Reasons for Action

Inauguraldissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Paulus Esterhazy
aus Bielefeld
2013

Erstgutachter: Prof. Dr. Johannes Hübner
Zweitgutachter: Prof. Dr. Stephan Sellmaier
Drittgutachter: Prof. Dr. Gerhard Ernst
Datum der mündlichen Prüfung: 1.2.2013

Reasons for Action

Paulus Esterhazy

The present book is the result of a PhD project in philosophy. The thesis was accepted in the 2012/13 winter semester by *Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft* at Ludwig-Maximilians-Universität München.

Gutachter: Prof. Dr. Johannes Hübner, Prof Dr. Stephan Sellmaier, Prof. Dr. Gerhard Ernst. Die Promotion erfolgte am 1. Februar 2013.

Author contact: pesterhazy@gmail.com

ISBN: 978-1-326-03768-0

Acknowledgements

This book is a revised version of my PhD thesis, submitted at the University of Munich. The completion of the thesis in 2012 was made possible through a scholarship provided by *Studienstiftung des deutschen Volkes*. Special thanks are due to Johannes Hübner, who, as my Doktorvater, supported the project, philosophically and with friendship. I benefited greatly from a research stay at the University of Pittsburgh; I'm grateful to Robert Brandom for his generosity in terms of time and patience. Thanks to Kasia Suchecka for the cover design, and to Stefan Brandt, Christine Bratu, Rachelle Esterhazy, Franz Knappik and Erasmus Mayr for their help and for their comments on earlier drafts. I also thank my family and, most of all, Anke Breunig for their continuing support.

Contents

Introduction

Reasons for action are ubiquitous in our thought and talk about what people do or should do. They explain, predict and justify our conduct. Without reasons, there would be no intentional agency. Yet for the longest time — with the exception of the perennial question whether reasons can be causes — contemporary philosophy of action has paid little attention to the nature of practical reasons. It is only in recent decades that practical reasons have emerged as a subject of philosophical interest in their own right. Reasons are considerations which motivate us and in the light of which we act. But just what is it that we attribute to a person when we credit her with a good reason? What sort of entity is on our minds when we deliberate about what we have reason to do?

Two comprehensive proposals — and two contrasting pairs of positions — shape the debate. In the first place, normative realism sees reasons as things or states of affairs in the world to which we respond when we act intentionally. On this view, the agent's reasons are largely independent of his subjective state. Psychologism denies this. It takes the opposing view that reasons for action are psychological states of the agent. A common psychologistic theory holds that when an agent acts for a reason, he is motivated by a belief and a desire which jointly rationalize his action.

In the second place, the 800 pound gorilla among theories of practical reasons is Humeanism. According to the Humean proposal, desires are the only source of our reasons. Furthermore, Humeans argue that practical reason as a faculty is powerless to assess our ends as irrational or rational unless they result from faulty causal or means-end reasoning. By contrast, rationalists or anti-Humeans assign practical reason a more extensive role, and they question the assumption that all our reasons are grounded in desires.

Some recent writers have championed normative realism as against psychologism while others have sought to defend Humeanism against ratio-

nalism. In this dissertation, I will argue that both normative realism and Humeanism fall short as comprehensive pictures of intentional agency. The first two chapters are devoted to the discussion of normative realism and Humeanism respectively. In the two chapters that follow, I develop and defend High Brow, an alternative view of agency that is superior to normative realism and, in particular, to Humeanism. In the final part, consisting of chapters 5 and 6, I proceed to putting High Brow to use. I take up important questions in the contemporary debate with the goal of showing that High Brow can help see the difficulties raised by these questions more clearly. If the conceptual tools developed in the course of the dissertation enable us to solve these difficulties, this may lend additional support to High Brow as a theory of reasons.

Throughout this dissertation, I make liberal use of ideas and approaches, distinctions and arguments that originate in the philosophy of Wilfrid Sellars. After fading into relative obscurity in the 1980's, Sellars's writings have recently regained a larger audience. The renaissance, however, has only been partial, being mostly restricted to his contributions to epistemology and the philosophy of mind. In addition to these fields, I draw on Sellarsian ideas from the philosophy of language and action theory as well. In particular, I rely on Sellars's inferentialist conception of semantic content and on his conception of intentions as the principal mental states of the practical realm. This thesis aims to highlight these lesser-known aspects of Sellars's view, though it should not be seen as an exercise in exegesis. Even if my view is not intended to represent Sellars's view, my positive suggestion, High Brow, is inspired by Sellarsian themes. I also rely heavily on Robert Brandom's elaboration and further development of the Sellarsian project.

Chapter 1 weighs the respective merits of normative realism and psychologism. I consider recent arguments for and against normative realism. As I argue, it is a general condition of adequacy of any theory of reasons that it must explain that judgments about reasons are related internally to motivation. Normative realism, as it turns out, fails to satisfy this condition. The upshot is that a theory of reasons must avoid the pitfalls of normative realism and psychologism alike while preserving the insights behind both views.

In chapter 2, I turn to a criticism of the Humean conception of agency, focusing on the claim that reasons are based on desires. Is it true, as Humeans hold, that wanting to do φ is sufficient for having a reason to do φ? Since what an affirmative answer would entail depends crucially on what exactly is meant by "wanting", I review a number of conceptions of desire that Humeans have traditionally relied on. Following a termi-

nology proposed by Peter Railton, we may say that the major Humean conceptions of desire are low-brow. According to this view — Low Brow — desires need not involve an evaluation of the object desired as good. The basic problem with Humeanism is that a low-brow interpretation of desire cannot support the far-reaching Humean principle that desires rationalize our conduct.

Criticism of Humeanism is hollow unless accompanied by an alternative. To supplant the Humean's Low Brow theory of desire, Chapter 3 motivates and develops a High Brow conception of agency. The key to understanding what it is for human agents to act for a reason — and the cornerstone of the High Brow view — is the concept of a practical commitment. In developing this view, I take my cue from Saul Krikpe's Wittgenstein-inspired critique of dispositional theories of meaning which, I contend, applies *mutatis mutandis* to dispositional theories of desires as well. I conclude that we should understand the principal mental state in the practical realm to be, not, as Humeans claim, desire but rather practical commitment, an essentially rule-governed, normative state.

As philosophers since antiquity have held, intentional agents necessarily act under the guise of some good. Despite its long pedigree, this doctrine — which is entailed by High Brow — has attracted criticism from various quarters. In chapter 4, I take up a number of arguments against the Guise of the Good. Some philosophers take issue with the assumption that acting for a reason is rule-governed activity. It may appear that acting for a reason cannot be accounted for in terms of guidance by a rule because any such account would generate an infinite regress. To counter this appearance, I explain in greater detail what it is to be guided by a conceptual rule. Next, I consider objections to the thought that intentional action invariably occurs with an evaluative view toward the good. I show that an agent may act on an evaluative stance without thereby acting reflectively by providing a High Brow account of practical reasoning.

Of the final two chapters, which apply High Brow to much-debated problems, the first focuses on two different theses commonly subsumed under the rubric "internalism": judgment-internalism and existence-internalism. While I agree with many writers that the former thesis is true, the problem is to show how High Brow makes sense of this intuitive truth. I undertake this task by showing how the upbringing of a rational being lays the foundation for the relevant conceptual connections. Next, I turn to Bernard Williams's internal reasons doctrine or existence-internalism. A review of his argument shows it to fail to provide independent grounds for believing the internalist conclusion. But Williams's presentation of the issue also contains valuable insights about our practices of reason-

ascription, which, I argue, High Brow is in a particularly good position to accommodate.

Finally, chapter 6 addresses the ongoing debate over the nature of so-called rational requirements. Two puzzles have emerged from the literature. First, it seems that principles of rationality such as the instrumental principle allow us to derive patently false conclusions about what reasons an agent actually has from seemingly harmless premises. Second, though it seems plausible that we must have some reason to be rational, it is hard to pinpoint the sense in which rationality can be normative. I argue that High Brow allows us to make headway toward a solution of both puzzles if we understand the process of practical reasoning and the role of rational requirements in it.

Chapter 1

Psychologism and realism

1.1 Psychologism and normative realism

When we act intentionally, we act for a reason. But what is a reason for action?[1] In one sense of the question, the answer is obvious. When we deliberate about what to do, we ask ourselves what we have reason to do. A reason, then, is the object of deliberation. Often in order to decide what to do, we weigh different reasons of varying strengths. We may not always do what we have most reason to do, yet we often let our actions be guided by our reasons, or what we take our reasons to be. We are sometimes wrong, sometimes right about our own reasons. As agents, we are interested not just in our own reasons but in the reasons of other agents as well. Every day, we take note of what other agents do, and when we do, we usually try to make sense of their doings in terms of reasons. The intelligibility of a person's actions depends on the possibility of discovering the reasons in the light of which he saw what he did when he did it. We ascribe reasons to others and justify our own conduct by revealing our own reasons. Reasons can be shared, but they can become the object of criticism as well.

We can, and often do, disagree about the existence of particular reasons for doing things, both our own reasons and those of other agents. Nonetheless, as agents we feel competent — even in the face of disagree-

[1] The topic of this dissertation is reasons for action, or practical reasons. Unless otherwise noted, I will use the term "reason" to mean practical reasons rather than reasons for belief, or theoretical or epistemic reasons.

ment — in our dealings with reasons in their various forms. It is not surprising that the various linguistic and nonlinguistic practices related to reasons for action are easy for us to follow, for the ability to do so is part of what it takes to be an intentional agent. We all know what reasons are, at least in the sense that we know our way around reasons. However, in a metaphysical sense, the question remains baffling: what kind of thing, what type of entity is a reason for action? To elaborate an answer to this philosophical question is to develop what we may call a theory of reasons. In this chapter, I compare and evaluate two theories of reason: psychologism and normative realism. After explaining the differences between the two competing views (§1), I consider fundamental arguments against the viability of normative realism (§§2–3). Next, I discuss a possible advantage of normative realism (§4). I continue by taking up the crucial point, i.e. the question whether normative realism can accommodate judgment internalism (§§4–5). I end the chapter by raising a final difficulty for psychologism (§6).

A prominent statement of psychologism, arguably the most commonly held theory of reasons today, is due to Donald Davidson:

> *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 2001a: 5)

According to Davidson, any action can be seen as the joint product of two intentional states of the agent: a pro-attitude and an instrumental belief. "Pro-attitude" is his catch-all term for all members of a heterogeneous class of desire-like attitudes, each of which constitutes a way of being attracted to an object or state of affairs.[2] Davidson conceives intentional action on the belief-desire model: roughly, whenever an agent performs an action for a reason, he desires some object, in the wide sense of having a pro-attitude towards it, and he believes that performing the action will help him obtain the object desired.

[2]Davidson includes under the umbrella term "pro-attitude" states as diverse as "desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed towards actions of a certain kind" (Davidson 2001a: 3–4). The Humean definition of the conative component of action will be elaborated in chapter 2.

For Davidson, to give the primary reason of a doing is to provide a full explanation or, as he calls it, "rationalization" by referencing a pair of psychological states. His view is paradigmatic for psychologism:

Psychologism Reasons for action are psychological states of the agent.

What does it mean to reject psychologism? Over the years, Davidson's view has been subjected to much criticism. First, a number of writers have taken issue with Davidson's contention that an action is bodily behavior caused by a belief-desire pair. Some writers have insisted that the agent's reasons stand to his action in a non-causal relation, while others have argued that it is in the agent himself, rather than in his intentional states, that we find the real causes of intentional action. Second, some philosophers disagreed with Davidson over the composition of primary reasons. Whereas on Davidson's view, in order to move an agent a reason must always contain a conative component as well as a cognitive one, some philosophers have insisted that a belief may suffice to motivate an action. Although both of these objections are interesting in their own right, they do not constitute ways of opposing psychologism. Critics of causalism argue that our beliefs or desires are not the causes of our action or that their causation of our behavior is insufficient to produce an intentional action, but they agree with Davidson that beliefs or desires are reasons for action. Similarly, pure cognitivists, as the proponents of purely belief-based reasons have been called, disagree as to the sufficiency of cognitive mental states to produce action, but they do not dispute Davidson's identification of reasons with mental attitudes.

Rejecting psychologism is the more radical supposition that reasons are not located in the agent's mental life at all. Recent writers have defended the anti-psychologistic view that psychological states, whether desire-like or belief-like, cannot, at least in normal cases, be reasons.[3] In one of the earliest explicit defenses of normative realism, Joseph Raz faces the question about the nature of reasons head-on.[4] He starts by mentioning three candidates for the job performed by reasons. As he notes,

[3]Psychologism has without doubt been the dominant position in analytic action theory at least since Davidson (2001a). In recent years, a growing number of philosophers have endorsed versions of normative realism as against the psychologistic alternative. Joseph Raz was one of the first writers to explicitly make the case for normative realism (Raz 1975). He further elaborates his position in his later (2002: especially ch. 2 and 3). For a more recent defense of realism, see Scanlon (1998: ch. 1) and Scanlon (2010). Bittner (2001) and Dancy (2000) are two recent monographs devoted to versions of what I call normative realism. Because Dancy's exposition contains the most worked out version of the view, much of the discussion in this chapter takes its cue from his book. Further defenses of normative realism include Parfit (1997) and Hyman (1999).

[4]Raz (1975: ch. 1).

reasons have been variously identified by different authors as statements, as beliefs or as facts. Raz immediately rules out statements as implausible candidates on the grounds that it would be highly unnatural to say that the reason for doing φ is the statement that *p*. The choice between beliefs and facts is more difficult. On Raz's view, reasons are to be conceived realistically:

Normative Realism Except for rare cases, reasons for action are not psychological items but facts or states of affairs in the world.[5]

To borrow a phrase by McDowell, normative realism holds that reasons do not stop short of the facts.[6] For normative realists, reasons are objectively out there in the world for the agents to discover. They hold that when Jim puts on his raincoat as the first drops start falling from the sky, his reason is not, as psychologism takes it, his belief about the weather but the non-psychological fact in the world that it is raining. His reaction is not to something happening in the domain of psychology but rather something that occurs outside the agent. Thus against the view that our reasons are constituted by beliefs, Raz writes:

> Beliefs are sometimes reasons, but it would be wrong to regard all reasons as beliefs. It should be remembered that reasons are used to guide behaviour, and people are to be guided by what is the case, not by what they believe to be the case. To be sure, in order to be guided by what is the case the person must come to believe that it is the case. Nevertheless it is the fact and not his belief in it which should guide him and which is a reason. If p is the case, then the fact that I do not believe that p does not establish that p is not a reason for me to perform some action. The fact that I am not aware

[5]The term "normative realism" is borrowed from Wallace (2006b), who calls a similar position "normative moral realism". Dancy, who expounds his anti-psychologistic views in a monograph with the title "Practical Reality" defends a kind of practical realism. Similarly, David Enoch calls his own view "robust metanormative realism" (Enoch 2007). The term "metanormativity" is apt because it derives from the view that moral realism, as defended by philosophers like G.E. Moore, is part of metaethics rather than normative ethics (Moore 1959). Accordingly, normative realism is a second-order view, one that remains silent on substantive questions as to the specific normative reasons agents have and instead confines itself to the question how we use expressions that ascribe reasons, what these expressions refer to metaphysically, their epistemological status, and so on. The topic of metanormativity is not just, for instance, the kind of objectivity or intersubjectivity found in moral judgments but normativity of reasons in general. The term "normative realism" is intended to convey a similarly broad scope.

[6]Cf. McDowell (1996: 33).

> of any reason does not show that there is none. If reasons are to serve for guiding and evaluating behavior then not all reasons are beliefs. (Raz 1975: 17)

In this passage, Raz makes two points against the view that reasons are, in typical cases, psychological states. First, psychologism seems to paint an implausible picture of how reasons are related to deliberation. Raz emphasizes as defining features of reasons their ability to guide our activity from the first-person perspective and their role in evaluating behavior. Explaining the agent's process of deciding what to do solely in terms of his own mental states conveys a mistaken picture of deliberation. What an agent saw in a given action is not confined to aspects of his inner life but includes the various features of his environment that could influence his choice. It would be solipsistic to think that the agent's sole guide of action is his own psychology. Intentional action properly speaking is a matter of being responsive to circumstances in the world rather than to a mere conception of them. Similarly, when we assess a person's actions, we see them as a reflection of events and objects in the world, rather than his own thinking. In Raz's eyes, the problem with psychologism is its blindness to the important evaluative or normative dimension of reasons. The world, not our conception of it, determines whether or not an action is justified.

For Raz, that deliberation and evaluation are guided by facts, not appearances of facts, shows that, at least most of the time, our reasons are facts. Raz's second point is that the existence of a particular reason does not depend on the agent's psychological state. To him, this contradicts psychologism. He seems to assume that a psychologistic view of reasons would be committed to saying that a person cannot have a reason without being consciously aware of having it, given that we are for the most part aware of our intentional states. Psychologism couldn't be true if it had this implication because we clearly are often unaware of the reasons we have, as for instance a child searching for hidden easter eggs may have a reason to look under the mailbox without having a clue that it is there he should be looking. As Raz's argument suggests, the fact that an agent is unaware of a reason to φ in no way shows that he does not have such a reason.

However, the two arguments in Raz's passage speak against psychologism only on a simplistic interpretation of the view. Consider a distinction between two types of reason-ascribing constructions. On the one hand, there are statements of the form "X has a reason to φ" or "That p is a reason for X to φ", on the other statements of the form "The reason

why *X* φ'ed was that *p*" or "*X*'s reason for φ'ing was that *p*"[7]. In the first kind of reason-statement, we ascribe *normative* reasons: we say that the agent has a good reason for φ'ing or that the fact that *p* counts in favor of doing φ. We do not necessarily mean that the reason compels the agent: good reasons can be overridden or defeated in some other way by different, stronger reasons. Nonetheless ascribing a reason in this way implies that there is something to be said for φ'ing, something that has weight on the scale of reasons, something that, other things being equal, makes it reasonable to φ.

By contrast, the second kind of reason-statement mentions, not necessarily what counts in favor of φ'ing, but the agent's reason for φ'ing — the consideration in the light of which the agent acted[8]. When we ascribe *motivating* reasons, we identify the consideration that was operative in the agent at the time of action. Motivating reasons are conceptually distinct from normative reasons. That *S* φ'd because *p* does not entail that the fact that *p* was a good reason for φ'ing. The ascription to an agent of a motivating reason to φ does not require that φ'ing was in fact reasonable in the light of the fact that *p*.[9]

Conversely, it does not follow from the fact that *S* has a normative reason *p* to φ that he in fact φ's or even that he is motivated to do so: just as motivating reasons do not in every case imply normative reasons, normative reasons do not in every case imply motivating reasons. For one thing, there may not be a good opportunity to φ or *S* may not notice that he has an opportunity; he may be distracted by other activities; he may have other stronger reasons than his reason that *p*. Importantly, *S* may not be aware of the fact that he has a normative reason to do φ. All these cases involve a gap between normative and motivating reasons.

This distinction shows why Raz's first argument is faulty. Psychologism may grant that, when deliberating, we do not examine our own inner mental landscape but look outside into the world to determine what we

[7]Not all statements of the form "the reason why X φ'd was that p" are reason-statements in my sense. Consider: "The reason why the building collapsed was that low-grade concrete was used in its construction." The building, not being a person, cannot perform actions intentionally, nor can it do anything for a reason. Evidently what is ascribed here is not an agential reason but a mere cause. To complicate things further, we sometimes ascribe mere causes to persons as well as to objects, as in "The reason why he tripped over the dog was that the streetlight was broken". The word "reason" as used in the text always refers to agential reasons in the emphatic sense, not simple causation.

[8]Cf. Dancy (2000: 2).

[9]Other writers use the term "explanatory reasons" instead of "motivating reasons". The term "justifying reasons" is common as an alternative to "normative reasons" or "good reasons".

have reason to do. But we deliberate about normative reasons, not motivating reasons. Psychologism need not argue that our normative reasons are psychological states. Making use of the distinction, it can agree with the normative realist that our normative reasons are located outside the agent but insist that our motivating reasons are beliefs or desires. On the psychologistic view, we are guided by the world as we are guided by our normative reasons, which are related to but distinct from our motivating reasons. However, the view goes on, when we explain an action in terms of its motivating reasons, we necessarily explain it in terms of the agent's psychology.

Equipped with the disambiguation, the defender of psychologism can also disarm Raz's second argument. It is true, he may say, that we are introspectively aware of our reasons, but only if we understand "reasons" to refer to motivating reasons. Of course an agent motivated by a reason knows of the belief or desire moving him. But this doesn't show that we have intimate introspective knowledge concerning what we have most reason to do. Normative reasons can diverge substantially from motivating reasons, so knowing the latter does not imply knowing the former. Psychologism holds that motivating reasons are beliefs and desires, but on a sensible interpretation, it does not hold the same view about normative reasons; it is free to interpret the latter in such a way as to explain why the child looking for easter eggs may have a normative reason to look under the mailbox without knowing that he does.

In other words, psychologism is best understood as a claim about motivating reasons.[10] Psychologism is not touched by the arguments found in Raz's passage, provided it comes with a satisfactory commentary which explains how it proposes to understand normative reasons. If it is to escape the simple counter-examples mentioned, psychologism must comment on the relation between normative and motivating reasons. In fact, it is clear that any theory of reasons must include an account of this relation. It has to answer two distinct but related questions:

1. What is the metaphysical status of motivating reasons or, alternatively, how should we understand statements about motivating reasons?
2. What is the metaphysical status of normative reasons or, alternatively, how should we understand statements about normative reasons?[11]

[10] Cf. Dancy (2000: 98).

[11] The twofold formulation is designed to leave open the possibility of answering either question by resorting to "semantic ascent", as is common practice in metathics.

Psychologism and normative realism give different answers to these two questions. The first question inquires about the *kind of item* we talk about when we ask what the reason was for which an agent acted. As we have seen, psychologism holds that motivating reasons are psychological states. For Davidson and many others, reasons necessarily include a conative state, although others hold that the reason behind an action may only consist of a cognitive state. Normative realism holds, instead, that motivating reasons are facts in the world or states of affairs. But a theory of reason should also tell us how to understand statements about good reasons for actions such as "The fact that it is raining is a reason for Tom to open his umbrella" and "Claire ought to sell her house because the real estate prices are falling." Normative realism does not only say that motivating reasons are facts in the world, it also posits the existence of facts of the form "X has a reason to φ". So its interpretation of statements ascribing normative reasons is straightforward: they are true if, as a matter of fact, it is the case that X has a reason to φ. Normative realism posits independent and at least potentially subject-independent reason-facts. By consequence, the view understands the act of coming to believe that one has a good reason to φ — that a reason-statement is true — as an act of grasping an independent reality.

Psychologism's answer to question (2), by contrast, is broadly reductionist. Whereas normative realism holds statements about normative reasons to enjoy independence from whatever particular motivations an agent has, psychologism explains normative reason statements in terms of motivating reasons. A general way of providing such an analysis is to paraphrase "X has a reason to φ" as a hypothetical statement about motivation.[12] In its account of normative reasons, psychologism is led by two considerations. First, it gives pride of place, in the order of explanation, to motivating reasons rather than normative reasons. Whereas normative realism focuses on the role of normative reasons in deliberation, psychologism bases its account on the third-person ascription of reasons in accounting for action. Second, psychologism is impressed by what has been called the placement problem concerning normativity.[13] Scientific naturalism holds that the entirety of reality is constituted by the world as described by the natural sciences, in particular by physics. On a common view, the list of constituents of the world as provided by science does not leave room for normative phenomena. Facts about someone's having a good reason to do something are normative phenomena in a sense

[12]The details of psychologism's hypothetical analysis of normative reasons are filled in in §4 below.

[13]Cf. De Caro and Macarthur (2010) and Darwall, Gibbard, and Railton (1992: 126).

that appears incompatible with the scientific naturalist contention that the world is devoid of genuinely normative significance.

If we accept naturalism in this sense, the question of how and where to place normative reasons leaves only two options.[14] The first is to conclude that, contrary to the common assumption, normative phenomena do not in fact exist, at least in the sense in which we suppose them to exist. However, eliminating reasons from our picture of the world is highly revisionist: it requires giving up many assumptions we make about agency, a prospect that is unattractive to most philosophers. The alternative is to attempt to give an account of the normative phenomena which shows them to be unproblematic. Psychologism takes this approach when it analyzes statements about normative reasons in terms of motivating reasons. On a hypothetical analysis of normative reasons, the only facts that have to be accepted as real are actual and hypothetical facts about psychological states. This reduces the problem of "placing" normative reasons to the less formidable one of finding a place in the natural order for psychological states.[15]

By contrast, normative realism, which does not share psychologism's naturalist conception of what there is, is happy to accept elements with irreducibly prescriptive authority in its picture of the world. As a result it does not take there to be a special problem of "placing" normative elements in our world view. Accordingly, it does not see a need to provide a reductive analysis of normative reason statements. It can understand "having a reason" as a relation between an agent and a proposed action. Furthermore it can understand this relation in a fashion similar to the way in which moral realism traditionally understands the property of goodness. It can understand having a reason as a non-natural relation.

This last difference between the two theories of reasons highlights the parallels between normative realism and the traditional metaethical doctrine of moral realism, which regards moral statements as expressions of cognitive truth-apt judgments that are true or false depending on whether the action-type in question really has the non-natural property of good-

[14]For further discussion of naturalism, see §4.

