b) A is disposed to experience pleasure upon coming to believe that Φ." (Ridge 1998, p. 160). How do these dispositional desires then underlie intentions?

Like early reductive theorists such as Davidson (1963) and Audi (1973), Ridge first invokes the concept of dominant desire to distinguish the desires that guide action: "to have a strongest, or dominant desire to Φ, there must be no desire Ψ such that the agent views Ψ as a real competitor with Φ, and the agent's desire to Ψ is stronger than her desire to Φ" (Ridge 1998, p. 161). A predominant desire would

simply be a dominant desire Φ that has no competing desires Ψ as strong as Φ, thus ruling out ties between competing desires (Ridge 1998, p. 163). Having done this, Ridge is ready to formulate his conception of intending:

> A intends to Φ if and only if
>
> (a) A has a desire to Φ,
>
> (b) A does not believe that Φing is beyond her control,
>
> (c) A's desire to Φ is a *predominant* one, which is just to say that there is no desire to Ψ, such that A does not believe Ψing is beyond her control, she desires to Ψ as much as or more than she desires to Φ, and she believes that a necessary means to her Φing is that she refrain from Ψing,
>
> (d) A has a desire not to deliberate any more about whether to Φ unless new, relevant information comes to light.
>
> (Ridge 1998, p. 163)

The first three elements of Ridge's account are very similar to many traditional BD accounts: a desire (a), a weak belief condition (b) and a set of predominance conditions (c) for the desire (a). The fourth element is novel. The desire not to deliberate (d) is designed specifically to answer Michael Bratman's (1987) worry about belief-desire accounts of intention, namely that the agent can have "a predominant desire to do something but not be settled upon it" (Ridge 1998, p. 173). If one is not settled upon Φ, in Bratman's sense of reasoning-centered and volitional dimensions of commitment (Bratman 1987, p. 18–19) I have described in an earlier chapter, it would seem odd to claim that one intends to Φ. Ridge accepts Bratman's argument that being settled is a prerequisite for intending, but unlike Bratman he argues that being settled can be formulated in reductive belief-desire terms. Ridge argues that to see an issue as settled is to have a potentially belief-mediated[3] disposition not to deliberate, which is simply a desire not to deliberate according to the Humean dispositional characterization of desire that we saw earlier (Ridge 1998, p. 163). Being settled on an issue could thus be expressed as a desire (d).

However, I believe there are two reasons to think that the desire not to deliberate (d) doesn't necessarily lead to seeing the issue as settled. First, Ridge hasn't specified that the desire not to deliberate should be predominant. This means that one could arguably intend to Φ, thus having a desire not to deliberate, but also have a stronger predominant desire to deliberate. This could of course be fixed by specifying that the desire not to deliberate should be predominant. Having two predominant desires

---

[3]This means that the dispositions may "lead to action via the influence of a belief or beliefs", but it could also lead to action directly (Ridge 1998, p. 175, n3).

shouldn't be a problem as (a) is not a "real competitor" (Ridge 1998, p. 161) to (d) or vice versa.

My second objection is more important as it shows that Ridge's account of intending doesn't hold up if we accept his account of desire. According to Ridge's account of desire introduced in the beginning of the chapter, a desire is a *either* a disposition to act *or* a mere disposition to experience pleasure upon coming to believe what was desired. It seems obvious that a disposition to experience pleasure upon coming to believe that one will not deliberate is a long way from actually not deliberating. This means that even if the desire not to deliberate was predominant, it wouldn't necessarily lead to seeing the issue as settled, as one can be merely disposed to experience pleasure rather than be disposed to not deliberate. If this is true, Ridge's account of intention fails in its explicit attempt to escape Bratman's critique.

To ameliorate this Ridge should either characterize the element of commitment in some other manner than (d) or somehow rework his account of desire. Whether or not this is possible in reductionist BD terms remains an open question. Simply dropping the disposition to experience pleasure from his account of desire might not be possible, as it serves a distinct purpose in Ridge's account (1998, p. 160).

In addition to his own account, Ridge proposes a general strategy to argue against two kinds of critiques directed against reductive belief-desire accounts (Ridge 1998, p. 164–165). Critiques of the first kind concern properties usually associated with intentions that the belief-desire accounts purportedly lack. For example, if desires lack the kind of commitment that is distinctive of intention, then reducing intentions to desires seems unjustified. Critiques of the second kind concern the difference between the normative constraints that govern intentions and the normative constraints that govern desires. For example, intending to have pasta now and not having it would make me irrational, while merely desiring to have pasta now is perfectly compatible with not having it (Ridge 1998, p. 157). If intentions are governed by different normative constraints than desires, it seems that intentions can't be desires.

Ridge's proposed strategy against both kinds of nonreductionist claims is simply that the belief-desire theorist can keep adding desires and beliefs to accommodate for any missing properties of intention that the nonreductive belief-desire-intention theorist brings up (Ridge 1998, p. 164–165). Ridge's own account of intending seems to utilize this general strategy against a critique of the first kind: Bratman's critique that being settled is a distinctive element of intending, but not a distinctive element of having a predominant desire. Reductive belief-desire accounts of intention that rely on predominant desires consequently lack the property of being settled. Ridge adds the desire not to deliberate, (d), into the account explicitly to answer Bratman's

critique (Ridge 1998, p. 163, 173).

Adding the desire not to deliberate (d) seems intuitive given Ridge's minimalist, dispositional conception of desire. Neil Sinhababu has argued that a belief-desire theorist wishing to employ a richer conception of desire, a desire not to deliberate would entail more than just being disposed not to deliberate and could easily lead to false phenomenological predictions (Sinhababu 2013, p. 690). Utilizing Ridge's strategy also has the danger of seeming ad hoc (Sinhababu 2013, p. 696, n27).

Those belief-desire accounts that attempt to give psychologically realistic accounts of intending (e.g. Sinhababu 2013), are even more vulnerable to the problem above. These accounts should presumably reflect our folk psychological intuitions about the desires and beliefs we hold while we hold intentions. Adding beliefs and desires on an ad hoc basis to answer particular nonreductionist critiques simply won't do. This doesn't mean that some reductive belief-desire accounts couldn't avoid these problems, but this does make drafting responses along Ridge's proposed reductionist strategy non-trivial.

Another possible problem in Ridge's account that relates to his emphasis on dispositions is that they make the account more vulnerable to accusations of being non-reductive. If we define desire and belief to be so epistemically thin that all relevant dispositions can be put in terms of desires and beliefs, then the reductive claim — that intentions are desires and beliefs and nothing else — seems somewhat less interesting, as characterizing intention as a collection of dispositions is not specifically a reductionist endeavour. The disagreement between belief-desire and belief-desire-intention theorists is not about whether or not intentions can be characterized as collections of dispositions, but what kinds of mental states underlie these dispositions.