[15]Michael Smith advances a hypothetical analysis of normative reasons that fits the description in the text. He does not, however, think that such an analysis can be made "fully reductive and explicit" (Smith 1994: 161). The trouble is that a statement of the analysis — in terms of full rationality — requires using other normative concepts. Still, he is committed to making "the legitimacy of moral talk depend on squaring such talk with a broader naturalism" and he adds that "[i]f there are any moral properties, such properties must just be natural properties" (Smith 1994: 57). Given this goal, it seems fair to say that, like most other psychologistic theorists, Smith advocates a genuinely reductionist theory.

ness.[16] Thus both views imply ontological commitments that are apt to cause worries for defenders of an austere naturalism, and both face parallel objections.[17] It is important to note, however, that although the two views are related in spirit and expressive of a similar philosophical temperament, they are not at all identical. Normative realism, as defined above, is a view primarily about the nature of reasons for action. Moral realism does not entail any particular view on what it is we talk about when addressing the topic of someone's reason for doing something. It is also possible to defend normative realism while at the same time defending a non-realist metaethical theory. To give an example, Thomas Scanlon's view of reasons for action is realist in this sense, but he also proposes a contractualist analysis of what makes an action morally wrong based on the concept of the action as reasonably rejected by other rational agents.[18] Thus although reasons having to do with, say, the person's well-being are independent objective existences, he does not propose a realistic account of specifically moral reasons. Moreover the normative concepts which are the subject of moral realism are specifically moral concepts such as obligation; what is right or wrong; and our duties towards others. The many non-moral reasons, including hedonistic reasons, prudential reasons, reasons of etiquette or beauty and so on, are within the scope of normative realism but outside that of moral realism. Thus although considerations originating in metaethical discussion can elucidate questions about the viability of normative realism, we have to be careful to keep separate what is specific to the moral sort of normativity, and careful not to confuse the topic of ethics with the topic of the normativity of reasons in general.

1.2 Arguments against normative realism

As a first step towards an assessment of normative realism as a theory of reasons, consider the observation that acting on a reason requires awareness of the reason. Suppose Jim quits smoking because it is damaging his health. If we attribute his action to this reason, we must also suppose that he has the belief that smoking is an unhealthy habit. Unless an agent is

[16] By "moral realism" I mean the non-reductionist doctrine defended by non-naturalists such as Moore (1959) rather than the different class of views also sometimes called by the same name which combines a cognitivist view of the moral with a program of naturalist reduction, e.g. Boyd (1989).

[17] I will return to these difficulties in §4 below.

[18] Cf. Scanlon (1998: ch. 4–5).

aware of a consideration at least in some sense, the consideration cannot be *his reason* as an agent to do what he does.

This observation is correct, though its statement requires some care as there are apparent counter-examples. It is true that sometimes we say that somebody did φ because of p without knowing (or believing) that p is true. In this sense, a person may buy a house *because* of the bubble in the real estate market without knowing of its existence. However, in such cases we are not attributing to the person an agential reason in the relevant sense but rather a mere cause, albeit a complex and psychological and social one.

It is true that one can act in accordance with a reason without noticing the reason if, for instance, without my knowledge, the motor oil is running low and this is a reason for me to take my car to the garage. If I then do drive my car there, perhaps because I think the tire pressure needs checking, I do what I have reason to do, viz. take the car to the garage, but I am not aware of the reason. But although I act in accordance with this reason, I do not act for this reason; the oil is not my reason for doing it. Acting for a specific reason requires that I have formed an intention as a result of grasping the consideration. Intentional action involves awareness of what one is doing and knowledge of the reasons for which one is doing it.

But can we not also act for a reason, in the emphatic sense, without the relevant belief? Skepticism may derive from examples of people acting from unconscious assumptions. A gambler bets a large sum on one horse in the race. Here we may say that he makes the bet because the horse is likely to win, while at the same time, perhaps, we say that he does not quite believe that the horse will win. In fact the man may protest vehemently that he does not have this belief, and later he may explain his own action in terms of rather different motives.

However, we should not see this as an example of someone unaware of p acting because of p. On the contrary, the example shows that agents can be mistaken about their own states of mind. The observation that acting for a reason requires awareness of p does not imply that the agent needs to be conscious of the reason in a clear and reflective way, in the sense of correctly being able to answer with authority all questions about what he thinks. There is a presumption in favor of self-knowledge, but there are cases, many of which involve irrationality or self-deception, where we are less than perfectly aware of our own psychological states. Moreover, when it comes to giving reasons for past actions, we enter the somewhat murky waters of the agent's retrospective interpretation of his own ac-

tions. Just as we can be mistaken in our second-order beliefs about what we believe, there is room for mistakes concerning what actually motivates our own actions. The complex psychology involved in examples of this type does not disprove the sound principle that, whenever we have reasons and proceed to act on them, we also need some degree of awareness of those reasons.

Coming back to the main topic, criticism of normative realism may start from an application of the awareness principle. Although according to realism reasons are facts in the world, whenever an agent responds to a motivating reason by performing an action, he must also grasp the reason as a reason — that is what the principle entails. A reason-ascription "X φ'ed because *p*" can be shown to be false if it can be demonstrated that at the time X didn't believe that *p* was the case.

This observation can be turned into an argument. If it is impossible to φ because *p* without being aware that *p* is the case, should we not conclude that motivating reasons are beliefs? It is not hard to see that this is not a sound argument for psychologism. The fact that one cannot act for a reason *p* without having the belief that *p* does not entail that the reason and the belief are one and the same. Holding that I can only do φ because *p* if I believe that *p* is perfectly compatible with the claim that the belief, far from being the reason, is only a causal concomitant of the action. In order to act for a reason, it is necessary to grasp the consideration, which itself is not a psychological state. It may be said that there is a relation of counterfactual dependence between having the motivating reason *p* and being aware of *p*: if one didn't have the belief, one would not have the motivating reason either. While true, again this does not show normative realism to be false. The most it shows is that having the relevant belief is an enabling condition of acting for a reason and of having a motivating reason. According to the normative realist, stating such a condition is just to point out something without which the action could not have taken place. It achieves no more than saying that if my heart didn't work, I would not be conscious, which does not in any way show that a properly working heart is to be equated with having consciousness. Similarly, the fact that the belief that the reason obtains plays an enabling role, causal or otherwise, in the process leading to an action has no tendency to show that the reason is constituted by the belief.[19]

A second question about the viability of normative realism is raised by the way Raz qualifies his view in the passage cited above, where he writes that

[19]Dancy appeals to enabling conditions to counter the psychologistic point in Dancy (2000: 127–8).

"[b]eliefs can sometimes be reasons, although it would be wrong to regard all reasons as beliefs" (Raz 1975: 17). Although Raz clearly thinks that, by and large, motivating reasons are facts, here he seems to concede that this is not universally true. However, it is not clear what kinds of exceptions he has in mind. Two types of cases in which it would be acceptable to say that the agent's reason for acting was his belief that *p* suggest themselves. To illustrate the first kind of case, suppose that an agent believes that he is being followed by the C.I.A. If the belief persists despite the fact that there is little or no evidence for this suspicion, a friend advises him to go see a psychiatrist, take anti-anxiety medication, etc. *because of his belief*. In this scenario, the agent's reason for doing these actions would be a psychological state. In a second type of case, an agent takes himself to have a reason *p* that speaks in favor of doing φ, but his assessment of the situation is mistaken because, as it turns out, it is in fact not the case that *p*. In this sort of case, as in the first, it is common to say that the agent's reason for φ'ing was his belief that *p*, although the circumstances as well as the implications are different.

Putting to one side for the moment the second type of case, what should normative realism say concerning the obvious fact that we sometimes refer to, as we might call them, mental reasons?[20] Compare two cases. Suppose Carl notices suspicious figures in trench coats following him, finds that his living room is bugged, and so on. Given these pieces of evidence, it is reasonable for him to take appropriate measures: change taxis often to shake off his followers, check into a hotel under a false name, and so on. We would recommend him to do these things. If he does, the reason why he acts in this way, we may say, is that he believes that he is being followed by the C.I.A., though the belief seems to be a reason only in an indirect way. But now change the example and suppose that Carl is paranoid and reads too many spy novels. In this case, we know that he is in fact not being followed by the C.I.A., and we would not advise him to assume a false name or to try to cover his tracks but rather to seek the services of a psychiatrist. If he does go see a specialist, we might again say that his reason for doing so is his belief that spies are chasing him, but this time our meaning would be different: we would mean that his being in a psychological state itself was a good reason to do something.[21]

Perhaps, then, in leaving open the possibility that a reason may, in some instances, be a belief, Raz is thinking of cases like Paranoid Carl's. It may be thought, however, that admitting cases such as these, although rare, as exceptions to the rule that psychological states cannot be reasons places

[20] The second type of case, what I call an "isolated case", is the topic of the next section.

[21] Cf. Ammereller (2005), Hyman (1999).

one on a slippery slope towards psychologism. If the theory admits beliefs as reasons in some cases, one may argue that it could equally identify reasons with beliefs across the board. It may be said that if having a belief can be a reason in Carl's case, it cannot be true, as the realist believes, that we can reject on principled grounds the possibility of psychological states as reasons in normal cases.

Such an argument, however, if it were put forward, would be unconvincing. To begin with, if it generalizes from a special case to the general thesis that reasons are beliefs, it fails to take into consideration that, if Paranoid Carl has a mental reason, it is a peculiar type of reason. In order to discover his reasons, he needs to reflect on his own state of mind and on his own lack of sanity, but agents in ordinary cases do not have to engage in this type of introspection when deliberating. The belief about spies comes into focus in a way it does not in everyday actions. In normal deliberation, beliefs are transparent to the agent, so that what he reacts to are features of his environment. His beliefs are only the medium through which he grasps the facts. Paranoid Carl's case is special in that his belief becomes opaque to himself.

Moreover, in normal cases when an agent acts for a reason that p and therefore has the belief that p, he acts because he takes p to be true rather than false. It is of course true that when an agent acts on a reason, what he considers a reason may or may not be true; we are all occasionally mistaken in taking something to be a reason which in fact isn't. Nonetheless, even if an agent's action is based on a mistaken belief it matters whether or not the agent believes what is in fact the case. If the agent acts because of what he believed was a reason, the action is bound to be considered, in some sense, a failure. Had Cautious Carl known that his belief about the secret agents was wrong, he would not have changed taxis. By contrast, Paranoid Carl's action is not sensitive to the truth or falsity of his own belief. The fact that he has the belief despite half-suspecting it to be false may be all the more reason to see a psychiatrist. What matters is not, as it does in ordinary cases, the truth of the belief but the belief itself *qua* psychological state.

The fact that in the opaque case, whether or not the belief represents the world as it is does not play the role it normally does is an indication that the agent in such a scenario is alienated from his own beliefs. He does not treat them as part of his epistemic apparatus, as he normally would. Instead he sees them as something separate from his epistemic self. Generalizing from the opaque case to normal intentional actions fails because to do so is to put both types of cases in the same category. Yet it is essential to preserve a contrast between the alienated and the

non-alienated reasons.

The normative realist need have no qualms about holding Paranoid Carl's reason to be his belief that he is being followed by spies. We can after all understand this as a variant of the standard realist formula: the reason is *the fact that* he believes that *p*. Although psychological factors enter into the explanation here, they do so in the same way that any other facts, psychological or non-psychological, do.[22] Realism need not deny that psychological factors can be reasons in this way, which resembles the way other people's beliefs or desires can be a reason for doing something. On the other hand, if psychologism holds that both Cautious Carl and Paranoid Carl act because of their belief about secret agents it obscures the distinction between transparent and opaque reasons. The former in fact does not react in the same way to his belief as the latter. What holds in the peculiar case cannot be what holds in the general case.

If anything, then, the existence of opaque reasons poses a difficulty for psychologism, rather than constituting a point in favor of it. To preserve the contrast, it seems, what psychologism needs to say is that the agent with the opaque reason is really acting because he believes that he believes that he is being followed by secret agents. That Paranoid Carl's reason is really a second-order belief may seem unintuitive at first glance. Still, as has been noted, the opaque cases involve taking a reflective attitude towards one's own mental state. Consequently it is not implausible that although reasons are always a psychological affair, opaque cases are distinguished by being as it were doubly psychological. The cases therefore do not constitute genuine counterexamples for either view: for psychologism, reasons are always inside as part of the agent's psychology,

[22]Dancy (2000: 102) draws a distinction between the psychological state of the agent and the fact that the agent has the psychological state. The point is that it is not quite the same to say that the agent's reason is his belief that *p* and to say that his reason is the fact that he believes that *p*. This distinction, however, seems too subtle. There is of course a simple metaphysical distinction here. Whereas in the one case, we point to a state of the agent, in the other case we refer to a state of affairs or fact. As Dancy points out, these two are not the same thing. However, this distinction does not seem of any consequence in this context. We have no problem translating from the one form of words to the other. Dancy thinks that we need to treat the theory that takes reasons to be psychological states and the theory that takes reasons to be psychological states of affairs differently. But if we say that the agent's reason is his belief, we necessarily mean that the reason is his belief at some particular time or period of time, i.e. the agent's believing that *p* at time *t*. For something to be able to serve as a reason, it needs to have propositional structure. If I say that my reason for selling the car is my belief that it would be too expensive to repair, then surely this implies that my reason is that fact, the fact that I believe that it would be too expensive to repair. Although states of affairs and psychological states are ontologically different, in the context of assigning reasons, we always really mean psychological states of affairs, for which the other idea is just a shortcut.

though in the special cases doubly so; for realism reasons are always facts outside in the world, though in special cases about the part of the world, the agent's psychology.

1.3 Isolated cases

Normative realism emerges unscathed from the argument from opaque cases. This leaves the second interpretation of Raz's admission that beliefs can sometimes be reasons. If true beliefs aren't reasons, perhaps mistaken beliefs are. As an illustration, take Christopher Columbus who, believing that there was a short westward route to India, decided to sail West across the Atlantic. The difficulty for normative realism is summed up in the question: When Columbus set sail, what was his reason for doing so? Or, to put things another way, what seems important for normative realism is apt to be expressed using the rhetoric of *being in contact with* what is in fact true or desirable. Acting for a reason, realism insists, is reacting to an event or object in the environment, rather than to the merely psychological veil of beliefs and desire. Our actions are motivated not just by a representation of facts but by the facts themselves. But sometimes, as for Columbus, the agent's factual mistake isolates his action from the reality. Call these cases, in which the false belief seems to play an important role, isolated cases. The objection is that realism cannot give an explanation of isolated cases.

The objection is a phenomenological argument based on observations about the way we use reason-assigning locutions in everyday language. Do we refer to mental states when giving a reason? The first thing to note is that, with respect to normative reasons, we rarely say that an agent ought to do φ because he believes that p.[23] Mention of mental states is more common only when we turn to explaining actions, i.e. giving an agent's motivating reasons. These explanations have two basic forms. On the one hand, we use explanations of the form

[23]This is true, at least, when we consider only what can be called objective reasons. We sometimes say that the reason why S ought to do φ is his belief that p, but in these cases we mean that he has a subjective reason to do so. As an example, we may say of an agent that he ought to do, or intend to do, φ because of his belief that p despite the fact that it is not the case that p. Rather than giving substantive advice, the recommendation would have the force of urging the agent to comply with requirements he is under in virtue of being a rational agent. The reason would be only subjective because, in such a case from the point of view of the speaker the agent does not really have a reason to φ. Because these interesting cases will be discussed in detail in chapter 6, we can disregard them here.

(A) He φ'd because *p*.

where the verb represented by the letter "φ" is used in the past tense because we are dealing with an occurrence that has already happened.[24] This statement contains no apparent reference to any psychological states of the agent. By contrast, the second form,

(B) He φ'd because he believed that *p*.

explicitly mentions the agent's belief that *p*. Both statements have various alternatives formulation, including

(A′) His reason for φ'ing was that *p*.

and

(B′) His reason for φ'ing was his belief that *p* (was that he believed that *p*).

Different formulations within the groups — (A) and (A′) or (B) and (B′) — are essentially equivalent in use and meaning, but there are substantial differences between what we may call A-type explanations and B-type explanations. Crucially, they differ with respect to their factive implications. The verbs "know" and "remember" are factive. To say that *S* knows that *p* or that *S* remembers that *p* implies the truth of *p*. Causal explanations are also factive. The statement "(The fact) that I opened the oven door caused the soufflé to collapse" could not be true unless it was also true that I opened the oven door. It can be disputed whether the rational explanation of intentional action is a type of causal explanation. Whether or not that is true, A-type explanations share with non-rational causal explanations the property of factivity. This makes A-type explanations unsuitable for what we called "isolated cases" above. Thus we cannot say without contradiction:

Jack offered the job to Mary because she is a Harvard graduate but in fact she doesn't have a college degree at all.

Because this statement is a contradiction, where the first part of the sentence asserts something that the second part denies, we cannot use A-type explanations in this case. With B-type explanations, on the other hand, we can express the thought:

[24] We can also explain actions which are, at this instant, in the process of happening. For simplicity's sake, present-tense explanations are ignored in what follows.

> Jack offered the job to Mary because he thought she was a Harvard graduate, but in fact she doesn't have a college degree at all.

We can say this because explanations of the form "because he thought that *p*" do not imply the truth of *p*. It is of course true that they are factive in another sense: they entail the existence of a belief that *p*. However, this is a feature shared by the non-psychologized version. Keeping in mind, then, that a B-type explanation does entail the fact that the agent believes that *p*, I will say that such an explanation is not factive in the important sense.

An argument for psychologism constructed on the phenomenology of language could run as follows. We frequently use both A-type and B-type explanations. Isolated cases, a subset of the set of intentional actions, are such that their account can be given only by using B-type statements. As far as ordinary cases are concerned, although we can use A-type explanations, we are also free to use B-type. Thus only B-type explanations cover both isolated and regular cases. Because it is more generally applicable, the B-type explanation represents the more fundamental form of explanation. The argument concludes that the more fundamental explanation of all cases, isolated or non-isolated, involves reference to an agent's psychological state. Therefore, whether or not the case is isolated, the reason with which an intentional action is done must be a psychological state.

According to this line of reasoning, B-type explanations are primary compared to A-type explanations because they are more universal; A-type explanations should be thought of as derivative. "X φ'd because *p*" is understood as elliptical, a mere shorthand for the more complete "X φ'd because he thought that *p*". Now it is certainly true that B-type explanations are more universal than their counterparts. But a defender of normative realism might point out that B-type explanations sometimes feel forced or unnatural. Thus, asserting that

> Don put on rubber boots because he thought that it was raining outside.

while not logically implying that it isn't in fact raining, seems to carry a strong suggestion that Don was mistaken in his belief. In any event, picking this form of explanation would be an odd choice if Don's case was not a known or suspected isolated case. Type B is used mostly in isolated cases, when the agent acted on a belief that we, as the attributer, know was mistaken. We also use it when we are skeptical about the

agent's motivation, when we suspect but do not know for a fact that the belief is not true. In general, we use B-type explanations only if we think that something is amiss with the action in question. By psychologizing the account, we signal reservations about the agent's action. Use of this style of action-explanation is unlikely to be left standing on its own since it invites further questions along the lines of "In which way do you think was the agent wrong to believe what he did? In which sense was the agent unjustified, acting as he did?"

Absent special circumstances, it is more natural to choose A-type explanations. Although we are sometimes forced to retreat to the more restricted B-type, the full A-type explanation always remains the default. If an answer to the question of what form of explanation is primary is a clue to the nature of reasons, this idea yields two opposed arguments. On the one hand, normative realism may appeal to the observation that A-type explanations are more natural to argue that they constitute the primary form of action-explanation, as well as the one that reveals the true structure of reasons.[25] The normative realist objects to the psychologistic proposal that it is bad policy to model one's theory of reasons on the form of explanation natural or required only in a special subset of intentional actions. Instead we ought to take the more natural explanation as the model case. But as the A-type mentions no psychology but instead includes a bare dependent clause, often prefixed with the expression "the fact that…", reasons for action are to be equated with facts or states of affairs. In this spirit Dancy writes that "the most revealing form, perhaps I should say, the form least likely to mislead philosophers, is the simple form which contains no visible reference to belief at all" (Dancy 2000: 135).

On the other hand, the defender of psychologism can appeal to the universal applicability of B-type explanations to support his claim that they are primary and conclude that, because this form explicitly mentions beliefs and other mental states, reasons for action must in general be mental states. Given that both sides claim primacy, how can we break this stalemate? Neither argument is strong enough to defeat its target as it stands, so if we are to pick between the two theories of reasons on offer on phenomenological grounds, it is necessary to strengthen one of the arguments substantially. The normative realist's argument relies on comparative naturalness of the two explanation types. The mere fact that

[25] See e.g. Jennifer Hornsby's claim that, while both what I have called A-type and B-type explanations generate reasons, the normativity of the reasons behind B-type explanations is inherited from the normativity of the reasons behind A-type explanations (Hornsby 2008: 258).

using one form is less natural than the other is not sufficient for rejecting the opposing theory outright. Although the argument is suggestive, it does not hold the prospect of a knock-down argument against psychologism. This leaves the psychologistic line of thought which, if sound, has as its conclusion that normative realism is not just unnatural but untenable. The next step, then, is to spell out the argument in greater detail. How do we explain Columbus's setting sail in 1492? According to the objection, the A-type explanation

> (1) Columbus sailed West because there is (or was) a short western route from Europe to India.

cannot be used because A-type explanations are necessarily factive and there is no non-psychological fact in the vicinity that we could appeal to. Instead, we have to resort to a psychologized form of explanation:

> (2) Columbus sailed West because he thought there was a short western route from Europe to India.

This example shows, according to the objection, that not all intentional actions can be explained using A-type explanations. But if this is true, realist explanation cannot reflect the true nature of the action even in normal, non-isolated cases. If a theory of reasons is right to insist on non-psychologized forms of explanation, these explanations should be able to deal with all cases, including the isolated examples. But this is shown false. According to the argument, if A-type explanations fail in some cases, they aren't fit to reveal the true nature of reasons. Therefore, the argument concludes, reasons cannot be facts. This argument relies on three important assumptions:

(i) Any adequate theory of reasons must be completely general.
(ii) A-type explanations are necessarily factive.
(iii) The structure of reason-explanations reflects the nature of reasons.

Each of the assumptions can be questioned individually. Let us examine each in order. The *first* proposition, which is clearly a principle in the psychologistic argument, is the assumption that any explanation that reflects the nature of reasons must be capable of being applied to all cases of intentional actions. Columbus's reason for going to sea cannot be the fact that there is a short westward route to India because there is no such

fact. In this case type A cannot be used because it implies a false proposition. The argument concludes that across the board type A is not the true form of explanation and that facts are never reasons.

One way to defend realism is to block this final step. The defense concedes that A-type explanations cannot be used in isolated cases but denies that the existence of isolated cases shows that realism is false in general. The objection shows that realism is false *for* isolated cases but it leaves untouched the view that in cases where all goes well the agent is still in contact with reality. In those cases, the reason is the fact in the world that *p*. When an agent is mistaken in his conception of the situation, immediate contact with the facts is lost. Isolated cases are precisely those in which the agent does not appropriately react to his environment. According to the defense, in such a case although the objector is right to say that the reason for action is only a belief, this has no tendency to show that reasons can't be facts in the vast majority of cases.

If viable, this proposal would permit realism to take isolated cases in its stride. Furthermore, it would do so without the need to question the factivity of reason explanations. The realist attributes to the objector a line of thought according to which a uniform type of explanation is applicable to all cases. In fact, no less a defender of psychologism than Bernard Williams spells out just this thought in his highly influential paper "Internal and External Reasons" where he states, as a general principle, that "[t]he difference between false and true beliefs on the agent's part cannot alter the *form* of the explanation which will be appropriate to his action" (Williams 1981a: 102).[26] According to the principle, for the question whether type A or type B is appropriate to a given case, it cannot make a difference whether or not the agent conceives his situation correctly. But a realist may insist that this is not a valid principle. He can protest that a comparatively small number of exceptional cases cannot undercut the realist theory as a whole. Rather than assuming that a single theory of reasons can cover all cases, we should accept that different explanations are appropriate for different cases.

According to this proposal, both psychologized and non-psychologized explanations have their place and neither is reducible or inferior to the other. In good cases, the reason-explanation shows the agent's connection with the facts. This proposal can take inspiration from disjunctivism, a conception in perception theory. Philosophers who defend disjunctivism are drawn to direct realism, the view that perception allows us to enter into direct contact with our environment. On this view, the content

[26]Emphasis in the original.

of our sensory perception is not, as has often been held, a state representing reality but an actual fact in the world. While this seems plausible in "good" cases where a subject, for instance, actually sees a red object in front of him, it has the problem of being unable to account for "bad" cases where, although the agent may think he sees the object, it only appears to him as if he saw the object. The existence of bad cases like illusions here leads many to reject direct realism. To disjunctivists, however, this assumes falsely that good cases should be assimilated to bad cases or that our job as theorists is to identify an element — the immediate object of perception — which is common to both illusion and veridical perception.

Jennifer Hornsby has proposed to adapt a view of this sort to action theory. According to her, perceptual disjunctivism allows us to understand statements about perceptual appearing disjunctively:

> It appears to X as if it was true that p if either X sees that p so that X is well-placed to know how things objectively are or it merely appears to X as if it was true that p so that X is not in such a position.[27]

If Hornsby is right, we can transfer this idea to action theory. She emphasizes the importance of knowledge for acting for a reason.[28] If X acts because p, this is adequately captured in the idea that X acts because he knows that p, but not by the idea that X acts because he believes that p. In Hornsby's view, the former, but not the latter, is something that can actually provide the agent with a reason. She introduces the disjunctive element by giving an account or analysis of acting for a reason in which "acting from knowledge belongs under the head of acting from belief":

> If X φ-d because X believed that p, then either X φ-d because X knew that p, so that X φ-d because p or X φ-d because X merely believed that p.[29]

Here "X merely believes that p" is intended to imply that X did not know that p. To Hornsby, this account explains the neutral type of explanation

[27] Cf. Hornsby (2008: 253). I have simplified Hornsby's formulation somewhat.

[28] Hornsby (2008: 251) points out that it is possible that (i) an agent X performs an action φ thinking that p and (ii) it is true that p yet (iii) it is still wrong to say that X φ'ed because p. This is true in circumstances in which the agent does not know that p even though he believes it and it is true. This can happen when the agent's belief is true although it is derived from an unreliable epistemic source such as the testimony of a speaker who cannot be trusted. This is an interesting point worth exploring which, as Hornsby points out, suggests that there is a connection between acting for the reason that p and knowing that p.

[29] Cf. Hornsby (2008: 252).

— B-type explanations are applicable whether or not the belief is true — in terms of more committive formulations that imply that the agent was right or wrong, respectively. Furthermore Hornsby argues that A-type explanations derive their explanatory power from B-type explanations so that we would not be able to understand the former without a prior understanding of the latter. If the neutral explanation is itself an amalgamation of good and bad cases, we can assume that the more substantive disjuncts reflect the nature of reasons better. In particular, acting because *p*, or acting because of the knowledge that *p*, is a way of being in direct contact with the facts, in a way that is similar to how according to disjunctivism we should say that seeing that *p* is having the fact that *p* manifest to one without worrying that a more universal neutral description of the perceptual experience is available.

Does a disjunctivist approach help normative realism find a place for isolated cases? The disjunctive account of the neutral explanations holds that there are two different accounts which are applicable depending on the nature of the case. The proposal concedes that Columbus's reason for going to sea is his mere belief that there is a short route to India while insisting that in non-isolated cases reasons are facts in the world. However, to grant that, in Raz's phrase, reasons are sometimes beliefs, is a major concession. Normative realism holds that it matters whether we say, in any given case, that the reason is a fact or a belief: the latter conveys a misleading picture of agency. To act intentionally is not to turn inward but to regard some external consideration as a reason. For this it does not seem to make a difference whether the agent is mistaken in his conception of the situation. The realist point, if it is valid, must be the principled one that whenever we act for a reason, our reason cannot be a psychological state. If reasons, except in opaque cases, cannot be beliefs, it seems impossible to make room for exceptions in isolated cases.

Furthermore, the normative realist needs to take into account the possibility of opaque cases. Paranoid Carl's reason for going to see a psychiatrist is his belief that international spies are out to get him. As we have seen, this is not an isolated case because Carl's reason, unlike Columbus's, is genuine. But if realism accepts the disjunctive proposal, it seems unable to make the distinction between opaque cases and isolated cases: in either case, the reason is constituted by a psychological state. Not to be able to make this crucial distinction is a major deficiency.

This suggests that Williams's principle that the truth or falsehood of the agent's beliefs ought not to be allowed to affect the form of the explanation applicable is in fact sound. We should be careful not to assume in general that good and bad cases necessarily share a common form, but the

argument for a disjunctive conception in the theory of action is weak. The proposal seems committed to the claim that there are considerable metaphysical differences between isolated cases and regular cases. A regular case of acting for a reason involves the agent's standing in a relation to a fact in the world, whereas in an isolated case, according to the proposal, the agent stands in a relation to his own belief or desire. What is more, the relation in the good case seems to be that the fact makes the actions in some way reasonable. The disjunctive conception appears to say that it is equally possible for a psychological state to stand in this relation. But it is incredible that, depending on the truth of the belief, one of the relata is either the belief that *p* or the fact that *p*. The connection between reason and action is a rational one, and it seems that good and bad cases have at least enough in common to support the notion that recognizably the same rational sort of connection holds in both types of cases.

It seems, then, that normative realism cannot escape the argument we are considering by attacking the assumption that the form of explanation must be general. To move on to the *second* assumption, is it true, as the psychologistic argument has it, that reason-explanations are necessarily factive? According to Dancy this is a mistake: although some A-type explanations are factive, we have no reason to think that all non-psychologized explanations uniformly imply the truth of what is taken to be a reason. Dancy concedes that some explanations of this type are incompatible with the agent's conception of the situation being mistaken.[30] However, Dancy holds, there is choice between a number of locutions, some of which do not have this property. He mentions two examples:

> (3) His reason for doing it was that it would increase his pension, but in fact he was quite wrong about that.
>
> (4) The ground on which he acted was that she had lied to him, though actually she had done nothing of the sort.[31]

Because the statements do not mention any of the agent's psychological states, they seem to belong to type A. According to Dancy these locutions differ from regular A-type explanations in that explaining the action does not commit one to the truth of the reasons. This is purportedly shown by the fact that, as in (3) and (4), a conjunction consisting of the reason statement and the negation of the proposition in question does not constitute a contradiction. Dancy takes the existence of these locutions as evidence

[30] Cf. Dancy (2000: 133).

[31] Dancy (2000: 132).

for his claim that "the purposes of explanation of action in terms of the agent's reasons do not require such explanations to be factive" (Dancy 2000: 133). In particular, he invites us not to feel forced by the alleged factivity of intentional explanation to abandon A-type explanations in favor of B-type explanations. Some of the locutions we use are factive, but others aren't. We ought not to think that action explanation is by its very nature factive.

What are we to make of Dancy's attempt to reinforce the realist position? The present dialectical situation is that there are isolated cases which cannot be explained using a factive form of explanation. Therefore to help normative realism escape the argument from isolated cases, it is necessary to show that a non-factive non-psychologized explanation is available in every possible case. The examples adduced by Dancy are intended to show that such an explanation is sometimes available, but they do not show that there is a systematic way of translating A-type explanations into non-factive B-type explanations. In fact we have reason to doubt that such a translation is generally available. Thus as Dancy's examples are in the past tense, it seems fair to ask for a non-psychologized A-type explanation of an action happening at the time of speech. Transposed to the present tense, the sentences become:

> (3′) His reason for doing it is that it will increase his pension, but in fact he is quite wrong about that.
>
> (4′) The ground on which he is acting is that she has lied to him, though actually she has done nothing of the sort.

While (3) and (4) are acceptable, (3′) and (4′) have an air of logical contradiction about them. The first half seems to imply the negation of the second. The difference that explains why (3) and (4) have no factive implications is that because the main verb of the sentence is in the past tense the dependent clause uses the verb forms "would increase" and "had lied", following the grammatical rules of the sequence of tenses, whereas (3′) and (4′) contain a main verb in the present tense and a subordinate clause verb in the future tense, "will increase", or in the present perfect, "has lied". If this is true, the fact that some explanation forms lack factive implications depends on the particular grammatical construction of dependent clauses used in these sentences, a construction that is only available when describing actions that have occurred in the past. For actions that are in the process of being performed, no such grammatical device is available: "He does it because it would increase his pension" is ungrammatical. At least with present-tense explanations, the only form

of explanation we are left with is factive. But then we can repeat the psychologistic challenge to ask the realist how to treat action-explanations like (3′) and (4′).

Noting the particular grammatical forms used in Dancy's examples leads to a second point. Given that (3) and (4) show the grammatical phenomenon of backshifting ("will increase" becomes "would increase", "has lied" becomes "had lied"), it is likely that, in the English grammar at least, reason-giving explanation take the grammatical form of indirect speech: "his reason was that …" has the same underlying form as "he said that …". Though it is clearly not correct that when we say that someone acted for a reason *p*, we claim that the agent *said* that *p*, it is more plausible that assigning a motivating reason asserts that the agent *believed* that *p*. Indeed the sentences (3) to (4) can be read as shortcuts for:

> (3″) His reason for doing it was *his belief* that it would increase his pension.
>
> (4″) The ground on which he acted was *his belief* that she had lied to him.

But if the grammatical form used in Dancy's formulations are indicative of the ascription of indirect speech, it does not seem that Dancy can appeal to these forms of words to support his view that the reference to psychological elements in non-factive explanations can be eliminated. Instead (3) and (4) can be seen as incorporating a covert reference to the agent's psychological states.

In a second attempt at demonstrating that non-factive and non-psychologized explanations are not mutually exclusive, Dancy appeals to the possibility of canceling the factive implications in A-type explanations using an apposition.[32] On this proposal, the explanation

> (5) He drinks the contents of the glass because, as he believes, it contains gin, although in fact in contains petrol

is both non-psychologized and non-factive. Here Dancy is right to claim that the explanation succeeds without self-contradiction; the use of "as he believes" and "as he thinks" is particularly common and useful when reporting in academic prose the reasoning of an author without committing oneself to its truth, often because one disagrees with him. But again the connection to indirect speech is strong: the locution clearly reports what

[32]Dancy (2000: 128–31).

is going on in the agent's mind. There is a reference to a psychological state in the dependent clause, although it does not seem to be straightforwardly part of the propositional content. Dancy's appositional account gives us little reason not to think of (5) as a mere notational variant of "… because of his belief that it contains gin", with its explicit mention of the psychological state.

We can conclude that appealing to Dancy's special examples and to his appositional account does not succeed at dislodging the second assumption we have been considering. As a matter of fact, the English language does not seem to contain, for any given set of circumstances, a means of explaining an action which both is non-factive and belongs to type A: in some cases of non-factive explanations, we have to resort to implicitly or explicitly psychologized explanations. If we take the reasons-statements at face value, this conclusion implies that reasons really are sometimes psychological states. Nonetheless, despite the linguistic appearances, Dancy is right to insist that reason-explanations are not essentially factive. In other words, we should reject the *third* assumption in the list above, according to which the linguistic structure of reason-explanations automatically reveals the status of reasons. Although we must of course partly take our cue from the linguistic structure of our explanations, we are well advised to keep a certain philosophical distance.

One reason to be careful is that our language, the product of an organic development spanning millennia and responding to various pressures, is not perspicuous. Some features of the language are contingent on extraneous factors, such as pragmatic considerations. Thus the fact that our action-explanations are either non-psychologized or non-factive but never both may be a historical consequence of conditions relating to usefulness or simplicity. Such speculation aside, it does not seem impossible to introduce, by *fiat*, a new word "because$\oplus$" into our language where "x φ'ed because$\oplus$ p" has the same use as the English word "because" except that it does not imply that p is the case. If this possibility is genuine, it suggests that the fact that the word "because" and its cognates have the factive implications they do is at least to some extent an artificial or accidental feature of our language.[33]

Although it is true that non-factive explanations contain a reference to

[33]It is another matter whether there could be a language that contained only "because$\oplus$" and not "because". Perhaps this is impossible because of the way in which talk about reasons, which is essentially a social practice, is learned. However, whether or not "because$\oplus$" as a connective could stand on its own, its possibility even alongside "because" shows that it is not an indispensable feature of reason-attribution that endorsement of the reason is shared with the agent.

the agent's psychological states, we should not rashly conclude on mere linguistic grounds that reasons must be psychological states in general or even in isolated cases. To draw this conclusion would be to take the form of words we use in the explanation of intentional action too literally.[34] Instead we should change tack and focus on what we do when we explain an action by reference to a reason. As has emerged from what has been said, the primary function of action explanation is not, as in the explanation of natural events in terms of their causes, to show a causal link between two events — though it may also do that — but to display a rational relation between a consideration, on the one hand, and the action, on the other. This rational relation does not have the intelligibility characteristic of explanations in the sciences, paradigmatically in physics, but the sort of intelligibility characteristic that is at home in the logical space of reasons, paradigmatically in the explanation of speech and action in terms of propositional attitudes.[35] Suppose that when asked "Why did John eat the tofu?" we answer, "Because it contains protein." What we do is to exhibit a sample piece of practical reasoning. Thus an explanation of John's action points toward a practical argument concluding in the intention to perform the action in question, along the following lines:

> This cube of tofu contains proteins. (p)
> Proteins are good for me.
> Thus, I shall eat this cube of tofu. (q)

In giving the reason we point out that, along with ancillary hypotheses, the consideration that is advanced as the reason, *p*, implies the conclusion, *q*. Thus given that proteins are good for John, *p* implies *q*, i.e. that this cube of tofu contains protein implies (for John) "I shall eat this cube of tofu". What the action-explanation does is to show how the move from *p* to *q* is, or could be seen as, a reasonable inference and thus to make sense of the action from the perspective of the agent.[36]

Over and above this first job, our most common form of explanation, type A, also performs a secondary function, which is to endorse the truth of the proposition picked out by the explanation as a reason. This has a parallel

[34] See Dancy (2000: 133ff) for remarks that go into a similar direction.

[35] See McDowell's remarks on the "understanding we can achieve by employing the conceptual apparatus that is governed by the constitutive force of rationality" (McDowell 1998b: 328). He also asks us to picture "a particular instantiation of deductive rationality as a more or less approximate grasp of a normative structure, determining what follows from what and thus what ought to be believed, given other beliefs, for deductively connected reasons" (McDowell 1998b: 327). The kind of intelligibility proper to phenomena in the logical space of reasons will be the topic of chapter 3.

[36] The topic of practical reasoning will be explored in greater detail in chapter 4.

in talk about implications in general. It is natural to read "p implies q" as committing the speaker, not just to the implication, but also to the truth of p. Similarly, when we use type A explanations, we not only point out the existence of an inference from p to q but also assert p ourselves. In choosing this type of explanation we choose to endorse the existence of the reason itself as a fact. But this further function is hardly essential and perfectly separable from the first. This can be seen from the fact that, as we have noted, there are a variety of means of canceling this secondary function. We can, with Dancy's examples, use grammatical constructions similar to indirect speech; we can use a parenthetical "as he believed"; and, most importantly, we can use B-type explanations. If B-type explanations lack the secondary function, they still have the same primary or essential task of pointing out a rational relation. We ought to think of them as a species of A-type explanation where the functionally inessential endorsement of the proposition p is suppressed.[37]

To conclude, we should not interpret the use of B-type explanations in Columbus's case as a sign that in this instance his reason must be a mental state. Even if cases of this sort require grammatically psychologized locutions, what is asserted is the same as in A-type explanations. If this is true, the phenomenological observations adduced by psychologism rely on the unwarranted assumption that the mention of psychological states in B-type explanations licenses immediate conclusions about the status of reasons and thus fail to show realism to be false.

1.4 Normative realism, pros and cons

We have seen that the psychologistic argument fails to establish that reasons must be psychological states. The next step is to consider an argument advanced by Dancy which purports to show that reasons cannot be psychological states. Although in my view this argument also

[37] As mentioned above, Hornsby (2008) suggests that there is a connection between acting for the reason that p and knowing that p. In this context, it should be noted that "X φ'ed because p" is similar to "X knows that p". In ascribing knowledge to an agent, we not only say that the agent believes that p but also commit ourselves to the truth of the proposition that p. In Brandom's terminology, ascribing knowledge is not just attributing a commitment that p but also undertaking a commitment that p (see Brandom 2000: ch. 3). A-type explanations have a similar social aspect involving two scorekeepers, the property of simultaneously attributing commitments to another and to oneself. Knowledge has an important role in the theoretical sphere which goes beyond that of mere belief (and even justified true belief). Similarly, it may be suspected that in the practical sphere A-type explanations have an essential role that goes beyond that of non-committive B-type explanations. The connection between reasons and knowledge is also stressed by Hyman (1999).

fails to achieve its purpose, seeing what is correct and incorrect about it brings into relief an important point that has been present in the preceding discussion. Dancy's argument begins by noting a contrast in how the two theories of reasons account for motivation. Normative realism straightforwardly holds that motivating reasons, like normative reasons, are facts. If S acts because of p, the fact that p can be invoked directly as S's motivating reason. Psychologism, on the other hand, holds that, although normative reasons do not necessarily correspond to the agent's beliefs and desires, motivating reasons are psychological states of the agent. This means that it has to give a slightly more complicated explanation of what happens in motivation. According to Dancy, it has to tell what he calls a three-part story. This story can be expressed symbolically as "normative reasons $\longrightarrow$ motivating reason $\longrightarrow$ action", where each arrow represents a relation of explanation.[38] Thus when an agent reacts to a fact in the world, the defender of psychologism needs to explain the existence of the motivating reason, a psychological state, by reference to the normative reason, which even according to psychologism may be a fact in the world; and the motivating reason for its part explains that the action is performed. For instance, if Joan climbs up the tree because of a nearby grizzly, then her motivating reason is the belief that there is a grizzly in front of her. Joan's belief is explained by the fact that there is in fact a grizzly in front of her; this fact both makes the belief reasonable and explains it. The motivating reason, in turn, explains Joan's actual climbing up the tree. The normative reason is involved in the explanation of the action only indirectly, with the motivating reason acting as a proxy.

Now for Dancy, the trouble is that a theory of reasons is subject to what he calls the "normativity constraint", viz. the requirement that

> a motivating reason, that in the light of which one acts, must be the sort of thing that is capable of being among the reasons in favour of so acting; it must, in this sense, be possible to act for a good reason. (Dancy 2000: 103)

In other words, Dancy's constraint requires that motivating reasons be the right sort of things to be normative reasons. The three-part story holds that beliefs are motivating reasons, but except in opaque cases, beliefs themselves cannot speak in favor of performing a certain action: they cannot be normative in the sense required. If this is so, then psychologism's story clearly fails the normative constraint.

[38] Dancy (2000: 101).

What do we make of the objection that psychologism fails the normative constraint? Intuitively there is some substance to the normative requirement: even if normative reasons on occasion diverge from motivating reasons, plausibly motivating reasons and normative reasons should often come together. But Dancy is wrong to set up a requirement for normative reasons and motivating reasons to be, at least potentially, identical. The trouble with this objection to psychologism is that the normative constraint is not an independent requirement: it is tailor-made to rule out psychologism. It begs the question against psychologism as we have defined it by assuming from the outset what psychologism denies: the potential identity of normative and motivating reasons. But it is precisely the relationship between the two sorts of reasons that is at issue between the two theories of reasons. It is hardly clear that Dancy is right to impose on theories of reasons the requirement that they must assume, as he does, a possible identity between the two sorts of reasons.

It does, however, seem a fair request to ask such a theory to provide an account of the relationship between the two kinds of reasons that does not pry them apart too far. If a theory drove a wedge between normative and motivating reasons which makes it hard to see how the two are related, this would be a ground for rejecting it. The requirement could be explained in this way. The two sorts of reasons have a different focus. Motivating reasons are primary retrospective. When we ascribe a motivating reason to an agent, we talk about an action that has already taken place (or perhaps is taking place right now), and our goal is to render the action intelligible. We take the action as a given and seek to understand it. Normative reasons, on the other hand, have a forward-looking goal. In ascribing a normative reason, our goal is to justify an action that has not yet taken place, and we can talk about normative reasons for actions that may never happen. Our goal is, not to give an account of something given, but rather to show what it is rational to do.

As a consequence, to understand reasons we have to understand them in relation to two projects or perspectives, a problem clearly seen by Bernard Williams, a defender of psychologism.[39] On the one hand, reasons figure in first-person deliberation about what to do: they are what makes doing one thing rather than another reasonable. We may call this aspect, in which we consider questions of justification, the rational dimension of reasons. For Williams, the conception of reasons for action he considers correct "is concerned with the agent's rationality. What we can correctly ascribe to him in a third-personal internal reason statement is also what he can ascribe to himself as a result of deliberation" (Williams

[39]Williams (1981a).

1981a: 103). The deliberative perspective is important but not exclusive. As Williams also rightly points out, "if there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action" (Williams 1981a: 102). Williams calls this aspect the explanatory dimension of reasons. From this intentional perspective, reasons are what we cite in the explanation of action.[40] Reasons do not belong to either of the perspectives exclusively but have their place at their intersection. If, as we can assume, the duality of two perspectives corresponds to the duality of normative and motivating reasons, we can conclude that it is a requirement of any adequate theory of reasons to show that normative and motivating reasons are two sides of the same coin, rather than disjointed elements.

We can, then, interpret Dancy's objection as raising the valid question whether psychologism can bring together the intentional and deliberative perspective. To answer this question, we need to turn to psychologism's three-part story. As we have pointed out, a theory of reason must, in addition to explaining the ontological status of motivating reasons, provide an account of statements attributing normative reasons. Psychologism is committed to explaining normative reasons in terms of motivation. Being motivated, however, is neither a necessary nor a sufficient condition of having a reason: on all accounts we are often motivated to do something we have little reason to do, as we also often fail to be motivated by a normative reason. Psychologism needs to acknowledge a certain distance between having a reason and being motivated. As has been mentioned, this means identifying normative reasons, not with the agent's actual motivation, but with his hypothetical motivation. Such an analysis has the following general form:

> *X* has a normative reason to do φ iff he would be motivated to do φ in conditions *C*.

Of course, everything here depends on how the conditions *C* are filled in. Williams gives a particular clear exposition of the idea. According to Williams, the gist of a normative reason statement is that "if the agent rationally deliberated, then, whatever motivations he originally had, he would come to be motivated to φ" (Williams 1981a: 109). More specifically, an agent has a normative reason to φ if he has some desire *D* and there is a sound deliberative route from *D* to doing φ. For Williams, three conditions have to be met: the practical reasoning leading to the decision

[40]For the distinction between the intentional and deliberative perspectives and its relation to normative and motivating reasons, see Smith (1994: 131f).

to φ must not be based on a false belief; the agent must have access to, and make use of, all relevant factual information; and he must not exhibit any faulty reasoning. For Williams, the condition of correct deliberation is to be construed narrowly: it consists chiefly of correct means-end deliberation and using one's imagination to discover ways in which a given goal may be satisfied — a style of reasoning that has been called "constitutive reasoning".

I will return to the details of Williams's account later.[41] The question now is not whether an account of normative reasons like Williams's is true but rather whether it would, if true, divorce normative reasons from motivating reasons in a damaging way. Williams's psychologistic view is that what we posit in explaining action from the intentional perspective is recognizably related to what we attribute to an agent when explaining his action. He conceives of having a normative reason to φ as being such that if one followed certain rational procedures, one would arrive at the decision to do φ. In explaining action, we base our account partly on the assumption that the agent is deliberating from motivation he already has. This does not mean that motivating reasons and normative reasons can be identical. But if having a normative reason is related, by way of rational deliberation, to being motivated, psychologism is able to explain how a motivating reason is the consequence of a normative reason. Far from divorcing the two explanatory and rational dimensions of action, Williams's account allows us to see how the intentional and deliberative perspectives are related.

Dancy's claim that normative reasons and motivating reasons must be, at least potentially, identical is too strong. Perhaps in recognition of this problem, he glosses his own objection in a different way. He says that motivating reasons must be "of the right sort to be good reasons" and that this is not true on the psychologistic identification of motivating reasons with beliefs as the latter are "the wrong sorts of thing to be good reasons" (Dancy 2000: 106–7). Here Dancy is making a metaphysical distinction. Facts in the world belong to one ontological category, psychological states to another; and he emphatically asserts that only the former category, and not the latter, contains things that can speak in favor of doing something.[42] Now although it is true that these are two separate ontological categories, it is hard to see how this distinction matters with respect to the question at hand. Dancy seems to be basing his claim on his observation of how we talk about these two kinds of things. However, as we have pointed out, something, *p*, is a good reason for φ'ing if there is a

[41] See chapter 5.

[42] Dancy (2000: 107–8).

good practical argument from p to the decision to φ. So in order to be a candidate for a reason, p must be something which can be said to imply φ'ing, assuming certain ancillary hypothesis. But if it is correct to say that φ'ing follows from the fact that such and such is the case, it seems equally correct to say that φ'ing is implied by the agent's belief that such and such is the case. If so, Dancy's point that motivating reasons must be among the things that can speak in favor of φ'ing is compatible with psychologism.

Rather than engage in ontological subtleties, we should consider a distinction in the vicinity that matters in the context of this argument. In the philosophy of mind, it is crucial to distinguish between the act that is a mental episode and its content or intentional object. This is the difference between the taking of an attitude towards an object and the object towards which one takes to attitude. Although the distinction is basic, it has proved surprisingly easy to commit the fallacy of confusing the one with the other. This possibility of confusion is related to what Sellars calls the "notorious -ing/-end ambiguity".[43] The word "representation" is ambiguous as it can either mean the act of represent*ing* a state of affairs or the content represent*ed* in the act. Accordingly, it is easy to be confused as to what one means when talking about representations.

This difficulty is not restricted to the word "representation" but is shared by much mental vocabulary. When we talk about a person's intentions, we may refer to what he intends to do, or we may refer to his psychological attitudes of intending. Similarly, "belief" may refer to what is believed but also to the mental act of believing. Concepts of contentful states seem to be systematically ambiguous. We leave it open to interpret the word in the sense supplied by the content. However, automatic disambiguation sometimes fails.

The act-content distinction can be seen as the core of Dancy's positive point. He says that, metaphysically speaking, reasons are states of affairs rather than states of mind because only the former can be good reason. But it is more accurate to say that reasons are the contents of beliefs rather than the believings.[44] The problem is not with the metaphysical features of states of affairs but rather with two aspects of mental states. Thus if we, as we should, understand reasons in terms of their roles in practical arguments, it is clear that what we reason from is not a belief in

[43]Sellars (1963a: §24, 154).

[44]Dancy himself distinguishes between the believing and the things believed when he says that "[t]he crucial point here is that believing that p is never (or hardly ever) a good reason for φ-ing. It is what is believed, that p, that is the good reason for φ-in, if there is one" (Dancy 2000: 107).

the sense of an act but in the sense of a content believed.[45] Reasons of any kind clearly fall on the content side of the act-content divide.[46] From the true observation that motivating reasons must be able to count in favor of doing something it does not follow that motivating reasons must be facts but only that they must be believeds rather than believings.

The amended version of Dancy's objection fails to rule out psychologism. Of course, some philosophers with psychologistic views may take reasons to be acts of believing. After all, it is natural to understand "His reason was his belief that *p*" to refer to the believ*ing* rather than to its intentional content. However, it would be rash to dismiss psychologism outright. After all, if psychologism holds that reasons are beliefs, this may also be construed as meaning that reasons are things believed. The lesson to draw from Dancy's normative constraint, then, is that, although act-psychologism is ruled out, content-psychologism is still in the race.

After noting the negative payoff of normative realism — its insistence on seeing reasons on the content side of the act-content divide — we can turn to what it offers on the positive side. Normative realism identifies the things *believed*, which constitute reasons, with facts or states of affairs. This raises two difficulties. *First*, we have seen that in isolated cases, there simply is no fact that could be cited as constituting the agent's motivating reason. Given that it is not the case that there is a short westward route to India, what could explain the fact that Columbus sails west?