5.2 Neil Sinhababu: A novel approach to Humean desire

The most recent defense of the Humean reductive view comes from Neil Sinhababu. He argues that intentions can be reduced to combinations of desires and beliefs and presents a thorough argument against two prominent critics of the reductive view, Michael Bratman and Kieran Setiya. In contrast to Michael Ridge's (1998) account, Sinhababu's account of intending utilizes heavily the many facets of desire, such as its motivational qualities and its attentional aspect. Sinhababu also argues that traditional decision theory systematically fails to provide solutions "when the complexity of the options outstrips our attentional resources, or when a slightly less-desired option is represented much more vividly", because it doesn't account for properties of desire and other psychological capabilities such as our limited attention

(Sinhababu 2017, p. 112). The basic structure of Sinhababu's theory of intending is as follows:

> A intends that φ[4] if A has a desire and belief such that for some behavior B and situation S:
>
> (1) A desires that φ.
>
> (2) The desire is combined with a belief that S will obtain, and that A's B-ing in S would make φ more likely.
>
> (3) If the desire were combined with a belief that S obtains now, they would without further reasoning produce motivational force sufficient for A to immediately initiate B-ing.
>
> (Sinhababu 2017, p. 100–101)

In other words, the sufficient conditions for intending are the desire that Φ (1), the belief that the situation S will obtain (2.1), the belief that A's B-ing in S would make Φ more likely (2.2) and two further background conditions: that the desire is combined with the beliefs (2.3), which I will call the combination condition, and that the combined beliefs and desire would make the agent initiate action without further reasoning if they were combined with a further belief that S obtains now (3), which I will call the contrafactual condition. This also holds for simple intentions where the agents goal Φ is performing behavior B (Sinhababu 2017, p. 101).

Sinhababu takes the actual intention to be the desire (1) while the background conditions (2) and (3) are necessary for the desire to be an intention (Sinhababu 2017, p. 101). He allows that the intention can also be taken to "consist of the desire from (1) and the belief from (2), positioned in [the agent's] psychology as described in (3)" (Sinhababu 2017, p. 101). Thus the intention to Φ can be either an appropriately situated desire that Φor composed of all the psychological states it involves.

The requirement that A's B-ing in S would make Φ more likely (2.2) allows for intending events that the agent deems highly improbable (Sinhababu 2013, p. 681; Sinhababu 2017, p. 102). As we can see, it is not a necessary condition of intending that the agent believes the contrafactual condition (3), though presumably the intending agent can believe it. On the other hand, the agent can also believe that the desire and belief would not produce motivational force sufficient for action. This seems to mean that in Sinhababu's account it is possible to have intentions unknowingly.

---

[4]Sinhababu doesn't explicitly define lowercase phi, φ, but he uses it to represent whatever we intend be it actions or states of affairs in the same way most writers use uppercase phi, Φ (Sinhababu 2013, p. 681).

According to Sinhababu, combining might be driven by many psychological forces, but one of them is the attentional aspect of desire (Sinhababu 2017, p. 85). The attentional aspect "disposes one to attend to things one associates with [the object of desire]" (Sinhababu 2017, p. 33). Attending to the the content of the instrumental beliefs (2.1) and (2.2.) combines them with the desire, thus producing motivation (Sinhababu 2017, p. 2, 35).

This formulation of intending seems to show some influence from Wayne A. Davis's account (1984), as the notion of combining is explicitly added into the account as a sort of connection condition, to use Davis's term. The most significant difference between Sinhababu's early (2013) and late (2017) formulations of intending is that in later account desire and belief must be combined in both (2) and (3). Sinhababu himself mentions that his 2017 formulation differs from the 2013 formulation "mainly by adding the notion of combining" (Sinhababu 2017, p. 101).

While the element of combining belief and desire is not explicit in the 2013 account, it is already present to some extent: Sinhababu mentions combining desires and beliefs while arguing his earlier formulation for the contrafactual condition: "(3) rules out cases where the agent desires something and has a belief about how to attain it, but hasn't yet combined these mental states and formed an intention. In these cases intention-formation requires a further step of practical reasoning in which the desire and belief are combined so they can motivate action" (Sinhababu 2013, p. 682). This is to say that uncombined beliefs and desires don't necessarily motivate action before the further step of practical reasoning, in which they are combined to form intentions.

The notion of combining is further tied with practical reasoning. In Sinhababu's account, pratical reasoning is combining beliefs and desires "to motivate action or produce instrumental desires", whereas theoretical reasoning is multiple beliefs combining to generate further beliefs (Sinhababu 2017, p. 83–86). One psychological mechanism for such reasoning is the way our desires continually direct our attention toward relevant beliefs unless hindered by our limited attention (Sinhababu 2017, p. 85). Limits of our attention is what limits our capability to reason: if we are distracted by sensory stimuli, our other desires, etc., our attention is divided and we can fail to combine desires and beliefs further (Sinhababu 2017, p. 85). If not, we will "attend appropriately and discover how to get what we want" (Sinhababu 2017, p. 84).

Let's turn to motivation, which Sinhababu defines as "the disposition to act" (Sinhababu 2017, p. 23). After the desire and the relevant beliefs have been combined through practical reasoning guided by the attentional aspect of desire or some other psychological mechanism, they produce a motivational force that is "proportional

to the desire's strength times the increase in subjective probability" from the belief
(2.2) (Sinhababu 2017, p. 23). According to Sinhababu, this kind of combining is
the necessary and sufficient condition for being motivated to A (A stands for action,
E for end):

> Desire-Belief Theory of Action: One is motivated to A if and only if
> desire that E is combined with belief that one can raise E's probability
> by A-ing.
>
> (Sinhababu 2017, p. 2)

In other words, weak desires and uncertainty about outcome produce less moti-
vational force than strong desires and certainty about outcome. The motivational
forces of the belief-desire pairs favoring action need to be greater than the moti-
vational forces against action, including motivational forces favoring other actions
that are mutually exclusive (Sinhababu 2017, p. 102). If the motivational forces for
action triumph, they will eventually cause action: "you do whatever you're most
motivated to do" (Sinhababu 2017, p. 2).

The belief-desire theory of action seems to rule out what the agent believes to be
impossible, as one has to believe that one can raise the end's probability by A-ing.
It also rules out being motivated through desires only and thus is stronger than the
belief-desire theory of action Sinhababu presented in an earlier paper (Sinhababu
2009, p. 465, n2).

Sinhababu's central claim against nonreductive belief-desire-intention models is
that his account can explain relevant facts about intention, utilizes the properties of
desire more thoroughly, explains some facets of intentionality more thoroughly such
as the hedonic phenomenology of intention (Sinhababu 2017, p. 114), while simplify-
ing our psychological theory by leaving out intentions as independent mental states
(Sinhababu 2017, p. 110). Three questions arise: first, does the model adequately
answer the challenges set by nonreductionists and does it otherwise model inten-
tions properly? Second, does it succeed in simplifying our psychological theory in
comparison to nonreductive belief-desire-intention models? Third, does Sinhababu's
use of the combining condition and the contrafactual condition lead to any novel
problems?

5.3   Classifying the arguments against reduction

In the last three chapters I have introduced multiple arguments for and against
reduction along with many accounts of intention. The most important arguments for
reduction have been the reductive accounts themselves: giving an exhaustive belief-
desire account of intention as an attempt to answer the challenges put forward by

the nonreductivists is the best defense for any reductive view. In chapter 6 we also took a further look at arguments for reduction that concerned theoretical virtues and explanation. The arguments against reduction on the other hand are more diverse; I suggest that they could be tentatively grouped four ways in order to clarify the situation:

The first class of arguments concerns the disconnect between intending and the motivating element — be it desire, strongest desire or predominant motivation. It would include Brand's argument on intentions without desire (Brand 1984, p. 122–123); McCann's argument against identifying intentions with the strongest desires or what is judged best, as they are both said to entail intending (McCann 1986, p. 193); and Mele's examples for the possibility of predominant motivation without being settled (Mele 1988, p. 155–156) and his argument focusing on akratic action (Mele 1992b, p. 122).