Second, according to normative realism's straightforward interpretation of statements about normative reasons, they are true in virtue of corresponding to facts about reasons in the world. To know that there is a reason to φ is to know a fact, so normative realism must posit normative facts. An agent who recognizes that he ought to do φ first becomes aware of the fact that he as a reason to do φ and then if all goes well this fact becomes his motivating reason. According to the realist, these facts not only make it reasonable that something is worth doing, they also serve to explain the action when it has occurred. By having explanatory significance, they play a rather important role in our picture of the world. As the normative realist might put it, in action, we react to the reasons we discover in the world as independent of us. This gives rise to worries about the kinds of facts we should posit. What makes it sensible to assume that there are non-descriptive facts? If reasons can be evaluative and reasons are facts, this theorist incurs the ontological commitment

[45] See also chapter 6.

[46] See Alvarez (2008) for a similar conclusion.

that there are evaluative facts. It can be objected that it is metaphysically dubious to assume the existence of evaluative facts.

In order to accommodate the first point, the normative realist needs to find an ontological alternative to the fact that there is a westward passage, something that can play the role of a reason. We can rule out facts about the agent's psychological state because, as we have seen, this choice would not allow the realist to differentiate between opaque and isolated cases. The realist's best option therefore seems to retreat to saying that Columbus's reason for setting sail was, not a fact, but the non-actual state of affairs that a westward route exists. This commits the theorist to the existence of non-obtaining states of affairs, an assumption that is controversial. Defending this assumption is not easy because realists regard intentional actions as *reactions* to the facts in the world that constitute reasons, so that Columbus's going to sea would have to be thought of as a reaction to a non-obtaining state of affairs. Like normative facts, non-obtaining states of affairs must be appealed to in explanations of purposeful behavior. The realist must claim that states of affairs can account for an action even if they are not actual.

Nonetheless, in order not to get caught up in questions relating to the precise nature of states of affairs, let us waive this ontological difficulty to focus on the second issue.[47] Even if normative realism can provide a satisfactory solution to the problem of isolated cases, it still faces a challenge regarding specifically normative facts. We have already noted that placing normative elements in a picture of the world largely determined by scientific naturalism is often found to be problematic. Perhaps the most prominent attack against the idea that we can reconcile a naturalistic world-view with the existence of normative facts is found in J.L. Mackie's *Ethics*. In this book Mackie criticizes what he calls "objectivisim", the view that there are objective values in the world independent of our subjective inclinations and goals.[48] In the book, he puts forward what he calls the argument from "queerness".[49] Mackie's central claim is that values are not part of the furniture of the world. The argument has an epistemological and a metaphysical strand. To begin with the former, if there are moral truths, clearly we must have a means of epistemically accessing

[47]Perhaps the issue is not as much a problem as it may be thought. In Dancy's eyes, the problem of isolated cases is one the realist can cheerfully take in his stride. He tells us that "something that is not the case can explain an action" is something that "we should countenance without too much reluctance. In doing so, we will be giving one sense to the idea that reality is practical" (Dancy 2000: 137).

[48]Mackie (1976: ch. 1).

[49]The second argument he proposes, the argument from relativity, is not relevant to non-moral values.

them. But defenders of realism have difficulties explaining how we can "be aware of [...] authoritative prescriptivity" (Mackie 1976: 39). Supposing that normative truths exist, what sort of knowledge do we have of them? It seems incredible to suppose that we have sensory perception of these facts in the same way in which we know about empirical or descriptive states of affairs. Neither can other well-known sources of knowledge, such as introspection, inferential reasoning or the testing of empirical hypotheses explain our knowledge. The defender of realism seems forced to accept the existence of a special epistemic faculty of moral or rational intuition. For Mackie the mental faculty thus posited would be peculiar as it deviates from the other sources of knowledge.

But to account for knowledge of causally inert, non-natural facts is not as great a challenge as Mackie seems to assume. The charge that our way of knowing moral facts is peculiar presupposes that we have a firm grasp of which epistemic faculty are natural and which are peculiar. But there is much disagreement about our sources of knowledge. Indeed Mackie assumes a strictly empiricist view of the world, a view that can no longer claim to be the uncontested default position. It is hardly clear, as Mackie appears to suppose, that a defender of the existence of objective values is committed to the existence of a faculty of normative perception which functions in a way analogous to sensory perception.[50] In particular, instead of adhering to the roughly foundationalist picture of knowledge presupposed by Mackie, it may be argued that we gain our knowledge of normative truths by considerations of coherence.[51] Moreover, as Mackie himself points out, moral values are hardly the only entities that are not easily accommodated in his empiricist scheme. For instance, our knowledge of the axioms of mathematics or of philosophical truths doesn't follow the model of the knowledge obtained through sense-perception or simple logical deduction. Thus it is possible to defend our knowledge of values by appeal to "companions in guilt": if the existence of values is ruled out because of their "queerness" then so are a good number of non-normative entities such as numbers or propositions. At any rate, with the epistemological argument Mackie accepts the burden of proof that these non-normative truths, if they are truths, do not equally constitute a difficulty for an empiricist epistemology.

The second, metaphysical strand of Mackie's argument is more convincing. Mackie argues that the fabric of the universe does not include values because they are "of quite a different order from anything with which

[50]Even the idea of rational intuition has had a recent comeback. Intuitionism in ethics has been the subject of a book-length defense. Cf. Huemer (2007).

[51]Cf. Schmidt (2004: 124–5).

we are acquainted" (Mackie 1976: 40). One aspect in which values differ from the paradigmatic properties that on Mackie's view do populate the universe — empirical, natural properties — is that they are not in the same way integrated causally. To mark values as peculiar based on their causal irrelevance, however, requires an extremely narrow view of what exists. As with the epistemological argument, this would rule out many other entities such as numbers. What, then, is the feature that leads Mackie to describe them as "entities or relations of a very strange sort, utterly different from anything in the universe"? (Mackie 1976: 38) The answer lies is the relation of these entities with motivation. As Mackie writes:

> An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. (Mackie 1976: 40)

If there was something like an objective good or value in the world, it would have to be able, as such, to incite our motivation. It would have to have a built-in property of to-be-pursuedness. But to Mackie this conclusion is absurd: nothing in the world has such a property. Having this property would make values vastly different from other things in the world, including mathematical objects.

Note that Mackie objects only to the existence of *objective* goods. A subjective good or value is one whose validity depends on the existence, in the agent, of a prior aim, an aim that is furthered by doing the good or valuable thing. For Mackie, the existence of things which are good in this sense is not problematic. Paying my debts is an action which is good relative to my desire to keep my contract. But the goodness depends on my wanting to keep my contract, a goal I have only contingently and as a matter of my preferences. Absent a subjective goal, it is no longer true to say that the action is good. For Mackie, then, the only reasons that exist have the force of hypothetical imperatives. To him, categorical imperatives are an illusion perpetuated by our tendency to objectify subjective values.

Mackie's goal is to argue against the viability of ethical theories that presuppose the existence of objectively valid values or an objective conception of goodness. He explicitly targets the Moorean view that moral properties are non-natural qualities. However, his criticism is also clearly applicable to the view that there are objectively valid reasons for action, which after all normative realism conceives as entities with prescriptive

authority. The argument from "queerness" applies also to normative realism. This is true, it must be added, only to the extent that the realist assumes that the reasons involve "oughts" which have the force of categorical imperatives. But clearly this is part of the conception of the version of normative realism we have been considering, which stresses the authority of reasons without regard to subjective aims. According to normative realism, reasons for action can, and often do, exist independent of any desires the agent has or may have. Our normative reasons are objective in just this sense.

To see what in Mackie's eyes is incredible about the alleged to-be-pursuedness built into goodness, consider the following line of reasoning. As we have seen, normative realism assumes that we can have knowledge of facts about normative reasons. Suppose then that it is a fact that I have a good reason to help my neighbor move, and suppose also that, perhaps because he asks me for help, I come to believe that I do; now I'm carrying a box. Given that I judge that this is what I have reason to do, how do we explain my actually coming to do so? It clearly is a feature of normative judgment, as opposed to empirical judgments, that they systematically cause us to act, and this feature demands explanation. As an explanation, we should say that there is a tie between judgments about reasons and motivation. This feature, which is often called internalism, is captured roughly by the assumption that it is true — and true in a non-accidental way — that if an agent judges that he has a good reason to do φ, he is apt to become motivated to do φ. The internalist principle is a description of the fact that we often become motivated to do as we take ourselves to have reason to do.[52]

That this principle poses a problem for normative realism is a result of the fact that on this view, whether *P* has a reason to do φ is independent of his actual state of motivation. This raises the question how realism can explain that the agent acts on his reason-judgments. If realism appeals, as it should, to the view that there is an internal connection between normative judgment and action, it needs to present us with an account of how normative judgments necessarily give rise to motivation. But its options are severely limited by the fact that it holds normative facts to be independent from the agent's aims. This is a difference between the two theories of reasons. Psychologism holds that what an agent has good reasons to do is a function of the agent's motivation, for, as we have seen, it assumes the primacy of motivating reasons over normative reasons. By contrast, realism has a realistic interpretation of normative reason

[52]More specifically, this could be called reason-motive judgment internalism. See also chapter 5.

statements, according to which these statements are true no matter what the agent happens to desire. Accordingly, it seems mere coincidence if an agent's motivation closely matches what he believes he has reason to do. After all, on this view beliefs about reasons are beliefs about an independent matter of fact.

Thus the realist view seems to imply that being motivated by normative judgments seems to depend on the agent's having some further antecedent desire which supplies the motivational force. For instance, the view could posit a general desire to do what one has a reason to do. Yet if this move can establish a connection between judgment and action, such a connection would not be internal. The existence of this link would be a matter of the individual psychological makeup, or perhaps it would amount to a tendency as a part of human psychology in general, with the possible exception of pathological cases. But even a desire that is embedded deeply in our nature would not explain that it is a non-accidental truth that normative judgment leads to action, for the connection would not be a consequence of the meaning of the normative vocabulary but depend on a contingent motivational disposition.

To meet the internalist condition, the realist may dig in his heels and stipulate that there is, after all, a link between normative belief and motivation which is not contingent on human psychology. As a last resort, he may proclaim that it simply is a brute necessary truth about the relevant concepts that an agent who grasps the fact that she has a reason is accordingly motivated. To do so is to point to an unexplained truth that normative facts have a built-in property of to-be-pursuedness. However if the realist is forced to say that, he does not provide a genuine explanation of the internal connection anymore. One cannot without circularity explain the necessary tie between normative judgment and action by saying that normative facts have a built-in property of to-be-pursuedness where to-be-pursuedness is the property of necessarily giving rise to motivation. Appeal to a property like to-be-pursuedness appears *ad hoc* and leaves the tie between normative thought and action utterly mysterious.

That realism divorces normative facts from motivation means that, on pain of resorting to a "non-explanation" involving to-be-pursuedness, it must admit that something further is required when moving from the realization that there is a reason to φ to the corresponding motivation. Compare the psychologistic theory we have examined. Recall that on this view statements about normative reasons can be understood in terms of motivating reasons. On Williams's view, X has a reason to φ just in case he would desire φ'ing if he were fully informed, had no false beliefs and deliberated correctly. If so, for an agent to come to believe that she has

reason to φ is for her to become convinced that she would desire to φ if she had access to the relevant facts and made all the inferences that practical rationality required. This realization is enough for her to immediately move to being motivated to φ; no further goal is required. She only has to follow out the consequences of what she already believes. One cannot consistently accept that one would be motivated to φ and still be unmoved to φ. To do so is to incur charges of irrationality. For psychologism, there is a conceptual connection between normative judgment and motivation.[53]

Mackie's argument that normative facts would be epistemologically or metaphysically "queer" is not by itself convincing because once we abandon an overly rigid naturalism, we no longer need to see normative facts as such as problematic. But insofar as Mackie's point is that realism must explain the internalist tie, the argument displays a decisive weakness of realism. If what was said is right, Mackie's point that a realist view would have to posit a peculiar property of to-be-pursuedness derives its persuasiveness from the need of every theory of reasons to account for internalism. The realist's metaphysical commitments, which alone would not constitute a decisive argument against the view, stand in the realist's way of an illuminating explanation of the internal connection between normative judgment and motivation. We should interpret Mackie's argument as pointing out that normative realism doesn't have the conceptual resources required to support the claim that there is an internal motivational link of beliefs about reasons. On the assumption that there is such an internal link, we can conclude that realism is not an adequate theory of reasons.

[53]Smith (1994: ch. 5) proposes a theory of normative reasons that has much in common with Williams's proposal. Building on Williams's view, Smith replaces the individual constraints by what he calls a "summary" analysis. He holds that an agent has a normative reason to do φ just in case she would desire to φ if she were fully rational. The conditions of full rationality, he says, cannot be made fully explicit. They include but go beyond the types of reasoning Williams countenances. Instead of the narrow view held by Williams, Smith allows for what he calls "systematic justification" of desires. Deliberation of this kind makes use of the Rawlsian idea of reflective equilibrium. In other words, we can rationally come to have a new belief by considering how well the new desire would fit in or cohere with the existing set of desires, an idea that also involves explanatory relations between general and specific desires.

This proposal is more liberal than Williams's because it allows for greater differences between an agent's normative reasons and his actual existing motivational set. But like Williams's view, Smith's more liberal conception of normative reasons can account for the fact that in reasons the intentional and deliberative perspectives come together. For him *X* has a reason to φ iff she would desire to φ if fully rational. Given this analysis, it is even plainer that failure to be moved by the thought that there is a reason to φ is impossible without incurring charges of irrationality. Smith explicitly takes this feature as an advantage of his theory (Smith 1994: 177ff).

1.5 Internalism

What has been shown by the preceding discussion is the conditional claim that *if* internalism judgments is right, realism fails to fulfill its task of explaining the conceptual connection. What can be said in favor if this view?

"Internalism" means different views to different writers.[54] In this section, we will be concerned exclusively with what is often called judgment internalism.[55] According to judgment internalism, an agent's judgments about reasons are related to his or her motivation.[56] It follows that a person cannot candidly assert, or believe, that he ought to do an action without, in a sense to be specified, being motivated to do so. Judgment externalists typically agree that agents who judge that there is a reason often are motivated to comply with it, but deny that this connection is as tight as internalism claims. Agents regularly comply with their own reason-judgments but the fact that they do is not a matter of necessity. For externalists, agents who are motivated by their own normative judgment have a further motivation that is relevantly related to their reason.[57]

It is intuitively plausible that prescriptive concepts such as "ought" or

[54]Cf. Darwall (1997).

[55]Judgment internalism must be distinguished from existence internalism, the view that whether or not an agent has a reason depends on his motivational states, i.e. on his desires. For a discussion of existence internalism, see §5.2–3.

[56]The internalist principle under that name was introduced into the debate by Falk (1947). However ideas in the vicinity such as the claim that goodness has "a certain magnetism" (Stevenson 1952: 417) were common even before Falk's article. As Darwall, Gibbard, and Railton (1992) point out, the topic is already present in an inchoate form in Moore's open question argument (Moore 1959: ch. 1). We can find the idea that rationalist theories of moral terms are open to objections based on judgment-internalism already in Hume's Treatise:

> In order, therefore, to prove, that the measures of right and wrong are eternal laws, obligatory on every rational mind, it is not sufficient to shew the relations upon which they are founded: We must also point out the connexion betwixt the relation and the will; and must prove that this connexion is so necessary, that in every well-disposed mind, it must take place and have its influence; though the difference betwixt these minds be in other respects immense and infinite. (Hume 1978: 3.1.1, 465–6)

Hume adds that "we cannot prove *a priori*, that these relations, if they really existed and were perceiv'd, wou'd be universally forcible and obligatory" (Hume 1978: 3.1.1, 466). Darwall (2002) calls the difficulty of rationalists moral theories to accommodate the internal links "Hume's challenge". For an attempt to develop a topology of the different internalist views, see Robertson (2001).

[57]"Internalism" in the rest of this chapter should be understood as referring to judgment internalism rather than existence internalism.

"reason" have an internal connection to motivation. It seems part of the essence of these notions that self-directed judgments involving "ought" and "reason" affect our motivation. Furthermore, we can observe that the connection between normative judgment and action is strong and reliable.[58] At least in the majority of cases, motivation reliably tracks normative judgment. If an agent changes his normative beliefs about what to do, we typically expect his motivation to shift accordingly. Internalism is in a particularly good position to explain the reliability because it posits an internal connection.

For this reason, judgment internalism has been regarded as the standard view for most of 20th-century moral psychology, a view that hardly requires argument. However, despite its intuitive plausibility, judgment internalism is sometimes challenged.[59] According to judgment internalism, an agent who judges that she ought to do φ has a motivation to do so; and it is not merely accidentally true. What does it mean, in this context, that the agent has a motivation? Clearly it cannot mean that the agent actually does φ. There are many circumstances that may prevent an agent from φ'ing. If these circumstances are present, the fact that the agent doesn't φ is no indication that he doesn't judge that he ought to φ. For instance, the agent may be physically prevented from φ'ing, through the actions of another person, physical force or even through paralysis. Here we would say that, despite his inaction, *S* nonetheless has a motivation to do φ. Still, motivation is close to behavior: if no preventing circumstances are present, there is a strong expectation that an agent with motivation to φ will in fact φ.

An action is typically the product of practical reasoning, in which the thoughts involved can be arranged on a scale, depending on how close to action they are. The one extreme is action or being motivated to act. This motivation which is apt to be turned into reality is the intention to do φ here and now. The agent typically doesn't form intentions spontaneously, however. Instead the process leading to a decision, the process which concludes in motivation and (typically) action, involves other thoughts. One such thought may be that one has a reason to bring about state of affair *X* and that doing φ is one way, perhaps the best way, to achieve *X*. This belief, too, is a normative judgment, albeit one which is one or more places removed from action. Of course not all normative

[58] As Smith writes, "it is a striking fact about moral motivation that a *change in motivation* follows reliably in the wake of a *change in moral judgment*, at least in the good and strong-willed person. A plausible theory of moral judgment must therefore explain this striking fact" (Smith 1994: 71). For Smith, externalism has great difficulties in accounting for this observation.

[59] For a discussion of the challenges, see Smith (1994: ch. 3).

judgments automatically move us to action. A mayor can judge that there is a reason to repair potholes in the streets without being so motivated if he also judges that there is another, stronger *prima facie* reason to do something else, perhaps to keep the fire department from running out of funds. A reason-judgment is only necessarily motivating, as demanded by internalism, if the reason involved is a conclusive reason. The deliberative chain leading to the decision may incorporate different *prima facie* goals and different means, in the form of another normative judgment at some distance from the action.

Now this array of normative judgments is linked to the action logically. Judgments about relative strengths of reasons lead to judgments about what to do, and we move from one judgment to the other by practical arguments. The thought that the only options available are φ and ψ and the thought that φ'ing is better than ψ'ing all things considered entails the all-out conclusion that φ'ing is what ought to be done. Sometimes, as when in reasoning about what to do this evening we decide between two attractive options, the choice is free, but often, through practical reasoning, we find ourselves compelled logically to do something. The force of these lines of thought is logical force. The various normative thoughts in this deliberative chains are linked by inferential connections. Now we might say that depending on the proximity to the action, the different normative judgments have a closer conceptual proximity to the action. There is a conceptual connection between a desirability-judgment and the corresponding judgment, but this connection is comparatively weak. A stronger connection exists between the judgment that φ'ing is better than all other alternatives and φ'ing, and yet a stronger connection between forming the intention to φ and doing φ. There are, then, various internal connections in the vicinity of varying degree.

To return to the internal connection we are discussing primarily, can an agent assert sincerely that there is a conclusive reason to do φ and at the same time show no sign of being motivated to φ? Externalists have affirmed this possibility.[60] They insist that being motivated requires, in the agent, a certain antecedent desire or motivational disposition to do what one ought to do. Without such a tendency, an agent would not in fact be motivated to do φ when he judges this to be what he ought to do. Moreover, the fact that human beings have such a tendency (if they do) is merely a contingent fact about human psychology, albeit one widely shared. If so, it may still be the case that our normative judgments are reliably efficacious, but such a psychological regularity would only

[60]Cf. Brink (1989).

support an accidental connection between normative thought and action; this would not license talk of an internal connection.

Here, however, it is important to clarify the claim of externalists. Externalists deny that there is a conceptual connection between *S*'s judgments about what she ought to do and motivation. Because internalism originates in the discussion of moral psychology, what is typically meant by the word "ought" in the statement of internalism is a distinctly moral obligation. Externalists are skeptical about the necessary nature of moral motivation: for them it is possible, if perhaps unlikely, for an agent to think that doing something is right without taking this as something that has normative force for him. This can be seen by the kinds of counterexamples which externalists have taken to support their claims. Thus we may imagine an agent who judges that φ'ing would be morally right but has stopped caring about doing the right thing. In order to be able to make the transition from judging that φ'ing is right to doing φ, according to the externalist, an additional conative state is required. An imagined amoralist does not possess this optional motivational tendency. There has been much debate about whether an amoralist is in fact a genuine possibility — whether an agent could be entirely unmoved by an acknowledged moral good. An internalist reply might be that we have reason to deny that the unmoved agent really is making moral judgments; one might claim that the agent is using the moral expressions merely in an inverted-comma sense. Such an agent, the argument might run, does not really believe the action to be right but only that it is "right" in the sense of being the consequence of certain moral standards — standards that the agent may not endorse. The externalist, in turn, can reply that, if we can imagine an unmotivated person who sincerely asserts that the action is right, such a person cannot be a conceptual impossibility.

Notice, however, that the question whether internalism in this sense — concerning judgments about what is right — is true is distinct from a rather different internalism, the view that judgments about what one has a conclusive reason to do are non-accidentally connected to motivation. This latter view, which concerns normative judgments in general as opposed to specifically moral judgments, is the focus of the present discussion. This latter view, which may be called reasons-motivation internalism, is more easily defended than its ambitious moral cousin. Consider an attempt to sketch a counterexample to this view. This would be an agent who judges that she has a conclusive reason to φ but does not show any sign of being motivated to φ. If the amoralist classifies actions as right or wrong but does not care about what it is right to do, the arationalist would classify things as the things it would be rational or irrational

for him to do but does not care about what it is rational to do. To do so would amount to opting out of the project of being responsive to reasons. It is much harder to imagine such a person, who would be indifferent to her own reasons while staying, so to speak, in the business of intentional action. Here, even more than in the moral case, it seems plausible to conclude that the arationalist does not really make the normative judgment we are tentatively attributing to him. If a person persistently displays motivationally inert reasons-judgments, we are likely to look for explanations for his behavior other than indifference to his own rationality. The agent's assertions may be insincere, or he may not have mastered the normative concepts properly.[61]

Internalism about morality has the ring of truth, but as the existence of counterexamples shows, we cannot assume the truth of this position without detailed criticism of externalist proposals. The answer to the question depends on whether it can be shown that moral demands are a species of rational demands — whether to say that φ is morally called for is to say that it would be irrational not to φ. Externalists are apt to deny this relationship. According to them, it is possible for someone, without loss of rationality, to hold that φ is morally correct while taking himself not to have a reason to do φ.[62] Nothing we have said rules this possibility out.[63] Rather than deciding the question one way or another, I will instead point out here only that, at least on many views, the possibility of moral externalism rests on an internalism about normative judgments more generally. Even among moral externalists, normative judgment internalism is a commonly held view.[64]

It has, however, been pointed out that even normative judgment internalism faces counterexamples.[65] The view we have been considering holds that, by conceptual necessity, any agent who judges that she ought to do φ is motivated to φ, where the ought in the claim is the generic ought of rationality rather than a moral ought. This view seems too strong, as can be seen from two types of counterexamples commonly given. First, consider a losing squash player who values fair play and sportsmanship.[66] In a fit of uncontrollable anger, however, he wildly thrusts his racket at his opponent. He does this although he believes that it is not the correct thing to do, even by his own standards, and that it would be irrational to do

[61]Cf. Scanlon (2010: 14).

[62]A view that takes this to be impossible can be called judgment-reasons internalism.

[63]Wallace (2005) rejects moral judgment internalism while maintaining that there are substantial conditions on moral judgments.

[64]Cf. Darwall (1997: 307).

[65]Cf. Smith (1994: 137ff).

[66]The example is adapted from Watson (1982).

so. Even though he has every opportunity to remain friendly, the player acts against his normative judgment. According to the strong internalism we have been working with, the situation must be badly described. The view says that it is impossible to genuinely make the judgment in question without being motivated to remain calm. Should we say that in this case, the agent did not really judge that he had a reason not to throw the racket? Second, consider a student who judges that he has a conclusive reason, vividly present to him, to study for his exams but instead keeps lying on his couch. Does that mean that his assertion that he ought to study is not genuine? This reproach may well be justified: he may be lazy or in fact not invested in his work. This would mean that his real preference is to do nothing. However, it also seems possible that the case is more serious. Thus the agent may find it impossible to muster the will to study because of a profound sense of despair or desperation. Such a psychological malady would in effect prevent the agent from developing the relevant motivation. Michael Stocker lists a number of conditions that can have this effect:

> Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. (Stocker 1979: 744)

According to Stocker, we can describe an agent who fails to be motivated to φ by a normative assessment because of some form of depression as one who still judges that φ'ing is desirable despite his total lack of motivation. Depression is a state where what is seen as good is not felt as having any motivating effect. It would be a mistake to say that because of his inaction, the agent doesn't really believe that he ought to φ. If Stocker's description of the case is correct, it is after all possible to judge sincerely that there is a conclusive reason to φ and yet lack a motivation.

Both cases point to a deficiency in internalism as we have conceived it. As the squash player and student show, it is not true without exception that an agent who believes he should φ is automatically motivated accordingly. Still, the cases are somewhat special in that in either case there seems to be something amiss with the agent. The agent's motivation is out of step with his normative assessment of the situation. The squash

player finds it impossible to resist the urge to thrust the racket at his opponent even though he thinks that this is not what he ought to do. This points to a rational tension in the agent: the akratic gap between judgment and motivation is not just the failure of a regularity but a case of the agent failing to do what he knows is rationally required to do.[67] The student, while not having an irresistible urge, also fails to be responsive to his own normative judgment, even if he is not physically prevented from doing so. To begin, this means that these cases deviate from normal situations. The examples are clear because we suppose that, in their normal non-akratic state, agents would be motivated appropriately. Suppose that the squash player is just a choleric person in general who never holds his emotions in check. If the player consistently acts angrily and aggressively, we would, after some time, stop treating him as weak-willed; we would, in effect, begin to treat his claims that he has a reason to be moderate as spurious. The depressed student is only in a temporary phase of indifference. The attribution of a genuine judgment that studying is the thing to do is apt to become doubtful and eventually cease if the agent continues in his ineffectual assertions for too long, remaining indifferent to his own professed good. Judging an agent to be akratic requires a background of successful attributions enough of which lead to motivation; otherwise the attributions become hollow. It seems essential that we have in mind a normal case with which to contrast the behavior.[68]

Nonetheless these cases involving weakness of the will illustrate an important point about our normative judgments. On the one hand, we have a justified expectation that these judgments are followed by the motivation they are connected with. The connection is reliable, but, more strongly, it is also a conceptual connection. It is essential to beliefs about one's own reasons that these beliefs are connected to action by way of inferential connections. Such a belief would not be the mental state it is if it didn't stand in the inferential relations to immediately motivating intentions.[69] There are, as we may say, implications running from reasons-judgments to motivation. The naive expression of such a principle is:

Strong Internalism If X judges that he has a conclusive reason to φ, then he is *ipso facto* motivated to φ.

We can call this version of the thesis *strong internalism*. However as is

[67] Akrasia will be the subject of §4.2.

[68] The relation to normal conditions is stressed by Dreier (1990).

[69] For an account of the puzzling idea that the relation between a reasons-judgment and the corresponding motivation can be conceptual, see §5.1.

now clear, this view cannot be true as it stands because there are agents who make the judgments without being appropriately motivated. The connection is not as strict as the formulation implies. The principle needs to be qualified:

Weak Internalism If X judges that he has a conclusive reason to φ, then, *ceteris paribus*, he is motivated to do φ.

According to *weak internalism*, there are exceptions to the general thesis. In other words, there are conditions that prevent the agents from making the correct step leading from the normative judgment to the motivation. Although there is a strong relation, the relation is defeasible if certain conditions are present. The exceptions, of course, include at least the two kinds of counterexamples we have been considering. An intensely emotional state such as anger may prevent the squash player from being able to stop himself from doing something that he deems wrong. In such an agitated state of mind, he fails to be rational by his own lights. He does not react to his own judgment in the way that, as he is aware, reason calls for. Similarly, the depressed student is prevented by his disaffected state from developing motivations that are rationally required, even by his own lights. Accidie is a form of irrationality.

To summarize, strong internalism is the view that if X judges that he has a conclusive reason to φ, he is *ipso facto* motivated to do so. This means that such a judgment *is* a kind of motivation. Strong internalism implies that the scenarios of the squash player and the depressed student are badly described, but we have good reasons to say that these cases are genuine. If strong internalism posits too rigid a connection between the two, we should replace it by weak internalism. On this view there is an internal connection between reason-judgment and motivation, but it is defeasible. Normative judgment and motivation are separate states which can come apart.

Hence it is impossible only for a *rational* agent to assert candidly that there is conclusive reason to φ without being motivated to φ. There is a requirement of rationality that the agent become motivated in accordance with his normative beliefs, a requirement that, like any genuine requirement, it is possible to violate.[70] One may object that the relation posited by weak internalism is not a conceptual connection. But this objection is groundless. The fact that statement has a *ceteris paribus* rider does not in any way affect its status as a truth about the logic of the concepts and states involved. Though it leaves the possibility of exceptions

[70]For an elaboration of the notion of a rational requirements, see chapter 6.

due to akrasia, it is still true that a *rational* agent is necessarily motivated by his normative judgments. It is a matter of conceptual truth that if you judge you have a conclusive reason to φ, either you are motivated to φ or you are irrational. Far from being implausible, it is in fact to be expected that there is the possibility of having the one state without the other. As these conceptual connections are *rational* links, they are not without exception, but neither are other conceptual connections such as "If you believe that X is a triangle, then also believe that it is a polygon". Even for someone who has mastered the concepts, the two beliefs can come apart, though not without impacting the subject's rationality.[71]

It may be thought that weak internalism presents a smaller challenge for normative realism than the stronger variant. But realism is in no position to explain weak internalism. For how can it explain that an agent who grasps a reason-fact without being motivated must be irrational? It can, of course, simply stipulate that these facts are such that it would be irrational to look them in the eye unmoved. But instead of giving an explanation, it assumes a property of to-be-pursuedness-on-pain-of-irrationality, which once again leaves the phenomenon of motivation by normative considerations mysterious.

1.6 A Sellarsian conclusion

In this chapter, we have compared two theories of reasons. We have discussed a series of arguments. The phenomenological argument against normative realism has been found inconclusive. On the other side, Dancy's argument against the three-part story did not survive critical scrutiny. Next, Mackie's objections against the metaphysics and epistemology of normative realism, though themselves not convincing, prompted us to examine the ability of realism to provide an explanation of the internal connection between normative judgment and motivation. Normative realism proved unable to explain internalism.

Even with normative realism ruled out, we should not uncritically accept psychologism. We have already seen a difficulty with some versions of psychologism: *act*-psychologism fails to reveal the sense in which normative reasons and motivating reasons are two sides of the same coin. Content-psychologism avoids this difficulty.[72] However, act-psychologim

[71] For more on this point, see §5.1.

[72] That the theory of reasons developed below is a version of content-psychologism will also be clear from the discussion of rational requirements in chapter 6.

has historically had numerous defenders, many which subscribe to some form of Humeanism. In chapter 2, we will explore the broadly Humean approach taken by most contemporary defenders of the thesis that reasons are constituted by states such as beliefs and desires. Chapters 2 to 4 constitute an extended argument that Humeanism, as it is mostly understood, is unsatisfactory.[73]

A short digression to Wilfrid Sellars's metaethics will be helpful to map out a course for the argument of the following chapters. In his article "Imperatives, Intentions, and the Logic of 'Ought'", Sellars sets up a contrast between two polar opposites and two of the most important metaethical views of his time, rationalist intuitionism on the one hand and empiricist emotivism on the other.[74] He discusses both positions with a view to the question of how they relate to judgment-internalism. Like normative realism as we have conceived it, intuitionism regards normative or moral facts as objectively independent states of affairs we can grasp, whereas emotivism provides a non-cognitive analysis of moral statements in terms of the expression of attitudes of approval or disapproval.[75] For the emotivist, the linguistic performance one produces when asserting that one ought to φ is not an assertion, or the expression of a belief, but the expression of a non-cognitive state of motivation. So on the emotivist view, there is a trivially internal connection between normative judgments and motivation. In fact, emotivism can claim with some justification that according to it judging that one ought to φ is a way to be motivated to do φ. The emotivist analysis has the great virtue of accounting for the phenomenon of judgment-internalism.

There is a strong resemblance between the dispute between intuitionism and emotivism, as reported by Sellars, and the dispute we have been considering, between normative realism and psychologism. Normative realism, just like intuitionism, is facing great difficulties regarding the relation of normative judgments to motivation. Now for Sellars, the argument has a flip side. Emotivism, even if it excels at explaining the relation of prescriptive thought to motivation, has difficulties of its own as its analysis of normative judgments sees them as "judgments" really only in an attenuated sense, indeed in a sense that bears little resemblance to full-blown theoretical judgments. It conceives sentences containing the word "ought" as the expressions of favorable attitudes towards a certain state of affairs which themselves are not integrated into our reasoning

[73] For a tentative return to the topic of psychologism, see §3.5.

[74] Sellars (1963b).

[75] Ayer (1990: ch. 6) was influential for the development of emotivism. Ross (1973: ch. 1–2) is a sophisticated statement of intuitionist non-naturalism.

processes. According to emotivism, as compared to descriptive discourse prescriptive vocabulary is a second-class citizen, which has only "emotive meaning" rather than descriptive meaning. But for Sellars, this is a profound mistake. He considers it important that "'ought' has as distinguished a role in discourse as descriptive and logical terms" (Sellars 1957: 282). He stresses that "we reason rather than 'reason' concerning ought". In other words, emotivism mistakes normative judgments for the expressions of brute inclinations or tendencies without *bona fide* rational relations connecting them.

Yet as Sellars points out this is manifestly false: we reason about the normative — after all, we know not just the theoretical syllogism but also the practical syllogism —, we argue and disagree among ourselves, our methods of reasoning are rich and we have at our disposal substantial resources of normative concepts which are no less inferentially articulated than their empirical counterparts. Judgments about reasons obey the standard rules of quantificational and propositional logic.[76] The part of the intuitionist in Sellars's dialectic is to insist on precisely these discursive features of prescriptive discourse. The intuitionist captures this point well by insisting "on the truly propositional character of prescriptive statements, as over and against the emotivist contention that ethical concepts are 'pseudo-concepts' and the logic of moral discourse is 'pseudo-logic'" (Sellars 1963b: 162). Intuitionism emphasizes the parallels between theoretical — e.g. mathematical — reasoning and practical reasoning and is thereby in a good position to explain the argumentative, discursive phenomenology of practical discourse. This is as much a difficulty for emotivism as it is a merit of intuitionism.

Emotivism is in danger of underestimating the cognitive nature of normative judgments and of practical attitudes in general. An analogous line of reasoning suggests that a similar danger threatens psychologism, and in particular Humean varieties of psychologism. There are, of course, disanalogies between contemporary Humeanism and the emotivism of Sellars's day.[77] The modest point now is that there is a structural analogy between the two polar opposites that Sellars compares, on the one hand, and the views discussed here, on the other. Lines of thought similar to those brought to bear by Sellars also apply, with the appropriate modifications, to normative realism and psychologism. We have already seen

[76] Cf. Scanlon (2010: 10).

[77] The views under discussion, psychologism and normative realism, are theories about reasons and the rational ought in general, whereas the views Sellars is concerned with are primarily interested in moral obligations. What is more, psychologism does not propose the same analysis of prescriptive discourse as emotivism, and in any event expressivist philosophy has come a long way since its emotivist roots in the writings of Ayer and Stevenson.

that normative realism suffers from inadequacies not unlike those Sellars found in intuitionism: the metaphysical and epistemological extravagances make a mystery out of the crucial relation between normative judgment and motivation. As I will attempt to show in the next chapter, we have reason to object to many versions of psychologism on grounds that are similar to those Sellars has for being dissatisfied with empiricist view of practical discourse.

Sellars presents the controversy between intuitionists and emotivists as a dilemma. If we choose the intuitionist's view of normative judgments, we are able to account for the discursive nature of prescriptive language. To do so, however, is also to accept metaphysical commitments that make it impossible to elucidate that the connection between thinking and doing is "a matter of strict logic" (Sellars 1963b: 162). On the other hand, adopting the emotivist analysis of prescriptive 'judgments' allows us to clarify the connection between the meaning of normative vocabulary and motivation. But if we accept this view, we obscure the rational integration of practical discourse, the fact that moral propositions logically entail, and are logically entailed by, other propositions. For Sellars, we do not need to embrace either horn of the dilemma. Neither extreme allows us to appreciate the idea that practical reasoning actually deserves its name by at once being *practical* in its outcome and, at the same time, a genuine form of *reasoning*. Intuitionism fails the first requirement, emotivism the second.

The suggestion made by Sellars is that there is a metaethical theory that combines the strengths of both positions while avoiding their respective pitfalls. Although the details of how his account of metaethics can achieve this dual goal are far from clear, he emphasizes the importance of an "adequate philosophy of mind" for this project and spends the rest of his discussion sketching an account of the inferential interrelations between normative concepts and concepts expressing practically oriented psychological states, in particular intentions.

The analogy furnishes us, not with an argument against psychologism, but with a plan of how to develop an adequate theory of reasons. For the present topic, the Sellarsian conception allows us to draw three tentative lessons. As in Sellars's discussion, the choice between normative realism and psychologism in its most common form represents a false dichotomy. There is a *via media* between the rationalist normative realism and empiricist psychologism. As in Sellars's story, there are two crucial requirements. First, an adequate theory of reasons needs to account for the internal connection between reasons-judgments and motivation.

As we have seen, normative realism falls short of this goal.[78] Second, an adequate theory of reasons needs to show how reason-judgments are essentially integrated into our practices of reasoning. If there is a structural analogy between Sellars's discussion and ours, we have reason to suspect that there are analogous problems with psychologism in this area. Along these lines, the second chapter will argue for the conclusion that Humeanism has difficulties meeting this second goal. If neither Humean psychologism nor normative realism are adequate, we need to look for an alternative that meets both conditions simultaneously. The third chapter offers an attempt to sketch an alternative theory of reasons of this sort. Finally, we can draw from Sellars's article the further lesson that in order to remove the appearance of a dilemma, we need to question assumptions in the philosophy of mind. As a reflection of this point, the criticism against psychologism in chapter 2 will proceed by examining critically the notion of desire, the psychological state Humeanism sees as central to its project. The goal of the rest of this thesis will be to meet Sellars's twin requirements — illuminating internalism and showing intentional action to be the result of reasoning in the full-blooded sense of the term — by developing, even if only in outline, a theory of reasons based on a more adequate conception of an action-rationalizing state of mind.[79]

[78] See §5.1 on how to meet this goal.

[79] A crucial part of this project will be the introduction of the notion of a practical commitment, which is defended in §3.4.

Chapter 2

Low-brow desires

2.1 Skepticism about practical reasons

Humeanism — perhaps the dominant form of psychologism in the theory of action — assigns desires, or desire-like states, a critical role in its view of intentional agency. On a Humean view, reasons for action are inexorably linked with desires. To explore the nature of reasons further, we need to assess the Humean conception of reasons as well as the nature of desires.

The present chapter begins this task by introducing (§1) the distinctive view Humeans take on practical reasons (in the plural) as well on practical reason (in the singular). Next we explore common motivations for the Humean view (§2). As the mental state of desire plays a crucial role in the Humean account, we proceed to two common conceptions of the mental state of desire. After showing inadequacies of the phenomenological conception (§3), we focus on dispositionalism, the most defensible conception of desire proposed by Humeans (§4). The emerging suspicion that even dispositionalism cannot support the Humean thesis is confirmed as we discuss a number of defenses of a dispositionalist version of Humeanism (§5). The discussion of a final attempt to leverage the instrumental principle to strengthen the Humean view of agency, which again turns out to fall short of achieving its goal, concludes our review of Humean ways to understand desire (§6).

As I understand it, Humeanism consists of two related claims, the first

about reason*s*, the second about reason:[1]

Desire-based reasons thesis Desires, and only desires, provide agents with reasons.

Instrumentalism There is a form of genuinely practical use of reason, but it is restricted to means-end rationality.

The first claim is that all our practical reasons have their source in the agent's desires.[2] This involves the negative point that an agent has a reason only when he also has a desire of the relevant sort, as well as the positive point that we have our reasons by virtue of wanting to do things.

What does the Humean mean when he says that desires provide us with reasons? To see this we have to turn to the second thesis associated with Humeanism. Humeanism is at least partly defined by a skeptical attitude towards reasons for doing things but also towards the process of reasoning and the associated mental faculty of reason or rationality.[3] It should be clear from the outset that talk about reasons cannot be divorced from talk about reasoning — and indeed from talk about the faculty of reason. An answer to the question "What reasons for acting do we have?" is also an answer to the question "How do we reason about action?" as well as the question "What is the scope of practical reason?"

Humean skepticism is directed at what is perceived as an overly inclusive conception of practical rationality. We can understand this, historically, as a response to rationalists who often took it for granted that we can arrive at practical conclusions through purely rational procedures, by being

[1] See the end of this section for reservations about the term "Humeanism".

[2] Citing Gilbert Harman, Stephen Darwall, who coined the expression but doesn't espouse the view, defines the desire-based reasons thesis as "the doctrine that the only reasons for an agent to act are those which [...] 'have their source in the agent's desires'" (Darwall 1985: 27). Defending a version of this thesis, Harman writers:

> If S says that (morally) A ought to do D, S implies that A has reasons to do D which S endorses. I shall assume that such reasons would have to have their source in goals, desires, or intentions that S takes A to have and that S approves of A's having because S shares those goals, desires, or intentions (Harman 2007: 38).

[3] Korsgaard lays stress on the phrase "skepticism about practical reason" in her seminal Korsgaard (1996). It should be noted, however, that while Korsgaard, who revisits the theme in Korsgaard (2008), calls Hume a skeptic about practical reasoning, in the earlier essay she does not fully acknowledge the full extent to which Hume is skeptical about the scope of reason. In the later essay, she sees Hume in a more radical light. Below I will distinguish skepticism *tout court*, which applies to contemporary instrumentalists as well as to Hume, from the more radical practical skepticism defended by Hume himself.

in contact with eternal truths and without thereby relying on any preexisting attitudes. The Humean queries the justification of these allegedly rational procedures. Instead of thinking, as the rationalist does, that reason alone can arrive at a particular practical conclusion, the Humean imposes strict constraints on what counts as the proper use of rationality.

The instrumentalist view of practical reason gives expression to Humean skepticism about excessive conceptions of what constitutes proper reasoning. Jean Hampton usefully proposes a definition of instrumentalism as making three claims:

> 1. An action is rational to the extent that an agent believes (reasonably) that it furthers the attainment of an end; *and*
> 2. Human reasoning involves the determination of means to achieve ends in a way described by the theory; *and*
> 3. These ends are in no way fixed by reason operating non-instrumentally, i.e., what makes them our ends is something other than reason. (Hampton 1995: 57)

Hampton's *second* point is a recognition of the fact that human beings engage in practical reasoning: they make inferences leading to practical conclusions. Virtually everyone accepts that we, as agents, frequently engage in this type of mental process:

> I shall do φ. Doing φ requires that I do ψ. So I shall do ψ.

is accepted by all parties. What is not uncontroversial is the exact way of demarcating practical reasoning. However, there is no doubt that "To kill my aunt, I need to poison her" counts as a piece of instrumental reasoning, so despite any quibbles, we have a fairly clear idea what instrumental rationality consists in.[4]

[4]The core or, for some, even the whole of instrumental reasoning is a species of (or application of) causal reasoning. Its primary or only purpose is to discover what it takes to bring about a certain state of affairs, and to answer this question, causal knowledge is required. In this way, instrumental reasoning is very similar to regular *theoretical* reasoning about causes and effects, so we could call it the flip side of theoretical reasoning. Some authors have a narrow conception of what constitutes instrumental reasoning (restricting it to means-end relationships strictly speaking), while others count other sorts of relationships as instrumental as well. For instance, Williams (1981a) argues that an important element of practical — thus for him, instrumental — reasoning is reasoning about what *constitutes* doing a certain activity. If the agent's goal is to spend a pleasant evening in town, his going to the opera may or may not conform to his understanding of this; but if it is, it would not be entirely correct to say that going to the opera is a means towards spending a pleasant evening. Instead, the latter consists in the former activity.

According to the *first* point in Hampton's list, an action is rational only to the extent that, on the agent's view, it is instrumentally useful to the attainment of an end. The instrumentalist view is that reason can only dictate or veto an action insofar as it is seen as the only way, or best way, to attain a given end. This principle limits the ways in which we can evaluate an action as rational or irrational, correct or incorrect. Instrumentalists regard it as inappropriate to criticize an action as irrational, or commend it as demanded rationally, unless we can point to an end that is purportedly promoted by the means. It follows that an action can be rational or irrational only in a restricted sense. Whereas instrumentalists countenance only rationality depending on predetermined ends, rationalists hold that there are further, non-instrumental ways for an action to be rational or irrational and, as a result, ways of criticizing or recommending action which are not relative to an end. Humeanism treats these claims with skepticism.

It is important to see, however, that in addition to the negative point, the second point also has a positive component. Rather than denying the applicability of questions of rationality to actions altogether, instrumentalism holds that it is possible to recommend or criticize an action rationally. On the instrumentalist view, reason *can* be practical, if only in a strictly limited sense: it can be practical insofar as it depends essentially on the starting point of a prior end. This, of course, is a severe limitation, but the view does not amount to a radical skepticism about practical rationality. A radical skeptic questions the idea of practical rationality in general. Instrumentalism stops short of the radical contention that practical reason is an illusion; it only asks us to agree that practical reason does not possess the scope rationalists would have us believe it does.

If, for instrumentalism, practical rationality is always contingent on the existence of prior ends, what kinds of ends are capable of making actions rational or irrational? How are they determined? The *third* point in Hampton's list stipulates that these ends cannot themselves be the product of reason acting in any way except through causal or instrumental considerations. It is easy to see that this provision is necessary to rule out certain rationalist views. Suppose that we could answer the question "Which ends does the agent have?" by pointing to facts about what it is reasonable for the agent to do, viz. facts that are themselves deemed independent of the agent's prior motivations. If so, the rationalist could easily short-circuit the first instrumentalist claim by arguing that what ends the agent has is a matter of their being rationally required. As a result, it would still nominally be true that only means-end relationships give a proposed action the cachet of being rational, but the ends in ques-

tion would themselves be open to further rational criticism and argument. Consequently, the question what it is for an action to be rational or irrational would be deferred to the question what ultimately determines ends, leaving it open to the rationalist to introduce a crucial active role for reason.

To block this possibility, Humeanism concedes that rationality can come into play concerning the question what ends an agent has, but hastens to add that the possible role of reason is very limited. There are derived ends and underived ends, a derived end being one that is held as a consequence of having another end, by way of its means-end relations. Thus even for the instrumentalist, a derived end *can* be unreasonable, but only in the limited sense that it is derived from a further end by way of improper instrumental reasoning. The agent may have made a mistake of fact in supposing that there is a relevant causal relation between φ'ing and ψ'ing. Derived ends are ends only insofar as they can be traced to an ultimate, underived end. On the other hand, an underived end is entirely outside the scope of rational criticism. If the agent has an end, not because of its being instrumentally useful to promote another end, but because he pursues it as an end in itself, it is inappropriate to criticize him rationally.[5] Criticism of ultimate ends would be entirely improper, akin to a category mistake. Nor can the ultimate end itself be recommended as the right end; this, too, is ruled out by the third point. Practical rationality does not fix ends except instrumentally.

Instrumentalism, then, amounts to a reassessment of the role of the faculty of reason in deciding what it is rational for an agent to do: the faculty of rationality — what Hume calls the understanding — has only a limited say in deciding what it is rational to do; the question is mostly decided, not by reason, but by predetermined extra-rational ends. Instrumentalists often take inspiration from Hume's famous remark in the section of the *Treatise* entitled "Of the influencing motives of the will" that

> reason is, and ought only to be, the slave of the passions, and can never aspire to any other office than to serve and obey them. (Hume 1978: 2.3.3, 415)

In this passage, Hume's topic is the comparative importance of two mental faculties vying for control: reason or the understanding, on the one hand, and sentiment or the passions, on the other. Before the famous passage, Hume mentions the tendency of earlier writers to emphasize

[5] This leaves open the possibility of other, non-rational forms of changing a person's ends such as bullying, persuading, brainwashing and so on. See chapter 5.

the struggle between these two faculties.[6] According to this traditional model, when we are faced with unpleasant choices, our passions often threaten to overwhelm our ability to control our actions. When this happens, we often give in to our passion although we are aware that doing so is contrary to what rationality tells us, thus becoming enslaved to our urges, with reason, the rightful master, losing control. Now with his remark, Hume turns the traditional imagery upside down. Against the rationalists, he insists that the situation is not one of struggling between following a passion and following reason. Reason is not the master of the passion in the sense of the faculty that controls the passions. Nor does it make sense to say, as a moralist might urge, that although the reality of our weak psychology may often be different, reason by rights *ought to be* the master of our passions. This conveys the wrong picture of agency. According to Hume, reason has, from the very outset, only an auxiliary role in determining the course of our action. Our passions determine the direction of our action; reason's role is simply to find the best way to get there.

If reason, which only adjusts means to ends, has no substantial, or non-instrumental, role in the process of setting the ends themselves, the agent's aims are a matter of extra-rational fact. Hume adds that the ends are constituted by the agent's passions or, as it is most often put today, by his desires. Again we plausibly have to countenance derived as well as underived desires, the latter serving as the ultimate anchoring point of the former. But it seems clear that, in the final analysis, we can trace back any intentional explanation of an action to the agent's ultimate desires. Hume explains:

> [T]he ultimate ends of human actions can never, in any case, be accounted for by reason, but recommend themselves entirely to the sentiments and affections of mankind, without any dependence on the intellectual faculties. Ask a man *why he uses exercise*; he will answer, *because he desires to keep his health*. If you then enquire, *why he desires health*, he will readily reply, because *sickness is painful*. If you push your enquiries farther, and desire a reason *why he hates pain*, it is impossible he can ever give any. This is an ultimate end, and is never referred to any other object.
>
> Perhaps to your second question, *why he desires health*, he

[6] "Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason, to give the preference to reason, and to assert that men are only so far virtuous as they conform themselves to its dictates" (Hume 1978: 2.3.3, 413).

> may also reply, that *it is necessary for the exercise of his calling*. If you ask, *why he is anxious on that head*, he will answer, *because he desires to get money*. If you demand *why? It is the instrument of pleasure*, says he. And beyond this it is an absurdity to ask for a reason. It is impossible there can be a progress *in infinitum*; and that one thing can always be a reason why another is desired. Something must be desirable on its own account, and because of its immediate accord or agreement with human sentiment and affection. (Hume 1975: Appendix I, 293)

According to Hume, an agent's ultimate ends are set by his desires. When we ask for a rational explanation for his aim, we may be referred to a further superordinate goal. But an appeal to a further end can only occur a finite number of times. At the end of such a chain of referrals, we find an end that recommends itself "entirely to the sentiment and affections" of the agent. This ultimate aim — perhaps to attain pleasure or to avoid pain — is itself not the appropriate target of rational challenges because its being an end is in no way attributable to the understanding but only to the brute fact of its being desired. We cannot give any justification for this. Nor is this fact to be lamented: for Hume it is an inevitable fact about human agency. Thus it is "an absurdity", when dealing with an ultimate desire, to ask for a justification: these states of the agent are simply there to be discovered.