The second class of arguments contends that the necessary conditions or rationality requirements of intending are not reproducible with beliefs and desires. For some arguments the claim could be more accurately put by saying that the necessary conditions of intending don't match those of beliefs and desires. This class would include McCann's four differences between desires and intentions (McCann 1986, p. 126–127); McCann's argument on how desires can be in conflict with one another, but intentions cannot (McCann 1986, p. 194–195); Bratman's initial challenge of the characteristics of commitment in future-directed intending (Bratman 1987, p. 16–17) and his Buridan cases (Bratman 1987, p. 22–23); and finally Setiya's belief about doing (Setiya 2008, p. 394).

The third class of arguments focuses on agency and how belief-desire accounts misrepresent it. This class includes Velleman's agency-driven model and the accompanying critique of belief-desire models (Velleman 1992b, p. 461–463).

The fourth class of arguments assert that the kinds of desire and belief that are needed to model intentions reductively are incorrect or that merely that the concept of desire as it is used in belief-desire theory is not plausible. This includes Sellars's argument about pro attitude as an omnibus term for long-term dispositions (1973, p. 190); Brand's critique of beliefs (Brand 1984, p. 147) and Velleman's "guise of good" critique of desire (Velleman 1992a, p. 15–17).

These arguments are often accompanied by examples that aim to show that it's possible to do one without the other, but at the same time most of these examples seem to give rise to both reductive and nonreductive intuitions and interpretations. This classification is not supposed to be definitive and not all arguments fall neatly into one group. However, it provides a heuristic framework and points to the possibility of a more systematic analysis.

6 Humean responses to nonreductivist critiques

In this section I will discuss Neil Sinhababu's responses to critics of reductive belief-desire accounts and argue against some of the notions in his Humean belief-desire theory of intention.

6.1 Response to Mele: Two lines of defense

Ridge also presents a response to Alfred Mele's (1992b, p. 155–156) example against belief-desire accounts by showing the disconnect between predominant motivation and intending that was introduced in chapter 4. In the example Alan is predominantly motivated to humiliate Carl later today at a party and based on previous experience believes that he will probably humiliate Carl at the party, but still decides to resist his desire to humiliate Carl. Mele's argument is that if Alan decides against his desire to humiliate Carl, he cannot intend to humiliate Carl. The belief-desire analysis however seems to imply that Carl intends to humiliate Carl as he has the predominant desire to do that and believes that he will. Hence the belief-desire analysis is mistaken.

Mele notes that the fact that Alan "is more motivated to try to master his desire to humiliate Carl than he is to make no such attempt" (Mele 1992b, p. 156) is an important factor. This is what Michael Ridge (1998) utilizes in his two answers to Mele's problem. His first answer aims to show that the example doesn't pose a problem for his account, if we separate first-order pro attitudes from second-order pro attitudes. Ridge doesn't define the term, but Alfred Mele has defined a higher-order desire as "a desire whose representational content encompasses a representation of another actual or possible desire of the person whose desire it is" (Mele 1992a, p. 238), in short second-order desires are desires about first-order desires. Unlike Ridge, Mele doesn't mention higher-order intentions at all.

Ridge claims that the fact that Alan has formed a second-order evaluation, desire or intention that is "at odds with his first-order intention to humiliate Carl" is an explanation that is "perfectly compatible with [Alan's] intending to humiliate [Carl]" (Ridge 1998, p. 168–169). It does seem easy to understand why a conflicting second-order evaluation or desire wouldn't be a problem for belief-desire accounts, as there is nothing wrong with conflicting evaluations or desires. However, I believe second-order intentions would be problematic if the first-order intention is retained, since having conflicting intentions would seem to violate an important rationality requirement of intentions. Ridge doesn't define second-order intentions or go into detail why having a second-order intention wouldn't lead to a situation where Alan has conflicting intentions or why that would not be a problem for his account.

Ridge does considers the possibility that the first-order intention about insulting Carl might be dismissed when the second-order intention about the intention of insulting Carl is formed, but he does this in the context of another argument. He also correctly notes that this would make Mele's example moot (Ridge 1998, p. 169). As there is a more charitable interpretation available, this interpretation doesn't seem relevant.

Ridge's second line of defense is that if Alan would not succeed in resisting his desire to humiliate Carl and thus humiliated Carl at the party, claiming that he had no intention to humiliate Carl would not be taken seriously as no-one would believe him (Ridge 1998, p. 169). If this is true in the sense that it would be after all unintuitive to grant that Alan did not have an intention to humiliate Carl, it seems to suggest that the example is somehow flawed. However, the example only assumes that Alan did not have the intention at the time, well before meeting Carl at the party. That Carl had no intention to humiliate Carl that morning would probable be plausible, but irrelevant for the party-goers. Another possibility not explicitly noted by Mele or Ridge is that contrary to Mele's intuition Alan actually does have two conflicting intentions and is irrationally conflicted. Yet another possible defense is similar to the one we gave for Hugh McCann's argument against reductionists in chapter 4. It might be that sometimes we simply cannot consciously decide what to intend against our strongest desires. When Alan utters in his mind the decision not to humiliate Carl, there is a possibility that he does not form an intention not to humiliate Carl in a similar fashion that a person declaring an intention to quit smoking might not actually intend to quit smoking. If this is correct, the example is flawed in suggesting that Alan does not intend to humiliate Carl.

Mele's argument seems to based at least partially on the idea that as creatures with free will we can decide anything even against our strongest desires. This is not necessarily so. There is no agreement after all that it is possible to decide to act against our strongest desires. Having a conduct-controlling intention without the strongest desire at the time could be one way for an agent to decide to act against our strongest desires, but presuming the existence of such intentions seems like begging the question as the conclusion would already have to be a premise. Even if we decide to do something totally at random and follow through, it doesn't seem too far fetched to assume that our strongest desire at the time of the decision would be to do something at random.

6.2   Response to Setiya: Belief about doing

We saw in a previous chapter that Kieran Setiya's argument against belief-desire models was that they cannot explain why belief about doing is a necessary truth for intending. While we noted that this argument does not concern all belief-desire models as some of them (e.g. Audi 1973) incorporate belief about doing, it does concern Sinahbabu's model. I'm not going to argue forcefully for Setiya's conception of belief about doing, rather I will argue against Sinhababu's counterexample. To refresh our memory, let's take another look at Setiya's principle of belief about doing:

> "If A is doing Φ intentionally, A believes that he is doing it or is more
> confident of this than he would otherwise be, or else he is doing Φ by
> doing other things for which that condition holds".
> (Setiya 2009, p. 129)

Sinhababu argues that belief about doing is not a necessary feature of intentional action, but rather a contingent though seemingly ubiquitous correlate of intentional action. The fact that we usually hold beliefs about doing is because to intentionally Φ we must have the desire to Φ, and the attentional aspect of that desire usually makes us infer that we're acting, but not always (Sinhababu 2013, p. 684).

To support this claim, Sinhababu provides an original example of intentional action for a reason without belief about doing (Sinhababu 2017, p. 90–91): Bridget tends to bounce unintentionally all the time. She is tricked by Zur the wizard to believe that if she ever tries to bounce intentionally, Zur's spell will make her not actually bounce, but only appear to bounce. While this is not true, she is 99% sure of it. She also believes that if she does not bounce, she will be eaten by the demon Yawgmoth, summoned by Zur. She has a desire to survive and believes that she can only survive by bouncing. Believing all this makes it impossible for her to continue bouncing unintentionally and thus she tries to bounce intentionally. Because Zur lied about the spell, Bridget does actually bounce, but she still believes that with 99% certainty she doesn't bounce. Thus Bridget is bouncing intentionally while believing that she is almost certainly not bouncing. Furthermore, as she knows of her tendency to bounce unintentionally, she can also infer that her intentional bouncing has made her bouncing less likely than if she hadn't formed the intention to bounce[5]. This means that Setiya's principle, belief about doing as a necessary condition of intending, is false.