We have now arrived at the sense in which, in the eyes of the Humean, desires are the only source of reasons. The two ideas out of which the Humean view is composed — the desire-based reason thesis and the instrumental conception of practical reason — naturally complement each other.[7]

The question we will be concerned with in this chapter is the question whether the Humean theory can be maintained and, in particular, whether

[7] They do not, strictly speaking, entail each other. It may be possible to be an instrumentalist while holding that the source of reasons is not the agent's set of desires but another extra-rational source — perhaps our ends may be thought to be set by nature or through the command to a divine being. Conversely, there may be ways to combine the desire-based reason thesis with conceptions of rationality that countenance more (or less) than instrumental reasoning. Nonetheless, the two ideas clearly lend each other support. In particular, the instrumentalist idea that ends are fixed by extra-rational items lends support to the idea that we can name a source of practical reasons, and combined with the seemingly straightforward idea that desires or passions are extra-rational items, it is not a far leap from the idea that practical reason is limited to instrumental operation to assuming that it is desires, and only desires, that give us our reasons.

it is true that desires give us reasons. My conclusion will be that — at least on a construal of the concept of desire that fits naturally with the Humean picture — the desire-based reasons thesis is false. But before proceeding to a detailed examination of this thesis, it will be useful to refine the sketch of Humeanism by extending the space of theoretical options. The Humean theory can be contrasted with a more permissive theory which rejects the Humean's main claims. This more permissive theory, which I will defend starting at the end of this chapter, holds that not all our reasons are grounded in extra-rational desires; and it holds that the role of reason in the practical sphere is more extensive than the instrumentalist view allows. This theory portrays practical rationality as comparatively rich. By contrast, the Humeanism we have considered is more skeptical about practical reason and considerably limits its scope by rejecting all non-instrumental use of reason as illusory. Although we cannot criticize an agent for the ultimate ends he accepts, he can be incorrect, rationally speaking, by failing to taking the necessary means to ends he already has.

Going further on the scale of practical skepticism, however, there is another position which rejects even the idea that reason can be practical in this limited sense. According to this more radical skepticism about practical reason, there is not even instrumental rationality strictly speaking. On this radical view, even the failure to take the required means to accepted ends does not license rational criticism of the agent.

As I said, this view exceeds Humeanism in the radical nature of its skepticism about practical reason because it constitutes a wholesale rejection of practical rationality. Perhaps surprisingly, as a number of scholars have argued, it is this radical skepticism of practical reason that Hume himself actually defended in his *Treatise*.[8] It follows somewhat paradoxically that Hume does not count as a Humean.[9] For Hume, a passion *qua* passion cannot be evaluated as either reasonable or unreasonable. He supplies a famously extreme example:

[8]Cf. Millgram (1995), Korsgaard (2008) and Hampton (1995).

[9]The term Humeanism, as I introduced it and as it is used widely, is not the view of the historical Hume but a view held by contemporary philosophers of action that, although inspired by Hume's writings, is not identical to his view. One way to mark the distance between Hume's doctrine and Humeanism would be to call the latter view "Neo-Humeanism", as Hubin (1999) suggests, or to use the somewhat artificial lowercase variant "humean" (see Brandom 2000: Introduction). Because the term "Humeanism" is so well established, however, with for the most part little regard to the question of whether it coincides with Hume's doctrine, I will continue to use the term to denote the theory committed to instrumentalism and the desire-based reasons thesis. Contemporary Humeanism takes its inspiration from the suggestive passages I have quoted, and even if it doesn't coincide perfectly with Hume's point, it is fair to say that the view is, in many ways, a theory in Hume's spirit.

> Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. (Hume 1978: 2.3.3, 416)

Here Hume countenances only two ways in which a passion may properly be called unreasonable. For example, an agent, mistakenly thinking there is an apple in front of him, may consequently develop a desire to get the apple. In such a case, the passion exists because the agent's desire was stimulated as the result of a factually mistaken belief. On the other hand, an agent who intends to bring about X may do φ because he thinks that φ'ing will promote X although in fact there is no causal relation between the two. In this second type of case, the passion is based on faulty causal reasoning. In both of these ways, it is appropriate to criticize the agent's passion, but in each case the criticism is only directed at an element of the desire in question which is attributable to theoretical reason. The desire did not come about because of faulty *practical* reasoning.

Hume's example of scratching my finger illustrates the point. If I have a preference of doing φ over doing ψ, there is nothing that can be said, from a standpoint of rationality, against my having this preference. Either I have the preference, or not. Hume contends that it is wrong in principle to describe an end *as an end in itself* as incorrect from the standpoint of rationality. Summarizing his position, he writes "a passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment" (Hume 1978: 2.3.3, 416). So far the position described is in line with Humeanism as we have introduced it. However, Hume's position is soon revealed as more radical. The instrumentalist we have described may agree with Hume in principle that, other things being equal, a brute preference of a trivial action over as undesirable an outcome as the destruction of the world is legitimate. But the instrumentalist would want to add that, normally, other things are *not* equal. In particular, an agent normally has a multitude of desires. What makes Hume's example so surprising is that the destruction of the world conflicts in the most blatant way with a great many desires almost universally shared. Thus surely a normal agent wants to preserve the lives of his close relatives and wants to enjoy pleasures in the future — desires that would certainly not be fulfilled if the destruction of the world were imminent.

Thus although in a preferential vacuum Hume's claim is true, the situation is clearly underdescribed. A typical agent has goals whose attain-

ment is precluded by the action preferred and he can be assumed to know it. More fully described, then, contrary to Hume's assertion, the agent is open to charges of irrationality — or so the instrumentalist would argue. The criticism would not be that the agent's goals are irrational in themselves. Recall the instrumentalist thesis that an action or passion is rational only to the extent that it (is known to) promote an end accepted by the agent so that practical irrationality is always instrumental irrationality. The instrumentalist criticism would be that the agent's action cannot be squared with his other, crucial goals. On any instrumentalist view, the agent's preference for scratching his finger would be instrumentally irrational because the agent knows that choosing it is incompatible with pursuing other ends.

Although Hume does not intend the "destruction" example in that way, the conflict described in the passage makes a strong case for the intuitive idea that an agent may be acting irrationally in pursuing one of his goals if other, more important goals are incompatible with it. For this reason, instrumentalists typically qualify their endorsement of the principle that an end cannot itself be assessed: they accept that it can be criticized on prudential or instrumental grounds. Hume, on the other hand, makes no such concession. Just after the passage cited, and before the summary, he writes:

> 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. A trivial good may, from certain circumstance, produce a desire superior to what arises from the greatest and most valuable enjoyment; nor is there any thing more extraordinary in this, than in mechanics to see one pound weight raise up a hundred by the advantage of its situation. (Hume 1978: 2.3.3, 416)

Here Hume clarifies his position. Given a choice between φ and ψ, there is nothing rationally problematic about preferring φ *even if* I judge at the same time that ψ is the greater good which promotes a value that is more important to me. Contemporary instrumentalist concede that choosing the "acknowledged lesser good" is an instance of practical irrationality because it conflicts with what I know to be the right choice, but Hume demurs. He does not even countenance blatantly failing to take the means

to an acknowledged end as a case that makes it appropriate to accuse the agent of practically irrationality.

In other words, Hume is skeptical, not just about non-instrumental practical rationality, but also about instrumental varieties of rationality or irrationality. Hume accepts that we very often do what promotes our ends. But in his view this is a psychological principle rather than a principle of rationality: it is a psychological regularity. This is shown by the fact that, if we fail to do what is required to achieve our ends or even do something that is incompatible with their achievement, we do not act against the dictates of rationality. If this is so, the instrumental principle is not in fact a principle of reason, and questions of rationality do no enter the equation.

It follows that Hume does not espouse all claims of instrumentalism. Recall the three points in Hampton's list. Hume certainly accepts the second point, the idea that we reason instrumentally: we make the kinds of inferences necessary to adjust means to ends. To do so, the ability to calculate causal relations is required. Hume also agrees with the third point that our underived ends are not fixed by reason operating in a non-instrumental way. But, as we can see from the passage, Hume does not accept the first point, the idea that our actions or passions are rational or irrational to the extent that they promote the attainment of our ends. While he emphasizes, with the contemporary instrumentalist, that it is improper to call a passion or action irrational on non-instrumental grounds, he disagrees about the positive point. He also thinks it improper to call a passion or action irrational even on instrumental grounds. An agent who fails to take the means to an acknowledged end is just someone who prefers to act on his present occurrent desires rather than to pursue other ends he already accepts. To Hume, there is nothing irrational about failing to follow one's ends, although it may be a bit odd. The refusal to conform to instrumental norms is just an expression of the agent's preferences at the time. Hume, then, is not an instrumentalist in our sense as he doesn't accept all the points in Hampton's list. Accordingly, insofar as Humeanism involves commitment to the instrumental conception of practical reason, Hume's own view does not qualify as Humean.

To summarize, we have considered three positions, listed in order of the degree to which they countenance a specifically practical sort of rationality:

1. Anti-Humeanism holds that, in addition to the possibility that an agent may be irrational by instrumental standards, there is a substantial sense in which an action or passion can be seen as irra-

tional, a sense that does not require the acceptance of further ends which are related to the action instrumentally. Actions can be criticized, or recommended, without reference to any prior motivational states of the agent. As a concomitant, Anti-Humeanism assumes that an agent may have reasons for action which are not the direct or indirect consequence of the agent's having underived desires.

2. Humeanism rejects the notion of practical rationality in its substantial, non-instrumental sense but accepts the idea that reason can be practical insofar as it directs the agent to take the means to an avowed end. As a corollary, it holds that the source of our reasons is our desires. Humeanism is skeptical about practical reason, but its skepticism is limited; it salvages instrumental reason as a genuine form of practical reason.

3. Radical practical skepticism is the view, endorsed by Hume himself, that, properly speaking, there is no *practical* rationality whatsoever. In particular, the instrumental principle does not have rational force. There are means-end relations between states of affairs, but it is not incorrect, from the standpoint of reason, for the agent to ignore them and to remain unmoved. Human beings are so constituted as to reason instrumentally, but they are not under any sort of rational pressure to conform to the instrumental principle.

Does radical practical skepticism accept that our desires give us reasons? This seems to depend on what we mean by this. It cannot mean that our desires fix what it is rational for us to do; for that can hardly be true without instrumental relations getting any grip. This question will be treated in more detail later. Our main goal for now is to consider the plausibility of Humeanism in the contemporary sense — in the sense, that is, in which Hume is *not* a Humean. But it will be useful briefly to review the reasons why practical philosophers today often prefer Humeanism, with its commitment to instrumentalist rationality, over Hume's own radical practical skepticism. The issue is that radical skepticism denies instrumental considerations any rational or normative force. It is intuitively plausible that there is such a thing as means-end irrationality. An agent with a given end who stubbornly refuses to take the required means seems to be making some sort of mistake. It seems a genuine failing to violate the dictates of instrumental rationality.[10]

[10]See §6.

Of course, the existence of instrumental rationality is just what Hume, as we interpret him, denies. But that such rationality has a claim on us is built into our conception of agency. Someone who does not display a certain degree of conformity to instrumental principles is a creature we find it difficult to construe as an agent in the proper sense of the term. An agent, as opposed to an automaton, is responsible, and can be made accountable, for failing to be prudent. A creature that, as Hume envisages it, exhibits a brute preference to an option with total disregard to means-end considerations is unrecognizable as an agent.

The goal of the following discussion is to cast doubt on the tenability of skepticism about practical reasoning. In this context, it seems clear that of the two versions, the contemporary version is the more defensible position of the two views. For these reasons, in what follows I will focus on the modern, more defensible view.

2.2 Why think that reasons are based on desires?

There can be no doubt that Humeanism is widely held among practical philosophers — one writer has called Humeanism the "default theory" in the field.[11] Let us focus for now on the desire-based reason thesis. To see what makes the view so compelling to many action theorists, we will review a number of considerations that make the view that desires enjoy the special privilege of providing us with reasons an attractive position.[12]

1. The belief-desire model: The belief-desire model is a widely shared and plausible view of intentional action. It may be thought that the dependence of reasons on desires directly follows from this model of action, which posits that, whenever an agent does something intentionally, we can distinguish two relevant elements in his mental state. On the one hand, action involves a conative element, something that pushes the agent to action. This element aims at changing the world. On the other hand, the cognitive element, rather than purporting to change the world, aims at representing the world as it is. Neither element is dispensable. In order to act, it is not enough to have an opinion what the world should be like because one also needs an opinion as to what would be a way of effecting this change. Conversely, an opinion about the way to effectively

[11]Nozick writes that "instrumental rationality is the default theory, the theory that all discussants of rationality can take for granted, whatever else they think. There is something more, I think. The instrumental theory of rationality does not seem to stand in need of justification, whereas every other theory does" (Nozick 1993: 133).

[12]Cf. Darwall (1985: ch. 2).

bring about a goal alone does not dictate an action unless it is accompanied by a desire-like state that fixes the goal.

The fact that intentional action is inevitably accompanied by an intentional state of the conative sort may be thought to imply that only desires have the power to give reason, but this is to overlook a distinction. Thus Thomas Nagel, while admitting the possibility that some desire have the reason-giving force claimed by Humeanism, argues that not all desires are *motivating* desires.[13] On his view, a desire may itself be the upshot of a process of deliberation. These *motivated* desires, as he calls them, are themselves the result of a process of practical reasoning, which in turn may or may not involve further desires. A motivated desire, then, is a state that, for all that has been said, may be held as the result of having a number of pure beliefs. With this distinction in mind, we can see that the Humean claim that reasons are based on desires may be interpreted in two ways. A merely causal interpretation of the claim implies that, if a piece of behavior is to count as an intentional action, desires must be among its efficient causes. This interpretation leaves open the possibility that whatever desire is causally responsible for the conduct is merely a motivated desire, a mere causal byproduct of the psychological process. Thus it may be that the agent judges, on purely rational grounds, that he ought to do φ, and that as a final step in the process, as a matter of causal consequence, this judgments brings about the desire to do φ. In such a scenario, it would still be true, nominally, that desires play a role in the production of action. But this interpretation is so weak that it can hardly be the intent of the Humean contention that desires provide us with reasons.

Clearly the Humean wants his claim to amount to more than the modest idea that a desire must be part of the psychological mechanism leading to action. This modest idea is compatible with almost any theory of action, to the point of being almost vacuous. Instead, Humeanism is committed to assigning desires a constitutive role in acting for a reason, so that they are in a substantive sense the source of our reasons. For the Humean, far from being a mere causal antecedent of action, desires are what makes it true that the agent has a reason to act. True, the Humean typically accords desires an important causal role as well, but his view should go beyond causal significance. We have seen that Humeans are committed both to the idea that an action is rational insofar as it promotes the attainment of one of the agent's ends and to the idea that the agent's ends are constituted by desires. It follows that whether or not an action is reasonable is a function of the agent's desires. The Humean idea must be that

[13] Cf. Nagel (1970: 29). See also Wallace (2006a: 22ff).

intentional action ultimately requires an unmotivated desire that "simply comes to us" (Nagel 1970: 29) which is nonetheless endowed with rational or normative significance, in addition to causal significance.

The notion of motivated desires shows that the mere acceptance in principle of the belief-desire model does not imply commitment to Humeanism. There is something clearly true about the thought that it is impossible to act for a reason without having both a representation of what one's action will accomplish and a desideratum that one's intervention in the world is intended to accomplish. We can appreciate this thought without committing ourselves to the ambitious Humean idea that a desire, conceived as an element of our natural psychology beyond the ken of our rational faculties, is what makes it the case that we have a reason to act.[14]

2. Naturalism: As we have seen in the first chapter, there are naturalistic worries about alleged reason-facts. In particular, the inherently action-guiding nature of reasons is seen as puzzling on the assumption of a predominantly scientific world-view. The idea that reasons are grounded in desires is often taken to be a way to solve this puzzle. Reasons, as one writer puts it, can be reduced to desires.[15] Unlike reason-facts, desires, as psychological states, are deemed unobjectionable from the naturalistic point of view. An analysis of reasons in terms of desires helps us find a location of reasons in a naturalistic picture of the world.

This thought relies on the assumption that desires are elements in our natural psychology, i.e. states that we can describe objectively. In the final analysis, it is assumed, desires are states that are "just there" for us to discover them. The thought is that we can build a superstructure of normative properties and relations on top of the bedrock of naturalistically

[14]Likewise there need be nothing problematic about insisting on a cognitive and a conative element of action so long as one does not take this idea of a dichotomy of cognitive and conative states to have far-reaching implications. In my view, the truth behind the belief-desire model lies in the distinction of two complementary directions of fit. However, it is problematic to take it for granted that the element with mind-to-world direction of fit is *not* cognitive in the sense that it is placed firmly outside the scope of reason. It is important to see that these two ideas are separable, although they correspond to two common uses of the word "cognitive". On the one hand, the word applies to states that share the mind-to-world direction of fit of belief. In this sense, it is clear that intentional action involves some element that is not cognitive. But the word also refers to states that are not amenable to rational explanations, explanations that ask for a justification. Here I dispute the assumption that intentional action needs to involve, in a constitutive fashion, a non-cognitive or conative element in this second sense. This, in fact, is what the Humean asserts and the anti-Humean rejects. The point is simply that the seemingly straightforward plausibility of belief-desire models of action in no way entails the desire-based reason thesis. The question of how to accommodate the distinction of two directions of fit will be addressed in §3.3.

[15]Cf. Schroeder (2007: ch. 4).

describable desires. In this way, the Humean hopes to construct normativity from the ground up without risking to make it appear mysterious.

Part of the appeal of Humeanism, then, lies in the fact that allegedly objectionable normative entities can be reduced to mere psychological states. Having already discussed this thought, we can leave it to one side.[16] However, another naturalistic consideration that is thought to support Humeanism is that it is hoped to account for the biological continuity between human action and the behavior of non-human animals. If we assume, as the Humean does, that reasons are provided by desires, then as long as we have a conception of desire that is applicable to non-rational animals, we are free to say that animals, too, act on reasons, albeit in a less sophisticated way. Sub-rational creatures are attracted to, or repelled from, objects in their environment, which raises the question why we should not say that animals have reasons in a pre-reflective way. On the Humean model, human agency and animal behavior may be understood along similar lines.

As tempting as it is, I think we should resist this line of reasoning. In what follows, we will develop a conception of acting for a reason that implies that, at least so far as we know, only human beings, as the only known concept-mongerers, are agents in the full sense of the term. It may be objected that this response amounts to a form of chauvinism. We will take up this objection later.[17] Whatever else may be true, any support that the two naturalistic considerations lend to Humeanism is conditional on the acceptance of far-reaching naturalistic assumptions, ontological, as in the idea that the normativity of reason may be reduced, or methodological, as in the assumption that acting for a reason is acting on a desire for humans beings and non-human animals alike.

3. Internalism: According to existence internalism, there is a very general relation between desires and (judgments about) reasons. A reason is essentially a consideration on which it is possible to act, and it is something that we must be able to appeal to in an intentional explanation of the action. Desire is an essentially motivational kind of state. Accordingly, a necessary condition for ascribing a reason to an agent may be thought to be that the agent has a desire of some sort or other appropriately related to the action. This idea, however, is predicated on the unwarranted assumption that only desires can motivate us. An anti-Humean may argue that a purely cognitive state could by itself move an agent to action. Rather than providing independent motivation for the Humean thesis, this idea

[16] See §1.4.

[17] See in particular §4.6.

presupposes it. Here it may be objected that beliefs or belief-like states like observations are by themselves unable to motivate us. Once again, however, even if this is true, this doesn't rule out the possibility that a belief — say, an ought-judgment — is by itself able to bring about or justify a motivated desire. In this sense, the desires would only be causally necessary intermediaries; and the line of thought would not support the claim that desires are the source of our reasons. The idea that Humeanism is the superior explanation of the genuinely motivational character of reasons is speculation. The true test of this argument is to see whether a competing, anti-Humean theory can explain the idea that reasons are essentially things that motivate equally well. Showing this requires demonstrating how an anti-Humean theory can account for our internalist intuitions.[18]

4. Raw preferential desires: A simple line of thought may lend plausibility to the desire-based reasons thesis. Suppose two customers in an ice-cream parlor are alike in all respects save for the fact that one agent, *A*, has the desire to eat strawberry ice-cream, whereas the other, *B*, has no such desire. It seems plausible to say, in this situation, that the one agent, but not the other, has a reason to buy a cone of strawberry ice-cream. But if the only difference between the two agents is that *A* but not *B* has a preference or desire for this flavor, it may be held that is must be the desire that engenders the reason. *A* only has the reason because of his desire, and nothing else could explain the fact that *A* has a reason where *B* doesn't.[19]

Having pointed to examples of this type, the Humean may go on to argue that desires *in general* give us reasons in this way. What holds in the ice-cream parlor also holds elsewhere. When we ask for the most basic explanation of the agent's action — and thus of what his reason was — we inevitably reach a point where the only thing we can point to is a state of preference or desire. Furthermore, the Humean may point out that, having found that a desire grounds a reason in a special case, our best theoretic option is to assume that the same form of explanation holds in the general case. Explanatory simplicity is a reason for us to think that the general case is similar to the special case.

This argument, however, is not convincing. First, the idea that minimal pairs of agents, only differing with respect to their raw wants, have different reasons relies on special features of the situation. Even if we grant that desires may have this role in the ice-cream parlor, typical action does not occur in a context of strawberry-or-chocolate choices. It is doubtful,

[18]The valid intuitions behind both existence-internalism and judgment-internalism are discussed in chapter 5.

[19]Schroeder (2007: ch. 1) offers an argument of this type.

to say the least, that we can transfer the conclusion in artificial brute preference-based choice situations to the general case. Second, there are alternative explanations of what it is that in such situations gives us a reason. What relevantly distinguishes *A* from *B* may not be his desire to choose a flavor but the fact that *A*, but not *B*, can expect to derive pleasure from tasting the strawberry flavor. If so, the agent's reason does not derive from his desire *per se* but from facts about the kinds of things that tend to give him pleasure.[20] The reply to the argument would be that, if the agent with the desire for *F* has a reason whereas the one without the desire doesn't, this is so because desires often track what would be pleasurable for the agent. In fact, if the agent didn't derive any pleasure from the experience and if no other needs were fulfilled, the desire would no longer provide any reason. Finally, even if no alternative account of *A*'s and *B*'s reasons were in the offing, the argument would not by itself be conclusive. True, other things being equal, a theory that explains the difference in multiple-choice situations would be preferable, but given the artificiality of these situations, explanatory simplicity falls into the "nice to have" category. Certainly we should not ground our entire theory of reasons on differences in situations of this type.

2.3 The phenomenological conception of desire

Having seen that some common motivations for the desire-based reasons thesis are not compelling, we can now turn to an assessment of the thesis and put it to the test by drawing out its implications. Let us start by noticing that the thesis implies the biconditional

> You have a reason to do φ if, and only if, you desire to do φ.

Splitting this claim into its two components, we can see that Humeanism is committed to two ideas: (i) wanting to φ is necessary for having a reason to φ and (ii) wanting to φ is sufficient for having a reason to φ. Consequently, we can criticize Humeanism either by raising difficulties about the necessity of desires for reasons or by questioning the sufficiency of desires for reasons. The latter strategy is pursued in the rest of this chapter. However, before turning to this task, I will explain briefly why I will ignore the former strategy. Many anti-Humean philosophers have taken the route of arguing that it is possible for an agent to have a reason to do φ without having a desire to do so. It is common to point to a situation

[20] Cf. Scanlon (1998: ch. 1).

where an agent has a reason to perform an action while at the same time having no desire at all to do so. It is easy to imagine a situation where the agent has a reason to do something although he finds it altogether unpleasant or tedious to do so. Examples are amply provided by cases of moral duty clashing with personal pleasure. In scenarios of this kind, it is pointed out, the difficulty for the agent lies precisely in the fact that he is morally obligated to do something even though the prospect is entirely undesirable for him.

I do not want to say that this anti-Humean complaint is without merit. However, any attempt to attack the Humean position by citing moral counterexamples is laden with difficulties.[21] In particular, any such attack relies on a particular understanding of moral reasons as imposing categorical demands on agents. The cogency of an argument against Humeanism is conditional on the plausibility of the underlying theory of moral reasons, and any such interpretation of morality is inevitably controversial. The goal here, by contrast, is to understand reasons for action without regard to what makes moral reasons special. The hope is that skirting the contentious topic of morality makes the case against Humeanism more compelling.

The following criticism of the desire-based reason thesis, then, will focus on the sufficiency claim:

Sufficiency Having a desire to φ is sufficient for having a *prima facie* reason to φ.

Three brief notes are in order. First, *Sufficiency* is clearly a simplification. A relevant, simple distinction needs to be drawn between two ways in which an action may result from a desire. (a) On the one hand, if I want to dance in the rain, I can satisfy this desire simply and directly by dancing in the rain, provided that it is raining. (b) On the other hand, if I want to make my grandmother happy, I cannot do so directly. When I make my grandmother happy by baking her a cake, I satisfy the desire by performing another action which I think is likely to bring about the desired state of affairs. Clearly many if not most of my desires — being comfortably warm, having friends, and so on — are of the second type.

Strictly speaking, the sufficiency claim should take this into account. We could modify the formulation:

[21] To name one such controversy, Foot (1972) is a classical defense of the claim that moral reasons are only hypothetically valid. McDowell (1998a) takes the opposing view that the force of moral reason is not conditional on the existence of desires that function as independent components in the explanation.

> If *S* has a desire that *p* and the desire *p* can be satisfied by his φ'ing, then he has a *prima facie* reason to φ.