How to make sense of this? According to Sinhababu's own account of intending, to intend to Φ means that the agent must have a belief "that A's B-ing in S would

---

[5]In the example Sinhababu assumes that Bridget both bounces intentionally and intends to bounce.

make Φ more likely" (Sinhababu 2017, p. 100–101). This seems to hold for Bridget's intention to survive as her bouncing would make her survival more likely. It also seems to hold for Bridget's intention to bounce, as her bouncing would make her bouncing more likely compared to not bouncing, assuming she has already stopped bouncing unintentionally.

Next we must ask if there's a way out. We can start by going over Setiya's anticipated answer to critiques of belief about doing of the same form as Sinhababu's example: "Couldn't there be an action that is normally automatic but which can be done intentionally with a lower chance of success? If one is aware of all this, one will be, on balance, less confident that one is performing that action when one is doing it intentionally — but still more confident than if one were not doing it intentionally and one's automatic system were shut down" (Setiya 2008, 391). Setiya suggests here that when comparing whether or not the agent is "more confident of [doing what they intend] than he would otherwise be" (Setiya 2009, p. 129), the relevant baseline for comparison is not the situation where one does the action automatically or unintentionally, but the situation where one's automatic system is shut down.

Sinhababu doesn't accept Setiya's anticipated answer: "It's hard to see how to modify his formulation to explicitly shut down automatic processes without rendering it trivial or ad hoc" (Sinhababu 2017, p. 91). However, it seems to me that in the Bridget example there would be no need to modify the example to include a shut down of automatic processes, because failing to bounce unintentionally is already part of the example: it is what prompts the intentional act of bouncing. This seems to imply that we should use the situation where Bridget stops bouncing unintentionally as the baseline when checking whether or not belief about doing holds for Bridget. If we do, we can see that when bouncing intentionally Bridget is 1% more confident of bouncing than she would be if she didn't bounce unintentionally and thus has belief about doing as per Setiya's definition.

There is another way to look at the example that might salvage the conclusion that Bridget's intending to bounce has made her bouncing less likely than if she hadn't formed the intention. What if Bridget's intention to bounce isn't based on the fact that she has stopped bouncing and subsequently the belief that bouncing intentionally will make her survival more likely, but rather what stops Bridget from bouncing unintentionally is the intention to bounce? Her desire and subsequent intention to bounce could still be based on the belief that it will aid her survival, it's just that that belief is mistaken as without that intention she would continue bouncing unintentionally. Only after forming the intention does she realize that bouncing intentionally has actually made her goal of bouncing and survival more unlikely. Had she realized that earlier, perhaps she could have tried to distract

herself and possibly continue bouncing unintentionally.

Then again, even with this interpretation Bridget does not seem to simultaneously bounce intentionally and believe that bouncing intentionally would make her bouncing less likely than not intending to bounce. After Bridget realizes that her bouncing intentionally has made her goal of survival more unlikely, her automatic system is already shut down and she should be comparing her odds of survival while intentionally bouncing with the situation where she gave up the intention now. Thus it turns out that this interpretation doesn't support Sinhababu's point either.

What if Bridget ends up being less confident about her bouncing because she is confused by the possible interpretations of her situation, like the reader might be? This makes no sense either as the question is about whether or not the agent fulfills the conditions of belief about doing and not whether or not the agent knows that she fulfills those conditions. If this analysis is correct, it's still possible that belief about doing is a contingent correlate of intentional action, but Sinhababu's counterexample doesn't necessarily show that.

6.3   Response to Bratman: Commitment and Buridan cases

Sinhababu's second response is to the critique that Michael Bratman (1987) has presented against reductive belief-desire accounts. What Sinhababu calls Bratman's initial challenge for belief-desire accounts is that they lack the four features of commitment that are crucial characteristics of intention: intentions are conduct-controlling, nonreconsiderable, they exclude other conflicting intentions and they dispose us to reason further (Sinhababu 2017, 104–105). According to Bratman these characteristics are deeply conjoined, closely related and explain more when taken together (Bratman 1987, p. 17). Thus treating them as completely separate arguments as Sinhababu does (Sinhababu 2017, 104–110) might not do full justice to Bratman's account.

Analyzing this discussion gives insight into how far the traditional "being settled" critique by Michael Bratman and Alfred Mele can be pushed against recent Humean belief-desire accounts. I will go through Sinhababu's responses one by one and elucidate Bratman's critique at the same time.

6.3.1   Conduct-controlling disposition

Sinhababu claims that the conduct-controlling nature of intention, which Bratman calls the volitional dimension of commitment, is easily explained: "[belief-desire

pairs] are sufficient for motivation, and thus conduct-controlling[6]. If there's some problem with belief-desire pairs controlling conduct, Bratman hasn't told us what it is" (Sinhababu 2017, p. 105).

Sinhababu seems to believe that Bratman is arguing against a different position altogether, one that has to do with desires and not belief-desire pairs. I think that would be is a misrepresentation of Bratman's argument, or at least a more sympathetic reading is more plausible. In the passage Bratman argues against the "reduction of intention to predominant desire" (Bratman 1987, p. 19). In the same passage Bratman considers predominant desires together with background conditions, including belief. Bratman's notes that adding a further belief condition, even a belief that one will act on the intention, does not guarantee that the issue is seen as settled. It's also noteworthy that Sinhababu's account of intention doesn't rely on any belief-desire pairs, but specifically combined belief-desire pairs (Sinhababu 2017, p. 100–101).

So it seems that Bratman is after all talking about belief-desire models. Bratman also states the problem relatively clearly: belief-desire pairs are not conduct-controlling, because they might not settle the issue for the agent. This is because beliefs and desires do not guarantee the same kind of commitment that intentions do. The underlying problem with other features of commitment is very similar if not the same: even when an agent has a predominant desire and a set of relevant beliefs, they might not see the issue as settled.

Even though Bratman's problem with conduct-controlling belief-desire pairs seems clear and it seems to be directed against actual belief-desire accounts rather than a strawman position of predominant desires without beliefs, I believe that the first part of Sinhababu's explanation still provides a kind of an answer to the problem. In his account motivation stems from belief and desire by definition and what is predominantly motivating is by Sinhababu's definition conduct-controlling. At the same time it is noteworthy that these definitions are not shared and the account and the way it is presented may be controversial even when the account itself is internally coherent. I will return to this point in the end.

6.3.2 Nonreconsideration

Bratman calls nonreconsideration the fact of being settled or the disposition to retain the intention (Bratman 1987, p. 16). Bratman's idea is that intentions have

---

[6]What is sufficient for motivation is not usually thought to be sufficient for controlling conduct, because we can be motivated to do a lot of things that we don't intend to do. Sinhababu probably means that some of those belief-desire pairs still are conduct-controlling, those that generate the most motivational force.

a sort of inertia that implies that the issue has been settled. Sinhababu argues that nonreconsideration doesn't require a robust explanation because "[d]esire and belief aren't states of mind that we constantly reconsider. When they combine, we don't constantly reconsider the combination" (Sinhababu 2017, p. 106).

The question remains: are beliefs and desires together as stable as intentions are? The way I understand Bratman's argument is that while we might not constantly reconsider our desires, they nevertheless can and do change more easily than intentions. If we were to say that the belief component is what stabilizes intentions it would be unclear how the belief would be able to retain or stabilize the whole combined belief-desire pair, especially as in the Humean tradition beliefs generally have no power over desires. If intentions were belief-desire pairs, reconsidering the desire would presumably change the intention, thus making the account vulnerable to Bratman's critique. Sinhababu doesn't explain how combined belief-desire pairs might behave in these kinds of situations, but if they behaved as a single unit it would be unclear what separates combined belief-desire pairs from independent intentions.

Sinhababu's way around this critique is that intentions and desires aren't tied to the principle of nonreconsideration as tightly as Bratman thinks (Sinhababu 2013, p. 696, n28). Sinhababu argues that some intentions are more stable than others because our desires vary in strength under different conditions. For example our desire for food can rapidly change in strength due to our "internal biological conditions like the fullness of our bellies" (Sinhababu 2013, p. 693). As Sinhababu explains in another example, a person might change her intentions about which drink to get several times between leaving their table and reaching the bar (Sinhababu 2013, p. 691–692).

Indeed, in Sinhababu's intentions account are not very stable. It seems that according to Sinhababu's view to hold future intentions from one point of time t_1 to another t_2, the contrafactual condition must be true from t_1 to t_2, i.e. the agent's beliefs and desires be able to produce sufficient motivational force for the agent to initiate action immediately given that they believed that the relevant situation S obtains now. If at some point of time the agent's beliefs and desires wouldn't produce sufficient motivational force, the agent no longer holds the intention. In this sense the account is far from stable. This is something that might be intuitive for others and counterintuitive for others, as I'm sure noreductionist belief-desire-intention theorists would counter that the bar customer is not settled on a drink and thus has no intention before they make up their mind.

Sinhababu has another example against Bratman's stability condition: Susie's premeditated breakup that goes wrong when she reconsiders her intention at the

moment she sees her lover even if she knew beforehand everything there is to know about the situation (Sinhababu 2017, p. 108–109). Sinhababu's explanation is that a belief-desire pair to stay together rises in motivational strength and takes the place of the previously predominant belief-desire pair to break up tonight. However, we might ask if this is really reconsideration in Bratman's sense or revocability that Bratman argues is distinct from reconsideration (Bratman 1987, p. 16). It might also be plausible that Susie's intention to break up had certain inertia as long as it lasted and that she was disposed to retain the intention. Intentions can fail even if we had all the relevant knowledge beforehand.

6.3.3 Exclusion

Exclusion concerns the disposition to constrain further intentions in light of prior intentions, in other words we tend to exclude conflicting intentions (Bratman 1987, p. 17). The explanation Sinhababu offers for exclusion[7] is that we often hold beliefs about our future actions and these beliefs constrain our thinking so as to exclude the formation of further conflicting intentions. As we have already seen, Sinhababu sees belief about future action as not a necessary feature of intentional action, but a contingent though ubiquitous correlate of intentional action (Sinhababu 2013, p. 682–685).

This stance seems problematic. If this were how exclusion works, those intentions that don't satisfy a belief condition either wouldn't exclude conflicting intentions or some another mechanism would have to be in play. In the Bridget example Bridget supposedly both intends to bounce and doesn't believe that she will bounce. If she was weighing her options for future action, her intention to bounce shouldn't exclude intentions that would be in conflict with her intention to bounce, or the exclusion would have to be explained. Let's assume that one cannot run fast while bouncing and thus intending to outrun the monster is in conflict with intending to bounce. Could she really form the intention to run with no regard to the intention to bounce? Probably not. On the other hand, Bridget might still have some beliefs about her future action: that she will initiate the act of bouncing, even if she believes that she will probably fail. This belief or something like it might explain exclusion in this case.

---

[7]Sinhababu categorises exclusion and nonreconsideration as two cases of the disposition not to deliberate (Sinhababu 2017, p. 104), but it is not clear why this is so. In Bratman's account exclusion seems to be related to the fact that intentions have a role in further reasoning. So I believe it is rather the disposition to reason further that is divided into two: the disposition to engage in further means-ends reasoning and the disposition to constrain further intentions in light of prior intentions (Bratman 1987, p. 17).

6.3.4 Disposition to reason further

Bratman argues that when we intend, we are disposed to reason from prior intentions to further intentions. Bratman also calls this the demand for means-end coherence, one of the two demands on plans and intentions. Even though plans are for Bratman characteristically partial, they must be filled in as needed for their successful execution before it's too late (Bratman 1987, p. 31).

Sinhababu explains this feature by reference to desire's attentional aspect: our desires direct our attention to things that are relevant to the desires, e.g. their causes and effects (Sinhababu 2017, p. 33, 109). The attentional aspect makes even our desires inputs to deliberation (Sinhababu 2017, p. 109).

6.3.5 Buridan cases

Finally, something could be said about Buridan cases. I'm not aware of direct answers to the kind of critique presented by Michael Bratman we saw in chapter 4, but it seems to me that Buridan cases might not present unsurpassable problems for Sinhababu's account. We had two problems: how to choose between route 101 and 280 in the belief-desire framework and how to let our prior intention to take precedence when approaching a crossroads.

In Sinhababu's account the belief-desire agent would simply form a further desire to take route 101, which would direct the agent's attention to combine the desire with the relevant beliefs so that they can satisfy the contrafactual condition. The original belief-desire reasons would remain of equal weight, but the agent now has a new reason to take route 101: their desire to take route 101 and thus be more motivated to take route 101. The exact psychological process that leads to the desire to take route 101 should warrant no more explanation than forming the nonreductionist intention to take route 101. Perhaps the desire would be due to practical reasoning, i.e. combining beliefs and desires to produce instrumental desires (Sinhababu 2017, p. 83), perhaps some lower-level psychological process such as being slightly more attuned to things on the dominant hand side. The agent could also be sufficiently motivated to make an arbitrary choice as they have a desire to reach San Fransisco. Bratman seems to think that choice in Buridan cases is something that is not explainable on the level of desires, beliefs or intentions anyway: "[w]hich box of Cheerios I choose may not be explainable by my other desires, beliefs, and intentions; though it will, presumably, be explainable at some other — perhaps, neurophysiological — level" (Bratman 1987, p. 11).

The second problem is trickier. For Bratman, the intention to choose route 101 is a reason over and above my relevant belief-desire reasons — they don't even compete

against each other. This can be viewed in the light of Bratman's initial challenge for reductive belief-desire models that involves the four characteristics of commitement in chapter 4. This seems to be a much stronger conception of intending than current belief-desire accounts would allow. One option might be to say that the intention to take route 101 would subsequently be based on the newly acquired desire to take route 101 rather than the reasons for taking route 101 that are of equal weight with the reasons for taking route 280. Taking into account also the limits to our attention (Sinhababu 2017, p. 86) this kind of a belief-desire agent begins to seem somewhat plausible. It could also be said that the second problem is not as dangerous for the belief-desire account as it relies on very specific intuitions about the inner workings of practical reasoning. In any case, it seems to me that it would be unnecessarily harsh to say that Sinhababu's belief-desire account couldn't solve Buridan cases.

6.4   Theoretical virtues

As we saw earlier, Sinhababu argues that his reductive belief-desire account excels over nonreductive belief-desire-intention accounts in two theoretical virtues: simplicity and explanatory power (Sinhababu 2013, p. 694; Sinhababu 2017, p. 110). I will look at the issue at the ligth of Michael N. Keas's (2018) account of theoretical virtues. My aim is not to compare the virtuousness of reductionist and nonreductionist accounts to show that one is more virtuous than the other, but to examine if there's reason to side with the Humean reductionists based on Sinhababu's argument.