On this formulation, my desire to make my grandmother happy gives me a reason to bake her a cake. Moreover, my desire that I dance in the rain gives me a reason to dance in the rain. While in the first case, I will only perform the action if I also have the instrumental belief that baking a cake will make my grandmother happy, no means-end belief is strictly required in the second case. Nonetheless, there can be no harm in assuming that in dancing for the sake of dancing, I do so because of a trivial means-end belief that I can dance by dancing. On the assumption that we have such "null" means-end beliefs, the amended formulation covers (a) directly actionable desires as well as (b) the more common indirect variety. Keeping this complication in mind, however, we can, for the sake of simplicity, go back to the original formulation even if it doesn't cover all cases. Surely at least for desires that can be fulfilled directly, *Sufficiency* must be valid if Humeanism is true. If we can find cases that disprove *Sufficiency* in the simple form, this will cast doubt on the entire Humean claim that desires provide us with reasons.

Second, we are not claiming that Humeanism is committed to the idea that having a desire gives you a conclusive reason to satisfy it. Of course it cannot be true generally that if you want to do φ, you ought to do so. Even if I want to stay dry, when walking along the river I see a man drowning, surely I ought to ignore my desire and dive in. *Sufficiency* is not the implausible claim that you ought to follow each of your desires all of the time. True, the desire to stay dry gives me *a* reason to keep out of the river, but I have a much stronger reason to save the man's life, perhaps based on some other desire to help people in need. Desires do not provide the agent with an all-in reason. Instead, desire-based reasons are *prima facie* and typically defeasible by other considerations. Even for Humeans, it is only when no competing or more powerful *prima facie* reasons is present that your desire is what you ought to act on.

Third, it may not be obvious that Humeanism is committed even to this modest *prima facie* sufficiency claim. Here we may simply ask how else the Humean would want to see his idea interpreted if not as implying that someone who has a desire thereby has a defeasible reason to act to satisfy the desire. If claims about desires being a source of reasons are to have any substance at all, it must be an essential part of the idea that reasons flow from desires. Perhaps it will be said that wanting to φ is only necessary for having a reason to do so but doesn't *by itself* give us a reason. This could mean that, although desires generally generate

reasons, they only do so under certain circumstances, thus maintaining the connection while introducing certain qualifications. Later on we will consider candidates for such additional conditions.[22] On the other hand, if the Humeans rejects the sufficiency connection entirely, the result is the idea that desires have only a secondary, causal significance, which is to give up a central component of Humeanism. We can conclude that Humeanism is committed to the claim that in the absence of special circumstances desires are intrinsically reason-giving.

At this point, however, the question before us — whether desires generate *prima facie* reasons — has become difficult to answer without knowing in greater detail what, according to the Humean, desires consist of. In the abstract, it doesn't seem unplausible that you should pursue what you desire, but what this means exactly — and thus its truth — hinges essentially on the way we interpret the talk of desire in *Sufficiency*. A desire, to be sure, is a psychological state of the agent, but what is it about desires, as opposed to other mental states, that endows them with a power not only to explain but to *rationalize* our conduct?[23]

Humeans have typically defended either of two conceptions of desires:

1. According to the *phenomenological* conception of desire, an agent who wants to do something experiences a certain qualitative feeling towards a certain object or state of affairs.

2. The *dispositional* conception of desire maintains that to desire p is to have a disposition, or cluster of dispositions, concerning various sorts of behavior toward p.

Beginning with the phenomenological view, we will now examine whether these conceptions of desire lend support to the Humean sufficiency claim. A well-known version of a phenomenological conception of desires can be found in Hume's *Treatise*. The stated purpose of this work is to develop what Hume calls a new science of man — a comprehensive theory of the human mind. Among mental episodes or "perceptions", he distinguishes impressions and ideas, which he characterizes, respectively, as a kind of feeling and thinking. Passions are a kind of impression of the mind, and, more specifically, impressions of reflection which arise as a consequence of original impressions or ideas. Passions and other impressions are distinguished by the ways they feel. In particular, Hume holds that ideas and

[22] See §§5–6 below.

[23] Following Davidson (2001a), I intend my use of the verb "rationalize" to carry no pejorative overtones.

impressions differ chiefly as to their "force and liveliness" (Hume 1978: 1.1.1, 1). The characteristic feature of passions, Hume assumes, is something we are intimately acquainted with. Hume sees little need to explain the kind of phenomenal feel that defines passions.

Still, we should be skeptical about the existence of a single distinctive quality that accompanies all of our passions and, hence, all of our intentional actions. For Hume, desire and aversion are direct passion which arise as a consequence of impressions of pleasure and pain.[24] For example, an agent who experiences pain after holding his hand into a flame will develop a tendency to experience a feeling of aversion when faced with the prospect of direct contact with fire. The hypothesis that the aversion to fire is identified with a certain feeling has some plausibility, as it is not hard to imagine the peculiar "force and liveliness" of the impressions.

Perhaps it will be said that what gives desires, or passions, the power of giving us reasons consists just in the forcefulness of this peculiar experience. It should be clear that the various kinds of passion mentioned by Hume such as fear, hope, pride and humility each have a very different phenomenological profile. What do these experiences share that gives them their extraordinary power? It seems that Hume has no easy answer to this question. To make things worse, it seems that many desires are almost entirely dispassionate. Mundane wants — to solve the crossword puzzle in the morning or to save money for one's retirement — are rarely accompanied by strong feeling, or any feeling at all. *Pace* Hume, often we would be hard pressed to pinpoint which feeling accompanied a given action or to ascertain that any such characteristic feeling exists.

Hume's idea of a passion as what grounds reasons in general is modeled on actions we perform in response to salient experiences, like the one of having a sudden craving for chocolate. True, such passions have a strong phenomenal component, but acting for a reason isn't always, or even very frequently, like this. This difficulty may be what prompts Hume to admit, somewhat paradoxically, calm passions, whose qualitative profile is mild, in addition to violent passions, such as hunger or thirst.[25] The difficulty remains, however, that Hume's explanation of passions in terms of force and liveliness does not explain what distinguishes them from other mental episodes such as beliefs or sensory impressions. The existence of less than poignant passions shows that the urging feel of physical desire cannot be the defining characteristic of passions in general. In short, Hume's theory doesn't explain what it is about passions that binds them

[24] Hume (1978: 2.3.9).

[25] Hume (1978: 2.3.3, 417).

together and gives them the normative *oomph* required for them to have normative significance.

Some recent writers have revived the idea of a distinctive phenomenology of pro-attitudes.[26] Some writers base their account on our ability to recognize what they call a "obligatoriness or"demand quality"" in objects and actions (Firth 1952: 327). Sharon Street mentions our

> knowledge of *what it is like* to have a certain unreflective experience — in particular, the experience of various things in the world "counting in favor of" or "calling for" or "demanding" certain responses on our part. (Street 2008: 240)

She goes on to explain:

> I believe it is impossible adequately to characterize this experience except in such primitive evaluative terms, yet I think we all know exactly the type of experience I am pointing to. We need only think of how we feel when, for example, a tractor trailer swerves toward us on the highway or we see a stranger threatening our child; we all know what it is like to experience (at an unreflective level that we surely share with many other animals) evasive action or a protective response as utterly "demanded" or "called for" by the circumstances. [...] one must already be familiar with the conscious experience I am talking about (Street 2008: 240)

Accordingly, perhaps we should understand desires as attitudes characterize by what could be called an experience of intrinsic normativity: a general feeling that an action is called for. The idea, then, would be that a normative experience that φ'ing is called for is sufficient for giving an agent a *prima facie* reason to do φ.

Even in its updated version, the phenomenological conception seems unable to sustain the Humean claim. One problem with the proposal is that its fundamental premise is doubtful. It is not at all clear that there is any characteristic experience unifying all desires or desire-like states. In particular, when agents do tedious things out of obligations, they don't seem to have any salient normative feeling. In any event, if a calm passion or weak normative experience is present, it is hard to detect. In particular, the idea that a particular intrinsic quality could be associated with

[26]See also Galen Strawson's arguments against neo-behaviorism in his Strawson (1994: ch. 9).

normativity is questionable. This is not to dispute that desires have a normative aspect — on the contrary, this is precisely the thesis that will be defended.[27] It may even be true that normative states sometimes are accompanied with a certain phenomenological content. But it is doubtful, to say the least, that normative states have this content essentially.

More to the point, we should ask whether making the "demand quality" the identifying characteristic of a desire helps explain how desires, as the Humean insists, are reason-giving. In this connection, it is useful to consider an argument against the phenomenal conception put forward by Michael Smith.[28] Although I do not agree entirely with Smith's specific conclusion, I think it puts us onto the right track. Smith's argument is based on the epistemology of desires and specifically on the idea that we do not always know what we want. The phenomenological conception assimilates desires to sensations as it defines the former, like the latter, in terms of their distinctive phenomenal quality. Accordingly, it makes it seem as if our knowledge of our own desires is similar to that of our own sensations. Now our self-knowledge of pains and similar sensations is all but infallible: typically, when you think that you have a headache you really do. From the current suggestion, it seems to follow that we typically know, in the same way, about our own desires. However, Smith insists, this is a mistake because we can easily be deceived about our own motivations. Sometimes an agent may have a desire and act on it and yet not be prepared to admit to having the desire because, as it is seen as shameful, it has become subconscious. Even after acting, he may remain unaware that this desire was his motivation. On the other hand, an agent may assume that he has a motivation to do something, perhaps because it is expected of him, even though on closer inspection he does not really exhibit the desire in question in his conduct. We may, then, be deceived both about the desires we have and about those we don't.

If a phenomenological conception applied to desires, we would expect it to be easy to know what one wants. But while we have near-infallible knowledge of our sensations, we are not in such a privileged positions with respect to our desires. The phenomenological view fails to explain that knowledge of what we want is sometimes hard work. By contrast, Smith's preferred view, the dispositional conception, explains how our self-knowledge of desires can sometimes be imperfect. As that view identifies desires with complex dispositional states, knowing one's desires requires knowing what one would do, say or think in certain counterfactual circumstances. Clearly, one may be mistaken about one's own hypothet-

[27] See chapter 3.

[28] Smith (1994: ch. 4).

ical behavior. For Smith, an adequate theory of desire needs plausibility as a self-standing account of desires, which rules out the identification of desires with their qualitative content because this view gets the epistemology of desires wrong. Smith concludes that we should reject the phenomenological view in favor of dispositionalism.

What is responsible for the superior ability of this alternative view to account for conative self-deception, according to Smith, is the fact that it emphasizes the propositional content, rather than the phenomenological content of the states. The epistemology of the propositional content of a desire is different from that of its phenomenological content. It is easy to be wrong about one's own propositional states. For Smith, then, dispositionalism is the key to giving an explanation of propositional content. He writes that

> [phenomenological conceptions] in no way contribute to our understanding of what a desire as a state with propositional content is, for they cannot explain how it is that desires have propositional content. They therefore in no way explain the epistemology of the propositional content of desire. (Smith 1994: 108)

Because the phenomenological conceptions don't help explain the propositional aspect of desires, he adds:

> they require supplementation by some independent and self-standing account of what a desire is that explains how it is that desires have propositional content and how it is that we have fallible knowledge of what it is that we desire (Smith 1994: 108)

According to Smith there are two distinct aspects of desires to account for: their phenomenological content and their propositional content. He admits that a desire sometimes may have a qualitative feel attached to it. But that is not always the case, and even when it is, it is the state's propositional content that is primary. If the phenomenological view cannot provide an explanation of latter — it models desires on sensations, which do not *per se* have propositional content — it needs a supplementary account of the intentional aspect of desire. This further aspect is required to make room for Freudian counter-examples, meaning that for explanatory purposes, the phenomenal content is secondary. Smith concludes that desires cannot be adequately understood on the phenomenological conception.

Turning now to the evaluation of Smith's argument, we can agree with his conclusion — that the propositional content of desires is primary and any possible qualitative content secondary, and that this is an advantage of the dispositional theory over its rival — while remaining skeptical about his way of arriving at the conclusion. The problem with his argument is twofold. First, as regards self-knowledge, the case for dispositionalism is not as clear-cut as Smith seems to see it. Smith argues that to understand desires as a kind of feeling is problematic because it implies, against our everyday experience, that we mostly know exactly what we desire. Whether you actually have a disposition of this kind implies knowledge about what you would do in various situations, and it is easy to be deceived about such counterfactuals. But in the same point also lies a weakness, for this account is in danger of making it *too difficult* to know what you desire. If self-knowledge implies complex knowledge of how you would react in hypothetical situations, it may be that you hardly ever know for sure that you have a desire. If this is true, then the dispositional account overshoots its target. To the extent that it explains well how we are sometimes mistaken about our desires, it fails to explain how we are still right in the majority of cases.

Second, Smith suggests a two-stage process. As a strategy to explain the role of desires for the Humeans thesis, we ought to start by looking for an explanation of the mental state of desire which is plausible as a self-standing account. On this suggestion, we consider the connection between desires and reasons only after we have established an independently plausible account of desires. It sees the first step as independent of the second insofar as we can discover the nature of a desire in an isolated way, without taking into consideration the way in which the reason-providing force may be essential to desires. However, desire is *essentially* the type of intentional state that can figure in intentional explanation of an action. On a Humean view, desire is the kind of state that can rationalize behavior. But this casts doubt on the idea that we could understand the notion of desire without regard to its role in practical reasoning and in intentional explanation. If a conception of desire is to sustain the desire-based reason thesis, a two-stage process is unlikely to succeed.

Now in order for something to be able to figure in intentional explanations — the type of explanation that make sense of the agent's behavior by revealing the light in which the agent saw it — it must at least meet the minimal requirement of possessing propositional structure. Even if a desire may also have qualitative content, it is its propositional content that determines what actions the desire counts in favor of, or which ac-

tion it can appealed to in an intentional explanation. Counting in favor is best understood as a matter of something's being an element in a potential piece of practical reasoning, and such elements must have a from corresponding to that of judgeable, i.e. propositional contents. We cannot divorce desires from its *rational* role, which is determined only by its propositional content. Whatever phenomenal content these states may have does not enter into the intentional explanations.

Although Smith's appeal to self-knowledge is problematic, then, our argument leads us to a similar conclusion: desires should be thought of as primarily having propositional content. This doesn't need to imply that desires are devoid of qualitative aspects. While a desire may or may not have qualitative content, it is the propositional content that fixes its identity. Rather than attempting to understand desires in isolation from their potential role in rational explanations, we see this role as partly fixing their content. As it gives pride of place to an intentional state's propositional content, a dispositional theory is clearly to be preferred over a phenomenological theory.[29]

2.4 The dispositional theory of desire

Although it is clear that the dispositional theory offers advantages over the phenomenological conception of desire, to see if it gives support to Humeanism, we need to ask what kinds of dispositions it identifies desires

[29]It might be said that the last paragraphs amount to an unfair criticism of Smith's position. Smith carefully distinguishes between theories of motivating reasons and theories of normative reasons, and while he defends a Humean account of the former, he takes pains to stress that he is defending an anti-Humean theory of the latter (Smith 1994: ch. 4–5). His theory of normative reasons is that, roughly, an agent has a reason to do φ only if he would desire φ if certain ideal circumstances were present. As a consequence, although normative reasons might be said to be desire-dependent, they are dependent on hypothetical rather than actual desires. Smith insists that this question is quite independent of the actual preexisting set of motivations that the agent happens to have. The question we are considering now — the sufficiency of desires for reasons for action — is a question about normative reasons, so the criticism doesn't seem to apply straightforwardly to Smith's own theory.

Smith, then, is not directly committed to the idea that desires furnish us with reasons in just the way we have ascribed it to Humeans. Perhaps, then, Smith need not take into account the essential involvement of desires in reasoning. He cannot make use of the idea that normative reasons must be propositionally structured. To the extent that this is true, it should be pointed out that the principal target is not Smith's own, rather complicated theory of normative reasons but the Humean theory of normative reasons. In doing so, we are considering Smith's insightful discussion as one of the various concepts of desire available to Humeans, a discussion that is valuable regardless of whether Smith's own theory qualifies as Humean in our sense.

with.[30] In his *Analysis of Mind*, Bertrand Russell proposes such a theory according to which we should understand the mental state of wanting in terms of a series of events he calls a behavior cycle, a "series of actions continuing, unless interrupted, until some more or less definite state of affairs is realized" (Russell 1921: 75). He explains:

> The property of causing such a cycle of occurrences is called "discomfort"; the property of the mental occurrences in which the cycle ends is called "pleasure". [...] The cycle ends in a condition of quiescence, or of such action as tends only to preserve the status quo. The state of affairs in which this condition is achieved is called the "purpose" of the cycle, and the initial mental occurrence involving discomfort is called a "desire" for the state of affairs that brings quiescence. (Russell 1921: 75)

For instance, such a behavior-cycle may be the result of hunger. Describing a hungry agent, he writes:

> [I]t seems clear that what, with us, sets a behaviour-cycle in motion is some sensation of the sort which we call dis-

[30]Surprisingly, like the phenomenological conception, the dispositional view of desires can also be found, in an inchoate form, in Hume's *Treatise*. When Hume discusses the distinction between calm and violent passions, he notes that, unlike to latter, the former do not necessarily have a salient phenomenology. If I refuse to eat another brownie despite my strong urge to do so, we can ascribe this triumph to my long-term goal of maintaining my weight. On Hume's view, because the rewards of this desire are far off in the future, it does not make itself felt in the same way as an occurrent appetite. He writes that calm desires

> are more known by their effects than by their immediate sensation (Hume 1978: 2.3.3, 417)

and adds that, because of their tendency to "cause no disorder in the soul", we can easily confuse these passions with the operation of reason. In other words, we are justified in ascribing passions or desires to agents even when they don't report the appropriate feeling if they show, in their behavior, a tendency that is best explained by the presence of the desire. The fact that on the whole Hume clearly favors a phenomenological conception over the dispositional account suggested by the passage cited is no doubt a consequence of his general theory of mind or "science of man". On this theory, ideas are copies of sensory impressions; by virtue of their pictorial character, they represent the world. Passions, on the other hand, are what Hume calls original existences; they do not have representational character. For Hume, this theory, which he announces at the beginning of book I of the *Treatise* as the basis of his new system of philosophy, raises the problem of explaining the differences between various non-representational states. His official view on this topic is that these differences are all explained by the various degrees of liveliness. Adding that some types of mental states can also have their contents fixed by their dispositional connections to behavior would imply a fairly important change of direction.

> agreeable. Take the case of hunger: we have first an uncomfortable feeling inside, producing a disinclination to sit still, a sensitiveness to savoury smells, and an attraction towards any food that there may be in our neighbourhood. At any moment during the process we may become aware that we are hungry, in the sense of saying to ourselves, "I am hungry"; but we may have been acting with reference to food for some time before the moment. (Russell 1921: 67)

The desire manifests itself in characteristic "searching" behavior, which continues until the purpose of the activity is realized. However, the purpose is not fixed by what the agent's thinks but rather in terms of his activity of trying to get the object. In other words, Russell offers a behavioristic analysis of the mental state of desire in terms of its typical behavior. He writes that desire "must be a causal law of our actions, not something actually existing in our minds" (Russell 1921: 60). The purpose of a behavior cycle, and consequently the object of the desire, is, not what the agent consciously intends, but the state of affairs that, if realized, would bring the cycle to an end. Thus the agent's wanting food rather than some other object is constituted by the fact that his searching behavior would come to an end if he found food. What desires one has is a matter of what one would do or not do in certain counterfactual conditions. To desire that p is the state which would go out of existence if the state of affairs that p were realized.

Russell's view offers two advantages. By defining a desire in terms of a "causal law", i.e. links to possible behavior, it highlights the propositional aspect of desire rather than its qualitative aspect, which Russell still admits in the form of disagreeable sensations. To define the propositional content of a desire for food, Russell could say that X has a desire to have food just in case he is engaging in behavior which would stop if the state of affairs "X has obtained food" is realized. Desires, on this account, meet the requirement for being a candidate for a reason. Moreover, Russell takes into account the possibility of imperfect self-knowledge by admitting the possibility of unconscious desires as well as conscious desires. If the object of my desire is determined by what would bring quiescence to my searching behavior, I can easily be wrong about what I really want.[31]

[31] In fact, when he writes that "[c]onscious desire is made up partly of what is essential to desire, partly of beliefs as to what we want", Russell defines a conscious desire in terms of unconscious desire, i.e. simply as being a part of the behavior-cycle (Russell 1921: p 67). See (Kenny 1994: ch. 5) for an exposition and criticism of Russell's theory.

The question before us now is whether Russell's theory of desire makes it plausible that if I have a desire, I *ipso facto* have a *prima facie* reason to act accordingly. A negative answer is suggested by the following considerations.

To begin with, we should note that a technical mechanism, as well as a living organism, can exhibit behavior that it would not be inappropriate to describe as a Russellean behavior-cycle. The function of a pressure-plate system may be described as wanting to open the super-market door when a customer walks toward the entrance. The cycle starts with the system recognizing a certain weight. The opening of the doors brings quiescence to the cycle, albeit only temporarily, as when the door is closed again, another customer is likely to reactivate the mechanism. In short, the ability that Russell identifies with desire — what Brandom calls a reliable differential response disposition — is not restricted to potential agents.[32] However, a pressure-plate system can hardly be said to have reasons. This casts doubt on any Humean claim that all it takes to have a reason is to be in the kind of dispositional state Russell points to.

The defender of Humeanism, of course, has an easy comeback. The idea that wanting may be as simple as a reliable disposition, combined with the view that desires give us reasons, does not commit him to the absurd idea that all systems with behavioral dispositions resembling a behavior-cycle have reasons. On the other hand, we may suspect that there must be an issue with any Humean proposal that links having reasons to an ability shared by mere technical mechanisms. Surely the presuppositions of having a reason are more extensive than Russell's theory suggests.

Next, according to *Sufficiency*, to have a desire *ipso facto* is to have a reason. We may now ask what it is about being in a Russellean cycle that makes it the case that one has a reason. Suppose a smoker craves a cigarette. On Russell's analysis, this means that a state of discomfort initiates a cycle of behavior on the part of the agent: he looks in his jacket only to find an empty packet, he goes to the store and buys a new one, he lights the cigarette, and so on. He has a desire for a cigarette because his restlessness, and thus his searching activity, would stop if he smoked a cigarette. However, although the agent has a habit of seeking out cigarettes, that does not seem to make it true that he has a reason to do so. The description of the agent as looking for a smoke implies that he is likely to smoke a cigarette in the future, but not that he ought to do so.

Surely the answer to the question — what aspect of being in a behavior cycle rationalizes one's action — must be that nothing does. Merely being

[32]Brandom (1994: 7–8).

in a habit to do φ in no way justifies the doing of φ. One may value having a routine, of course. But if it is possible to appeal to one's habit in justifying one's action, this is only so because one espouses a normative principle to the effect that it is better to stick to one's habits. We clearly cannot presuppose a substantive normative judgment of conservatism in explaining the purported connection between desires and reasons. That I tend to smoke cigarettes does not make it sensible for me to do so.

Third, the complaint is not just that having a desire for a cigarette, in Russell's sense, does not give one a conclusive reason to do so. It is true that health hazards may outweigh individual reasons one may have for smoking. But merely being in a habit does not even give the smoker a *prima facie* reason to smoke. The disposition does not rationalize the action, even defeasibly.

This is true not just from the perspective of a third-person onlooker. This possibility exists from the first-person deliberative standpoint as well as from the external perspective. The agent may say, "It is true that I tend to pursue *X*, but even admitting this, do I really have a reason to do so?" It does not seem a sign of confusion to ask this. Instead, the possibility of such a question indicates that even for someone in the grip of a disposition, it is not evident that the disposition engenders a justification. This is what happens when one notices that one is in a rut. Such a person might say, "I know I always do this; but I have no justification for doing it."[33] The mere fact that one is in a rut doesn't by itself constitute a reason to continue the pattern at all, as all justification must result from additional considerations.

Although it sometimes thought that the idea that desires-as-dispositions give us reasons is intuitive, it is, to the contrary, not plausible at all that everyone who has a Russellean desire to φ *ipso facto* has a reason to φ. Hailing from 1921, of course, Russell's theory is hardly up-to-date. Perhaps it will be thought that the worries mentioned are peculiar to his theory and that a more contemporary dispositional will fare better. As we have seen, when Smith criticizes the phenomenological conception, he proposes his own dispositional account in which he identifies desires

[33]The problem brings to mind Moore's Open Question Argument (Moore 1959: ch. 1). In that argument, Moore argues that naturalistic definition of goodness inevitably fail. Take any natural property *F* that is proposed as an *analysans* for goodness. No matter what the property, we can grant that a proposed action *X* has property *F* while at the same time doubting, or denying, that *X* is also good. Crucially, we can doubt this without thereby betraying a confusion or misunderstanding of the terms "F" or "good". In other words, granting that *X* is *F*, the question whether *X* is also good always seems to remain an open question. For Moore, any attempt to identify goodness with a natural property is an instance of what he calls the Naturalistic Fallacy and should as such be rejected.

as states occupying a certain functional role. Desire is a "complex dispositional state":

> [W]e should think of desiring to φ as having a certain set of dispositions, the disposition to ψ in conditions C, the disposition to χ in conditions C′, and so on, where, in order for conditions C and C′ to obtain, the subject must have, *inter alia*, certain other desires, and also certain means-ends beliefs, beliefs concerning φ-ing by ψ-ing, φ-ing by χ-ing, and so on. (Smith 1994: 113)

On Smith's account, ascribing a desire for an end to an agent is to make a counterfactual claim about the agent, a claim about what the agent would do depending on his beliefs about what would be a way to achieve his end. He explains that, in this respect, desires differ fundamentally from beliefs:

> a belief that p tends to go out of existence in the presence of a perception with the content that not p, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that p. (Smith 1994: 115)

Smith improves upon Russell's view because he explicitly takes into account the relevance of means-end reasoning for the way in which a desire is realized behaviorally. What an agent who desires φ will do depends on what he thinks is necessary in order to achieve φ. Relatedly, Smith acknowledges the fact that the content of a mental states crucially depends on relations to other mental states. The content or meaning of a desire to φ depends on the belief that ψ'ing, or χ'ing, etc. is a way of φ'ing. What a desire to φ disposes an agent to do depends, importantly, on the agent's beliefs about what actions would be ways of achieving the desired result. A dispositional theory implies, correctly, that we cannot understand desires apart from these essential connections.