Keas distinguishes twelve major virtues of good theories: evidential accuracy, causal adequacy, explanatory depth, internal consistency, internal coherence, universal coherence, beauty, simplicity, unification, durability, fruitfulness and applicability (Keas 2018, p. 2762–2763). For the purposes of this discussion I will bundle evidential accuracy (fit with empirical evidence), causal adequacy (causal factors produce the effects in need of explanation plausibly), explanatory depth (causal history depth, range of counterfactual questions, etc.) and universal coherence (fits other warranted beliefs) together under accuracy following Kuhn (1977) instead of considering them separately.

Internal consistency (lacks internal contradictions) and internal coherence (components form a plausible whole that lacks ad hoc hypotheses) of the theories are evaluated rather thoroughly in the literature and in this thesis. As the issue is contested, I won't argue the issue further here. I will also leave aside beauty (evokes aesthetic pleasure) as it has no epistemic content apart from its association with simplicty and unification (Keas 2018, p. 2772). Past durability (has survived test-

ing) perhaps shouldn't affect our current evaluation much, but it could be granted that nonreductive accounts have fared fairly well in this regard. For example Sinhababu accepts that "[d]esire-belief accounts of intention are no longer widely accepted, largely because of [Michael Bratman's] influential arguments" (Sinhababu 2017, p. 100). The situation is of course somewhat skewed as Humean accounts of intention (1998) and Sinhababu's account in particular (2013; 2017) are much more recent than e.g. Bratman's nonreductive account (1987).

Sinhababu explicitly appeals to simplicity (explaining the same facts with less theoretical content), which he calls theoretical parsimony. In Keas's terms he could also be appealing to unification (explaining more kinds of facts with the same amount of theoretical content) to some extent as his reductive goals seem to be unificatory. Two important theoretical virtues are still left: fruitfulness (if the theory has generated additional discovery) and applicability (if the theory has guided science-based technology or similar).

Sinhababu's argument seems to be that as belief-desire accounts do not appeal to intention as an independent notion, they are ontologically more parsimonious by utilizing one mental state less than the nonreductionists (Sinhababu 2013, p. 694; Sinhababu 2017, p. 110). The goal of theoretical parsimony is replacing superfluous terms: if we can leave something unnecessary out from our account, we should. As folk psychological terms don't map well into lower-level (e.g. neurological) terms and as intentions have many points of contact with beliefs and desires, seeing intentions as superfluous in the way reductive belief-desire accounts do is certainly a possibility, even though there are evidently good nonreductionist arguments against this as well.

On the other hand, theoretical parsimony or simplicity need not be psychologically realistic and as such could reduce from the accuracy of the account. Michael Bratman has famously argued that intentions as plan-states help us to navigate the world despite our resource-boundedness, i.e. our limited capability to reason, and the changing environment. The more we deliberate, "the more chance there is that the world will change in important ways — ways that will undermine the very assumptions on which the deliberation is proceeding" (Bratman et al. 1988, p. 349). Calculating the best way to act as a function of the agent's beliefs and desires might be plausible for an idealized resource-unbounded agent, but not for humans. The cost of theoretical parsimony is added complexity in our mental processing. This is something that Sinhababu doesn't address.

Second, Sinhababu argues his reductive account has more explanatory power as it explains why the necessary conditions for intentions as described in Bratman's initial challenge don't always apply, i.e. why features like nonreconsideration and stability aren't as ubiquitous as Bratman says (Sinhababu 2017, p. 106–110). This

claim of course depends largely upon whether or not Sinhababu's criticism of non-reconsideration and stability are justified.

Looking at Keas's account we can see that in addition to theoretical parsimony and explanatory power there are other relevant theoretical virtues, such as internal consistency and coherence, accuracy, fruitfulness and applicability. Since Sinhababu's account is supposedly an account of how intention is psychologically realized in human beings (Sinhababu 2013, p. 682), psychological realisticness or accuracy is definitely one important virtue that should be weighed against theoretical parsimony. Michael Bratman has argued that we should judge theories by their theoretical fecundity or fruitfulness (Bratman 1987, p. 20). This might be hard to evaluate, but the fruitfulness and applicability of belief-desire-intention accounts is hard to dispute (e.g. Bratman et al. 1988; Bratman 2009, 2014), but maybe the same goes for belief-desire accounts if we think of them as connected with decision theory. Perhaps there is a case for pluralism here.

For those who are not familiar with decision theory, it deals with the way agents make decisions. The central concepts of decision theory and belief-desire accounts are very similar: in decision theory preference is a theoretical concept somewhat similar to volitive desire. Beliefs are expressed as probabilities. In orthodox decision theory agents prefer outcomes that have the greatest expected value and maximize their expected utility in similar fashion to a belief-desire agent that acts according to their predominant motivation. While the endeavours are similar, there is no direct link from most reductionist accounts to decision theory, some belief-desire theorists even criticize decision theory (e.g. Sinhababu 2017, p. 112).

6.5 Sinhababu's theory of motivation again

When first looking at Sinhababu's account it almost seems as if the old problem of predominant motivation doesn't apply to it. Predominant desires or predominant motivation do not feature in Sinhababu's formulation of intending explicitly, but rather their function is performed by the contrafactual condition. Sinhababu still has a theory of motivation that is very similar to old accounts of intending and is based on predominant motivation, or what "one is most motivated to do" (Sinhababu 2017, p. 23). This theory of motivation is relevant to intending too, as motivation features in the contrafactual condition of Sinhababu's account of intending: "(3) If the desire were combined with a belief that S obtains now, they would without further reasoning produce motivational force sufficient for A to immediately initiate B-ing" (Sinhababu 2017, p. 101). We know that combined desires and means-ends beliefs produce motivation, but what is the relationship between motivation and

action in Sinhababu's account?

In Sinhababu's earlier paper the motivational aspect of desire includes the notion that "[i]f at any time there is some action that they are the most motivated to do, they will initiate that action" (Sinhababu 2009, p. 468). The same notion also applies to his recent work, but only as a background assumption: "[m]otivation is the disposition to act. ...If at some moment there's an action that one is most motivated to do, one does that action" (Sinhababu 2017, p. 23).

Note that this seems to directly concern only acting *now* as there is no similar formulation for future-directed intending. In other words, Sinhababu doesn't explicate whether or not having predominant motivation that Φ will directly lead the agent to have a future-directed intention that Φ.

On the other hand Sinhababu's belief-desire theory of action implies that desires combined with means-ends-beliefs produce motivation (A stands for action and E for end): "[o]ne is motivated to A if and only if desire that E is combined with belief that one can raise E's probability by A-ing" (Sinhababu 2017, p. 2). In the contrafactual condition the sufficient motivational force is produced only when the combined belief-desire complex is further combined with the belief that the situation S obtains now.

This indefinite relationship between the contrafactual condition and predominant motivation leads to two questions: first, is Sinhababu's contrafactual condition doing any actual philosophical work or could it be replaced by putting the traditional relationship between predominant motivation and action more explicitly? Second, if the contrafactual condition does actual philosophical work that is independent of predominant motivation and presuming that the contrafactual can involve a multitude of psychological states and processes, is this is deep down a reductive account at all?

To pose the latter question differently: is it possible to understand the contrafactual condition as a way to position desires and beliefs in the agent's not necessarily belief-desire reductive psychology? In this sense it seems that we might be blackboxing parts of intentionality and the account might not be reductive. A nonreductivist might allow that desires and beliefs specified by Sinhababu are necessary conditions of intention and even sufficient for intending if they are situated in the agent's psychology in a way that would lead to action. Things are even better if we grant that the beliefs and desires need to be combined to form intentions.

In an earlier paper Sinhababu seems to suggest that this is exactly what the contrafactual condition does: "(3) rules out cases where the agent desires something and has a belief about how to attain it, but hasn't yet combined these mental states and formed an intention. In these cases, intention-formation requires a further

step of practical reasoning in which the desire and belief are combined so they can motivate action" (Sinhababu 2013, p. 682). If we can take seriously Sinhababu's justification for the contrafactual condition (3) as ruling out cases where desires and beliefs are uncombined, it seems probable that the contrafactual condition doesn't do much philosophical work as Sinhababu has since added combining explicitly into his account of intending.

Returning to the former question, what would happen if we were to understand the contrafactual condition as a way to position desires and beliefs in the agent's necessarily belief-desire reductive psychology? I suggest that if the contrafactual condition could be replaced with a traditional model of predominant motivation, the account would be vulnerable to old critiques of predominant motivation. There seems to be a real possibility that Sinhababu's account faces a dilemma: either yield at least partially to the nonreductivist view or end up with a slightly more traditional account of intending and face the old arguments that concern the disconnect between predominant motivation and intention, such as that of Alfred Mele in chapter 4, and the charge of utilizing an omnibus concept of desire, presented by Myles Brand in chapter 4.

One of the most promising parts of Sinhababu's work seems to be his analysis of desire. Sinhababu, unlike Audi and Davis, tries to explain how properties of intentions can be reduced to a properties of desires. There is one thing that seems to lack explanation. It seems that in some sense the reductionist accounts are forced to see desire as a very thin, but wide concept in the footsteps of Davidson and Ridge. If we want to claim that everything we do is based on desires, we are clearly not talking about desires in the appetitive sense. On the other hand, Sinhababu's case against many critiques is based on the thick, substantive properties of desire. It is not clear how desire could be used as both a general pro attitude and at the same time possess the many properties of appetitive desires. This case has been made earlier by many anti-Humeans (Sinhababu 2017, p. 22). Sinhababu doesn't offer a direct solution, but acknowledges the challenge and submits his Humean theory to empirical evaluation (Sinhababu 2017, p. 22). One possibility to take this critique further is to take a closer look at one aspect of desire (E stands for end):

> The Hedonic Aspect: Desire that E combined with increasing subjective probability of E or vivid sensory or imaginative representation of E causes pleasure roughly proportional to the desire's strength times the increase in probability or the vividness of the representation. (With decreasing subjective probability of E or vivid sensory or imaginative representation of not-E, it likewise causes displeasure.)
> (Sinhababu 2017, p. 28)

In other words, when we desire something, that desire causes pleasure when we imagine that desire. This conception of desire might seem thicker than thin volitive conception of desire that allows desires to motivate everything we do, e.g. allow us to desire unpleasant things. As explained in chapter 2, there are a multitude of things we desire in a volitive sense, but wouldn't call desirable in an appetitive sense: taking a son the the zoo reluctantly, committing one's spouse to a mental institution or filing for bankruptcy (Davis 1984, p. 46). How do these conceptions fit into the same theory of desire?

Let's assume that you're being successfully blackmailed to commit fraud. You're conflicted about committing fraud, but your predominant motivation falls on the side of going through with it. If you weren't blackmailed you would never commit fraud and you are very averse to the thought. In the volitive sense, you still instrumentally desire to commit fraud and that desire is stronger than the aversion. One might wonder how the agent might feel pleasure from vivid sensory representation of the fraud that he feels forced to commit? Wouldn't we rather feel bad in committing the fraud?

The answer lies in that "the properties of desire ...are best understood only as properties of intrinsic desire. The hedonic effects of instrumental desire don't add to the effects of the intrinsic desires from which they're derived" (Sinhababu 2017, p. 27). The motivation behind this is to see that "[i]nstrumental desires don't generate motivation independently of the intrinsic desires from which they're derived" (Sinhababu 2017, p. 27). This is to disallow long chains of instrumental desires to generate more motivation than short chains of instrumental desires.

In our example the instrumental desire to commit fraud generates no pleasure on its own, rather the uneasiness we might feel in committing the fraud is not only due to our aversion to being blackmailed but possibly also to our other desires to be honest, avoid crime, etc. Now it seems that while intrinsic desires in Sinhababu's account are more substantive and thick, instrumental desires are left somewhat thinner. This is not necessarily a problem for the account.

6.6 Further arguments against the contrafactual condition

There are still several issues that we haven't dealt with in relation to Sinhababu's account. In this section I will go over a few of them. One interesting issue is if Sinhababu's contrafactual condition could be vulnerable to Mele's critiques that involve surprisingly strong fears (Mele 1992b, p. 142–143). If one was afraid of snakes without realizing the intensity of the fear, one could arguably have a desire to touch a pet snake if someone brought one up to them. Yet, if the desire to touch a snake

was combined with a belief that a snake had been brought up to them now perhaps they could be instantly terrified and not motivated at all to even initiate action or at least there would be plenty of further reasoning involved. It seems plausible that someone like that could nevertheless intend to touch a pet snake without eventually even initiating the action.

One could argue that what is combined with the belief that a snake had been brought up to them would be a desire to stay away from the snake, not the desire to touch the snake. On Sinhababu's account it seems that there is no need for the desire to actually combine with the belief that situation S obtains now, instead it is enough for intending that if the desire were combined with said belief they would produce sufficient motivational force. This seems to follow Mele's intuition that someone could intend something they thought possible even if it weren't actually possible, so in this regard Sinhababu's account could to be safe from Mele's critique.

For some future intentions that are sketched up only partially it seems that assessing whether or not the contrafactual condition (3) applies is not an easy task. A teenager might intend that they be an astronaut when they grow up with no further thought put into the matter. If they believed that they were an adult now, would they be sufficiently motivated to immediately initiate being an astronaut or that they be an astronaut? For most people this probably seems like a bizarre question, while the intention should be perfectly understandable. The problem is that the agent might believe that they will make up the rest of the plan along the way and execute multiple instrumental intentions before S obtains.

Sinhababu seems to allow for these kinds of very partially sketched intentions as (for agent A, behavior B, situation S and action φ) "A need not characterize B or S precisely. An earnest college student may fervently intend to do something, someday, to reduce global poverty. All he must believe is that S will obtain at some time, and that B-ing then will make φ more likely" (Sinhababu 2017, p. 102) and "A's goal may just be performing behavior B" (Sinhababu 2017, p. 101). If we apply the same problem to Sinhababu's own example, we can ask what kind of a contrafactual condition should be satisfied if one intended to reduce global poverty somehow, someday? What would the combined desire to reduce global poverty and the vague belief that something can be done motivate the college student to do if they were to believe at the time of deliberation that the moment to act is right now? I suggest that they would begin reasoning further rather than immediately initiate doing something or reducing global poverty.

Yet another critique to look at[8] is causal interference of the contrafactual condition: "If the desire were combined with a belief that S obtains now, they would

[8]Suggested to me by Kirk Ludwig in conversation.

without further reasoning produce motivational force sufficient for A to immediately initiate B-ing" (Sinhababu 2017, p. 101). What if something happened between the "if" and the "then" of the contrafactual? If there was a possibility of a meteor striking the agent in between the desire and belief combining and them producing motivational force, wouldn't the contrafactual be rendered strictly speaking untrue? This seems to depend on the contrafactual describing a causal process that takes time. Instead, if the relationship between combining and motivating was not causal but constitutive and the only causal process was the actual combining, then Sinhababu's account should be safe as the starting point of the contrafactual is that the mental states have already combined. The agent need not immediately initiate B-ing either, since it's enough for intending that the motivational force is sufficient for the agent to immediately initiate B-ing, it should make no difference if they have no time to actually initiate action. When desires and beliefs are combined in the manner described by the contrafactual they produce motivation by definition. Even though Sinhababu doesn't describe combining in detail, this might be a good way to conceive its relationship with motivation if one wants to avoid counterexamples to contrafactuals.

7 Conclusions

The aim of this thesis has been to examine the arguments for and against the reduction of intentions to beliefs and desires. I have introduced a number of reductive and nonreductive accounts together with a large number of arguments for and against reduction and specifically examined the recent Humean answers to nonreductivist critiques. I've also brought up further questions about Neil Sinhababu's Humean account and examined some further possibilities for both defending and arguing against the belief-desire view through my own examples. Sinhababu's account seems to fare rather well against most critiques, but not unscathed. For example, the critiques by Kieran Setiya and Michael Bratman are left partially unanswered. Points in dire need of further clarification include the exclusion of conflicting intentions when the agent has no belief about doing, Sinhababu's renouncement of the disposition to retain the intention, i.e. nonreconsideration, the relationship between the contrafactual condition and Sinhababu's motivational theory, and how to think of future intentions that are sketched up only partially. On the other hand, the answers given to most other questions cleared up some ambiguities concerning e.g. combining, the use of contrafactuals and the concept of desire Sinhababu utilizes in his account.

If we now compare the reductive and nonreductive views based on the brief examination of their theoretical virtues and the examination of the arguments and

counterarguments for and against reduction, it seems that there is not much conclusive to say about the main question at hand: if intentions are reducible. This study might still give some indication on where to look next. While comparing classes of arguments I also noted that most of the examples seemed to give rise to both reductive and nonreductive intuitions and interpretations. This may indicate that the folk psychological concept of intending could be given several equally intuitive interpretations. Their characteristics, relationship and accuracy could be subject of further study.

There might also be a need for discussion about whether or not there are limits for the belief-desire reductive view. Pragmatically there should be nothing wrong with having multiple models of intending with different conceptual frameworks. However, as the discussion concerns folk psychological concepts whose ontological status is controversial to begin with, it's debatable how seriously we should take projects of ontological reduction. The possibility of conceptual reduction on the other hand seems to depend a lot on how far we're willing to stretch the concept of desire: some theorists prefer to use desire as a thin, dispositional omnibus term, whereas others are inclined to utilize many of the properties usually associated with appetitive desires.

Here a further systematic review of accounts of intending might be beneficial. Charting the properties underlying different accounts of intention, such as their conception of desire, belief about doing, the relationship between intending and intentional action, etc., might reveal if there are some sets of properties that are especially plausible and which combinations are clearly too weak or too strong. In other words, what it takes to provide a plausible account of intending.

Bibliography

Anscombe, G. E. M. (1958). *Intention.* 2nd edition.

Aristotle (1934). *The Nicomachean ethics.* Harvard University Press, Cambridge, Massachusetts, 2nd edition.

Aristotle (1937). *On the motion of animals.* Harvard University Press, Cambridge, Massachusetts.

Audi, R. (1973). Intending. *Journal of Philosophy*, 70(13):387–403.

Brand, M. (1984). *Intending and acting: toward a naturalized action theory.* MIT Press, Cambridge, Mass.

Bratman, M. (1987). *Intention, plans, and practical reason.* Harvard University Press, Cambridge, Mass.

Bratman, M. (1999). *Faces of intention: selected essays on intention and agency.* Cambridge studies in philosophy. Cambridge University Press, Cambridge, U.K. ; New York.

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

Bratman, M. (2014). *Shared agency: a planning theory of acting together.* Oxford University Press, New York, NY.

Bratman, M. E., Israel, D. J., and Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4(3):349–355.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78(2):67–90.

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, 60(23):685–700.

Davidson, D. (1970). How is Weakness of the Will Possible? In *Essays on Actions and Events.* Clarendon Press ; Oxford University Press, Oxford : New York.

Davidson, D. (1971). Agency. In Binkley, R. W.; Bronaugh, R. N. M. A., editor, *Agent, Action, and Reason.* University of Toronto Press, Toronto and Buffalo.

Davidson, D. (1978). Intending. In *Essays on Actions and Events.* Oxford University Press, New York.

Davidson, D. (1980). *Essays on actions and events.* Clarendon Press ; Oxford University Press, Oxford : New York.

Davis, W. A. (1984). A Causal Theory of Intending. *American Philosophical Quarterly*, 21(1):43–54.

Dennett, D. (1991). Real patterns. *Journal of Philosophy*, 88(1).

Goldman, A. I. (1976). *A theory of human action.* Princeton University Press, Princeton, New Jersey.

Hampshire, S. (1959). *Thought and action.* Chatto and Windus, London.

Harman, G. (1986). *Change in view: principles of reasoning.* MIT Press, Cambridge, Mass.

Harman, G. (1999). Practical Reasoning. In *Reasoning, Meaning, and Mind.* Oxford University Press.

Hume, D. (1978). *A Treatise of Human Nature.* Oxford : Clarendon Press, 2nd edition.

Keas, M. N. (2018). Systematizing the theoretical virtues. *Synthese*, 195(6):2761–2793.

Kuhn, T. (1977). Objectivity, value judgment, and theory choice. In *The Essential Tension: Selected Studies in Scientific Tradition and Change.* University of Chicago Press, Chicago and London.

McCann, H. J. (1986). Rationality and the Range of Intention. *Midwest Studies in Philosophy*, 10(1):191–211.

Mele, A. (1992a). Akrasia, self-control, and second-order desires. *Noûs*, 26(3):281–302.

Mele, A. R. (1987). *Irrationality: an Essay on Akrasia, Self-Deception, and Self-Control.* Oxford University Press, New York. OCLC: 476008157.

Mele, A. R. (1988). Against a belief/desire analysis of intention. *Philosophia*, 18(2-3):239–242.

Mele, A. R. (1992b). *Springs of action: understanding intentional behavior.* Oxford University Press, New York.

Ridge, M. (1998). Humean Intentions. *American Philosophical Quarterly*, 35(2):157–178.

Searle, J. (1975). A taxonomy of illocutionary acts. In Gunderson, K., editor, *Language, Mind, and Knowledge*, pages 344–369. University of Minnesota Press, Minneapolis.

Sellars, W. (1973). Actions and Events. *Noûs*, 7(2):179–202.

Setiya, K. (2007). *Reasons without rationalism.* Princeton University Press, Princeton, N.J.

Setiya, K. (2008). Practical Knowledge. *Ethics*, 118(3):388–409.

Setiya, K. (2009). Practical Knowledge Revisited. *Ethics*, 120(1):128–137.

Sinhababu, N. (2009). The Humean Theory of Motivation Reformulated and Defended. *Philosophical Review*, 118(4):465–500.

Sinhababu, N. (2013). The Desire-Belief Account of Intention Explains Everything: The Desire-Belief Account of Intention Explains Everything. *Noûs*, 47(4):680–696.

Sinhababu, N. (2017). *Humean Nature: How desire explains action, thought, and feeling.* Oxford University Press, New York, NY.

Smith, M. (1995). *The Moral Problem.* Philosophical theory. Blackwell, Oxford, UK; Cambridge, Mass., USA.

Smith, M. (2012). Four Objections to the Standard Story of Action (and Four Replies). *Philosophical Issues*, 22(1):387–401.

Velleman, J. D. (1992a). The Guise of the Good. *Noûs*, 26(1):3–26.

Velleman, J. D. (1992b). What Happens When Someone Acts? *Mind*, 101(403):461–481.