A further difference to Russell is that Smith doesn't think that desires are just "causal laws" or mere behavioral dispositions. Smith's proposal is based on a functionalist picture of the mind that holds that the identity of an intentional state is determined by its causal role, i.e. by the type of event that tends to cause the state, the type of event that is typically caused by the state and by the type of other mental states that both cause and are cause by the state. Furthermore, he holds that having a desire involves a disposition only *inter alia*, explicitly leaving open the possibility that a desire may also incorporate a phenomenological content, although

for Smith a state's phenomenological feel remains secondary to its propositional content.

Despite these improvements, however, Smith's functionalist view is in fundamentally the same position as Russell's behaviorist account with respect to the question whether the conception of desire supports the desire-based reasons claim. To see this, consider an example by Warren Quinn.[34] Quinn asks us to imagine a person with a peculiar fascination with car stereos — call him the Radio Man. This person, albeit normal otherwise (let us suppose), has the habit of turning on every stereo in every open parked car as he walks along the street. He just does it out of habit and does not at each stop pause to wonder whether or not he ought to turn on the radio at every instant. As Quinn puts it, Radio Man is in a "strange functional state": he is disposed to react to the sight of a parked car by turning its radio on.

As the case is described, the only salient aspect of Radio Man is his disposition to act. He doesn't expect to derive pleasure from this activity and he doesn't get any. Nor is his activity useful to any other further purpose. He does not turn on the radios *in order to* listen to music or to hear the news, or to do anything else. So in judging whether he has a reason we have only his functional state to go on. Quinn asks whether, given this disposition, the Radio Man has any reason to turn on radios. The answer he suggests is that strange functional states such as Radio Man's do not provide an agent any reason whatsoever to do anything. Such a disposition doesn't really seem relevant to the question what reasons the agent has.[35]

If it is true that, in spite of his functional state, Radio Man doesn't have any reason to turn on stereos, the thesis that desires conceived as dispositional states rationalize cannot be true. It is one thing to conform to a patter of behavior, but it is quite another thing for the pattern to make sense of the behavior. If the intuition elicited by Quinn's examples is that it would be wrong to speak of Radio Man's reasons, what features of the

[34]Cf. Quinn (1995). G.E.M. Anscombe discusses a similar case: "But is not anything wantable, or at least any perhaps attainable thing? It will be instructive to anyone who thinks this to approach someone and say: 'I want a saucer of mud' or 'I want a twig of mountain ash'. It is likely that the other will then perceive that a philosophical example is all that is in question, and will pursue the matter no further; but supposing that he did not realise this, and yet did not dismiss our man as a dull babbling loon, would he not try to find out in what aspect the object desired is desirable?" Her point is that, unless we can supply a desirability characteristic, we cannot properly understand the agent as *desiring* that object, despite his declaration (Anscombe 2000: §36, 70).

[35]For a similar conclusion, see Heuer (2004).

case explain this intuition? Radio Man's salient feature is that his functional state ultimately remains unintelligible or alien to us. It is rationally isolated in a way that makes it impossible to see the state as reason-giving. It also strikes us as arbitrary. Perhaps it will be said that on Smith's view, if a desire, as a functional state, forms part of an intricate network of connections between mental states, it is not arbitrary or isolated. However, although desires are functionally integrated, their integration is not of the right kind to trigger our intuition that they *count* as rationalizing. That someone has a desire is still a mere descriptive, psychological fact about the agent. It is not clear at all why the agent should care about a mere functional state as having normative force for him.

Smith's functionalist conception of desire is in no better position than Russell's behaviorist view to explain why a desire matters rationally. What the dispositionalist and phenomenological conception have in common is a deflationary view of what it is to want something. They all hold that you can desire p simply by showing certain dispositions, simple or complex, or by having a certain phenomenological experience. But if a desire that p is to justify, and provide reasons for, an action, it needs to be such as to make the action intelligible. The important difference is this. What separates a normal agent from Radio Man is that the former does, whereas the latter doesn't, see the action in a positive light. We are unsure why Radio Man should care about his dispositional attitude towards stereos precisely because he doesn't see anything good or desirable in the act of turning on the radios. The action strikes us as arbitrary because, even from the agent's own perspective, there is nothing good to be found in it.

So we can explain our intuition that, although Radio Man is in a strange functional state, he does not have any reason to conform to it by reference to an important distinction between two general perspectives on action:

Low Brow Intentional action is acting on a desire. In order to desire an object or state of affairs, it suffices to have a certain feeling or to be disposed to pursue the object or state of affairs, and nothing more. It need not entail any positive evaluative stance on the object or state of affairs.

High Brow Whenever we act, we necessarily aim at something good. Intentional action requires regarding an object or state of affairs as good or desirable. Intentional action requires positive evaluation of its goal.[36]

[36]The distinction between High Brow and Low Brow originates in Railton (1997). Railton's commentary on the difference will be discussed in §6.

The Humean fails to establish that desires rationalize because he sees action from a low-brow standpoint. To identify desires with dispositional states, however complex they may be, is to adopt a low-brow stance on agency. On a low-brow view, roughly, a state counts as a wanting of p when it is constituted by a functional state which involves pursuing p. It follows that Radio Man can follow his strange habit without seeing anything desirable about what he does. The consequence is that while we imagine the agent pursuing his strange functional state, that state cannot rationalize the behavior.

But we don't have to accept the consequence that the action remains mysterious. We can imagine Radio Man only if we of conceive intentional action in a Low Brow way which completely brackets its evaluative element. If we presuppose High Brow, we get a different picture. According to Hight Brow, an agent can be said to be intentionally doing φ as a response to a reason to φ only if he also takes a positive stance towards φ'ing, in a way that makes his action recognizable as aiming at some good or other. On this view, the action no longer appears arbitrary because the agent's evaluative judgment places it in a rational context. It is no longer isolated, rationally speaking: taking something to be good amounts to integrating it in a general view of the good. If the agent regards something as good, he doesn't have a choice but to care about it in some way: he cannot rationally remain indifferent about it. Moreover, the existence of a positive judgment on the agent's part enables us to provide an intentional explanation of the action which makes sense of it by the agent's own lights. Whereas a disposition only affords us with a causal explanation, we can now offer a *rational-agent* explanation of the behavior.

The consequence to draw from the Radio Man example is to drop Low Brow as an inadequate picture of rational agency. Russell's or Smith's dispositional theories are superior to a phenomenological account because they emphasize the propositional content of desires, but, like the latter, they are low-brow accounts which fail to bring into relief the evaluative nature of such states. As mere dispositions, functional states do not have the power to rationalize behavior: they lack any relation to the good. By conceiving action as incorporating an evaluative element, High Brow provides the missing element which allows us to see high-brow desires as rationally making sense of behavior.

2.5 Two Low Brow defenses

Summarizing the argument so far, the question we have been trying to answer is what truth there is to the Humean claim that if you desire φ you also have a reason to φ. As this depends on what we mean by "desire", we have plugged in the two most promising conceptions of desires to see if doing so makes the claim true. The shared explanation why either of the conceptions fails as an interpretation of the claim is that both conceptions take a low-brow view of intentional action. But if we think of wanting as merely experiencing a certain feeling or as exhibiting a pure behavioral disposition, we fail to account for the evaluative element of action. As High Brow insists, all intentional action involves some form of positive evaluation of the expected outcome on the agent's part. If a desire is to provide a reason, it must have an evaluative component.

The tentative conclusion is that low-brow views fail to validate the Humean claim. The conclusion is only tentative, however, as we still have to consider possible comebacks to the objection voiced in the previous section. The present section addresses two broad ways of defending the view that low-brow desires are reason-providing, and the next section a third one. To begin, when the argument concluded that a desire need not have normative force even for the agent himself, it relied on the assumption that a low-brow desire is something that an agent need not care about. Such a desire need not be seen as part of the agent's identity as a person. Seeing this assumption as an opportunity, the Humean could invoke Harry Frankfurt's view on personal identity. Harry Frankfurt asks what it is for an agent to act freely in the sense of acting as the expression of his own personal identity. His answer is that in order to do so, it is not sufficient merely to act out of a desire one has. The reason is that the agent may disown his first-level desires. The example he gives is that of an unwilling addict, who has frequent desires to inject drugs but prefers not to act on these desires. So although he frequently acts on his urges, he does so, in a sense, despite himself. He is not ready to stand behind his own choices.

The unwilling addict is, not an agent in the full sense of the word, but a "wanton" whose actions flow from desires he does not identify with. By contrast, an agent who identifies with his choice to φ not only desires to φ but also has what Frankfurt calls a second-order volition, a kind of desire to desire to φ. In effect the free agent wants to want to do φ and to be motivated by his first-order desire. The agent's freedom consists in his ability to be reflective by forming second-order desires that back up — or oppose — his base-level wants. With no second-order desires to

support his choices, the wanton fails to act in expression of his identity as a person.

A defense of low-brow Humeanism could posit that what is missing from the dispositionalist account so far is precisely this element of identification. On this proposal, only desire buttressed by a higher-order desire, as an expression of the agent's person, is capable of generating a reason. By contrast, wanton desires would not be intelligible as reason-providers. By deploying the idea of higher-order desires, the Humean could hope to avoid Quinn's counterexample.

We can agree that, if the strategy was sound, it would successfully rule out Radio Man's functional state as wanton. Unfortunately, the Frankfurt-style strategy does not prove workable, as becomes apparent in Gary Watson's criticism of Frankfurt's view.[37] For Frankfurt, an agent is wanton if he doesn't have any second-order volition to back up his first-order inclinations, but, as Watson points out, it is equally possible to be wanton with respect to one's second-order volitions. If Frankfurt asks whether this or that desire is what the agent really wants as an agent, we can equally ask whether this or that second-order volition is a genuine expression of the agent. The natural response is to allow for desires of a higher reflective level that back up the second-order desires in question. On this proposal, the agent's second-order desire is really the agent's own only if there is a further desire to back it up. However, this raises the obvious worry that such a theory generates an infinite regress. An agent identifies with his desire D_1 only if there is a further desire, D_2, which takes as its object the first-order desire. But then the original desire doesn't *really* seem the agent's own unless D_2 itself is not a wanton desire. Other than arbitrarily cutting off the chain, the only way to secure this further identification seems to be the postulation of a desire D_3 that shows that the agent stands behind D_2; and so on. But the assumption of infinite hierarchies of desires is a psychologically impossibility. Furthermore, the agent never seems to reach a point where he genuinely endorses a course of action. Aware of this type of worry, Frankfurt notes that

> [it] is possible [...] to terminate such a series of acts without cutting it off arbitrarily. When a person identifies himself *decisively* with one of his first-order desires, the commitment 'resounds' throughout the potentially endless array of higher orders [...] The fact that his second-order volition to be moved by this desire is a decisive one means that there is

[37] Watson (1982).

> no room for questions concerning the pertinence of desires or volitions of higher orders. (Frankfurt 1982: 91)

So Frankfurt allows the possibility of taking a decision as to which of one's desires are to count as genuine expressions of the person's identity. To do so is to undertake a commitment or to identify decisively with one's base-level desire. Accepting such a possibility, of course, allows Frankfurt to halt the regress: the act of identification brings further questions to a halt. But as Watson points out, this amounts to an admission that the notion of a higher-order desire cannot really perform the task Frankfurt assigns it. For according to the passage cited, desires, even higher-order desires, do not ultimately explain what it is for a person to identify with an inclination: it is the act of committing oneself to the inclination in question. Furthermore, once decisions of this kind are admitted, it is no longer clear why we need to assume that identification and commitment take place on some higher level of a putative hierarchy of desires at all. An agent can decide to commit himself directly to a given course of action as well as to a preexisting inclination. It follows that what an agent really wants, as an expression of his agency, is not what we wants to want, but the outcome of a different kind of attitude towards the state of affairs.

Contrary to Frankfurt's suggestion, the idea of a higher-order desire cannot in fact be what constitutes endorsement as an agent. Adding more elements of the same kind — more desires, albeit of a higher order — only adds further contenders with psychological force to the scene. Instead we should conclude with Watson that this endorsement is, not a low-brow attitude, but a *sui generis* attitude of valuing. But then we can simply accept valuing as applicable on the ground level, which renders a hierarchy of higher-order attitudes unnecessary. What is required for the agent's action to be really his own is not more desires of the low-brow kind. Instead an action the agent identifies with must be the outcome of a high-brow state such as Watson's attitude of valuing.

This insight is relevant to the Humean idea that desires give us reasons.[38] The sufficiency claim is sound only if we understand desires as evaluative attitudes. If instead we conceive them as, at base, nothing but mere inclinations, it is hardly clear why such states should be such that the agent necessarily needs to care about them, given that he does not necessarily consider them his own. But if even the agent himself, despite his own inclinations to act according to the patterns, doesn't in fact care about

[38]Watson makes this point in a footnote: "it is not the case that, if a person desires to do *X*, he therefore has (or even regards himself as having) a reason to do *X*" (Watson 1982: 100).

them, they can hardly provide him with genuine reasons. Genuine reasons can only flow from what the agent's positive evaluations of states — from what he holds, in some sense, to be good.

Reflections on Frankfurt's doctrine have led us straight to a high-brow conception of agency, but we may be declaring victory too quickly. Taking a different tack, the defender of a low-brow conception of desires may argue that Radio Man does not constitute a genuine counter-example to his thesis because, although desires normally generate reasons, there are exceptions, which include his strange functional state. This defense is a way for the Humean to retreat from the maximal position while keeping the spirit of the desire-based reason thesis. This idea offers the Humean the possibility to admit that the counter-examples are real; but he will hasten to add that they are counter-examples only on an overly strict interpretation of the sufficiency thesis. The Humean doesn't want to claim that any old desire suffices to generate a reason. On this proposal, not all desires are the same, and some have features that disqualify these particular states from actually providing a reason.

The Humean's task, then, is to identify a subset of the agent's desires which have the right features to generate reasons and to restrict the thesis accordingly. In effect, He proposes a filter acting on the class of the agent's functional states that removes objects that the agent doesn't desire in a narrower sense of the word, leaving only those that are desired in a privileged way. The challenge lies in specifying the kinds of desires in a non-arbitrary way. To give a simple example of an obviously inadequate specification, one might propose that desires give us reasons unless they resemble the case of Radio Man in some arbitrary way, say, in that the person acts on the inclination only once. The conditions the filter puts on what makes an object something the agent *really* wants must be general principles rather than ad-hoc responses to counter-examples.

According to the present proposal, only rational desires give us reasons, with the implication that Radio Man's desire to turn on car-radios is in some sense irrational, which would bar it from being able to give rise to a reason. Keeping in mind that the desire-based reason thesis has as a companion thesis the instrumentalist conception of practical reason, the Humean holds that desires cannot be criticized as irrational *per se* but only on some formal charge. What the Humean needs to do, then, is carefully to specify purely formal criteria of rationality, in a limited sense of that word, that pick out some desires as privileged or, alternatively, filter out some desires as only apparent.[39] The Humean may appeal to three princi-

[39]There are pitfalls associated with this approach. After all, the Humean intends to ex-

ples: rational desires must be (i) based on correct factual information, (ii) internally coherent or (iii) stable under reflection. Let us examine each proposal individually.

1. Rational desires must not be based on faulty factual information: Sometimes we develop desires as a consequence of misinformation. Suppose I want to fix myself a drink and believe that the clear liquid in the bottle in front of me is gin, so that I develop the desire to pour myself a glass of that liquid.[40] But I am mistaken about the bottle, which in fact contains petrol. As a consequence, despite my desire it doesn't seem true that I have any reason — any objective reason at any rate — to pour me a glass of the liquid.[41] According to the present proposal, the desire does not count as rational because it is based on faulty theoretical thinking. The proposal holds that this irrationality is a matter of defective theoretical or empirical reasoning. Because of this purported theoretical mistake, the proposal excludes this desire from the class of reason-providing states.

To say that a desire can be called unreasonable on account of its being derived from a mistaken empirical premises seems an unusual way of using the word "irrational".[42] Take a theoretical example. Suppose that, standing in the African savanna, I see a striped animal in front of me

plain the notion of having a reason by reference to the concept of a desire, so if we interpret "rational" as meaning or implying "reason-giving", the Humean account becomes obviously circular. This claim, amounting to the idea that desires give us reasons unless they don't, would be devoid of content. The current proposal must therefore include a different, independent definition of "rational". Furthermore, such a definition must not be substantive: saying that a desire is rational in the required sense should not embody particular normative advice about what is a good reason to do something. To do so would be to propose a substantive normative theory about reasons for action. His point of departure, remember, is that what we have reason to do derives not from any substantial normative theory — which always embodies particular rationalistic assumptions — but instead is a function purely of our contingent inclinations.

[40]For the example see Williams (1981a).

[41]Although arguably one has a subjective reason. For the discussion of the distinction between objective and subjective oughts, see chapter 6, particularly §5.

[42]Hume already speaks of affections as unreasonable because of mistaken empirical or causal judgments. In fact, these are the only ways in which he admits talk of irrational passions at all:

> 'tis only in two senses, that any affection can be call'd unreasonable. First, When a passion, such as hope or fear, grief or joy, despair or security, is founded on the supposition of the existence of objects, which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the design'd end, and deceive ourselves in our judgment of causes and effects. Where a passion is neither founded on false supposition, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. (Hume 1978: 2.3.3, 416)

and judge "This is a zebra". In fact the animal looks like a zebra, but unbeknown to me someone has painted stripes on a horse. My original belief, based simply on my visual perception and plausible assumptions, was wrong, but it wasn't strictly speaking irrational. Similarly, a further belief inferred from the first is likely to be false (though it could be true by accident) but it would be unusual to call the derived belief irrational on account of its provenance. But if a belief derived from a bad assumption is not irrational, we should not assume that desires that have come about in a similar way can be called irrational. A belief in a proposition may be irrational if it has been derived from another proposition by means of a bad rule of inference such as reverse modus ponens, the Fallacy of Affirming the Consequent. But simple factual mistakes need not be the result of bad reasoning. Similarly, for the question of whether a desire is rational it seems irrelevant whether the causal history of a desire includes untrue beliefs.

A desire deriving from a mistaken belief is such that the agent is likely to drop the desire if he became aware of its origin in a factual mistake.[43] Perhaps we can say that the desire, because of this fault, is not what the agent *really* desires. If we accept this, we can also admit that desires with this property constitute an exception to the sufficiency claim. However, clearly only derived desires can have this property. Desires that have not been derived from a factual premise cannot be "irrational" in this sense. Accordingly, the defense filters out only a small subset of desires, and in particular it does not apply to Radio Man's strange functional state, which we have described as in no sense derivative. In fact, the way Quinn describes the example all but rules out this interpretation of the case. For suppose that the agent has made a mistake about what turning on radios would accomplish — he thinks, perhaps, that doing so will bring him pleasure. Then what he really wants is to enjoy the expected pleasures of listening to the news or hearing music. Unlike many, Radio Man does not desire to turn on a stereo in order to listen to music or the news. He simply has the disposition to do it, with no ulterior motives; that's precisely what is strange about him. The idea that desires only give reason insofar as they are factually blameless does not mark Radio Man's functional

[43]Hume writes, perhaps too optimistically: "The moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without any opposition. I may desire any fruit as of an excellent relish; but whenever you convince me of my mistake, my longing ceases. I may will the performance of certain actions as means of obtaining any desir'd good; but as my willing of these actions is only secondary, and founded on the supposition, that they are causes of the propos'd effect; as soon as I discover the falshood of that supposition, they must become indifferent to me" (Hume 1978: 2.3.3, 416–7).

state as eligible.

2. Rational desires must not be incoherent: A natural way for a desire to be rational is for it to exhibit some form of coherence. Plausibly, when compiling a list of the things an agent *really* wants, we should omit the things he wants incoherently. However, it is not clear how the notion of incoherence, a property of a system of intentional states, applies to desires. Humeanism admits the possibility of means-end incoherence, the fault of an agent who fails to take the required means to an acknowledged end. But it cannot appeal to instrumental irrationality to filter out wayward desires because our task is precisely to explain what makes desires unfit to provide the starting point of means-end considerations.[44]

According to David Gauthier, an agent's wanting p can itself be incoherent in a formal sense, explained in terms of the notion of agential preference common in the social sciences and in economic theory.[45] Much work in these fields is predicated on the assumption that we can understand an agent's preferences primarily in terms of his observable choices. Preference, of course, is a cognate of desire, so that we can presuppose that it is possible to translate talk about the former into talk about the latter, or vice versa. The great advantage of preference-based theories is that they make preference amenable to measurement: we can observe what a consumer wants, for instance, by presenting him with the choice of a number of products. On a particularly strong version of this view, a person has just those preferences he exhibits in choice, so that preferences that don't show up in conduct are not taken as real. A rational-choice conception is often taken as a mere model of our behavior.[46] Versions of the view, accordingly, which, like Gauthier's, aim to represent our actual goals do not restrict their assumptions to revealed preferences — the behavioral component of preference — but also posit an attitudinal component of preference, which is taken as persisting even in the absence of relevant behavior.

The relation of preference holds between an agent and two options. It is one of the assumptions of this account of preferences that they are transitive. For instance, a child who prefers bananas to apples and cherries to bananas, when given a choice of cherries and apples, will be expected to choose the former. Now, however, suppose we attribute to the child also the preference of apples to cherries. If so, there seems something

[44]This is not to say that the Humean may not in some other way invoke instrumental rationality as the ingredient of low-brow conceptions of desires that allows desires to exert normative influence. This possibility will be explored in §6.

[45]Gauthier (1986: 40ff).

[46]Cf. Gauthier (1986: 27).

wrong with his set of values. His cyclic preferences make it impossible for him to make a rational choice between different types of fruit. Cyclic preferences are puzzling to formal theories of rational choice, which often assumed that we can, from a given number of two-place preferences, form a preference ordering, starting from the preferred option and leading to the least preferred. Behavior resulting from cyclic preferences can be described as inconsistent in itself. In the light of these formal difficulties, we may grant that a desire fails the requirement of coherence if its preferential equivalent is part of a group of preferences which share the structure of the fruit example.

Once again, however, if the Humean argues that preferences that rationalize must be coherent in this sense, this filters out only a small fraction of our desires. This particular kind of irrationality seems to be a relatively rare occurrence. In particular, this defense doesn't help to explain away the counter-example we have been considering. If Radio Man has a preference for turning on car-radios over doing something else, there is no reason to suppose that it is part a series of cyclic preferences. Radio Man's irrationality, if he in fact is irrational, does not seem to consist in a formal difficulty of this type at all. Even if the idea of cyclic preferences allows the Humean to rule out certain pro-attitudes as irrational, unless the Humean can extend the applicability of charges of incoherence, this minimalist concept of rationality does not address the problem.

3. Rational desires must be stable under reflection: Whereas the two preceding principles fail to apply to Quinn's counterexample, this further idea is applicable. Perhaps Radio Man's functional state seems arbitrary or *ad hoc* because it strikes us as capricious. Thus the Humean may say that in order for a desire to rationalize the behavior it produces it cannot be a one-off urge or whim. Again the idea is proposed by Gauthier, who writes that "[i]n the absence of full reflection on one's preferences choice tends to go astray" (Gauthier 1986: 31). On this proposal, a desire which generates reasons must have persisted through a period of reflection. The idea seems to be taken from everyday situations: if I come to want a motorcycle in the heat of the moment, this may not in fact reflect what I *really* want, even if that is how it feels at the moment. To ascertain my real or rational desires, I need to sit down in a quiet hour and reflect on the desire. Perhaps if I give it some thought, use my imagination to picture the possible implications and dangerous consequences, my rash desire will evaporate. However, if the desire remains in existence even after closer scrutiny, this shows that the desire is mature and stable. The Humean may suppose that only in the latter case is it correct to say that the desire gives me a reason or, as Gauthier puts it, reveals my values.

In an example, he envisages a young woman who, pressed for a choice, accepts a marriage proposal, only to regret it later:

> The young lady's acceptance of the proposal (or proposition) is equally not a basis for determining her values. She may not lack the experience to form a considered preference. But she fails to reflect before accepting. When she does come to reflect she realizes that her firm preference is expressed by 'No'. Does she change her initial preference, or does she correct it? Neither alternative is acceptable; the first would imply that her successive preferences stand on an equal footing; the second would imply that initially she misstates her preference. A better account is that without due consideration, she forms a tentative preference revealed in her acceptance, which she then revises. Any choice reveals (behavioural) preference, but not all preferences are equal in status, so that the rejection of a tentative preference is not to be equated with the alteration of a firm or considered preference. Of course not every rejection of a tentative preference need involve the formation of a firm one. The young lady's 'No' might have been followed by yet another 'Yes'. But we may suppose that in this case further reflection only confirms her second preference. And so we treat it as affording a basis for determining the young lady's values. (Gauthier 1986: 30)

In Gauthier's view, although the initial acceptance of the proposal expressed a preference, it was only a tentative preference. Not all preference are on a par. A tentative preference is valid only if affirmed in further contemplation. If thinking things over does not change the initial decision, the agent develops a firm, considered preference; otherwise the preference is dropped. To Gauthier, only firm and considered preferences form a basis for determining the woman's values and, as we may add, her reasons, as behavioral or even attitudinal preferences only count as an expression of what she really wants if they have been reflected upon.

Now it is certainly true that we sometimes form a desire or preference hastily, without thinking about it enough, and that this can be a ground for discounting its importance or relevance. On the other hand, it doesn't seem correct that everything we genuinely want is the outcome of long reflection on the subject. Sometimes we act spontaneously and still for genuine reasons, and there need be nothing wrong with such actions. Often, in particular in cases like the above that involve planning one's future life, reflection will improve the overall coherence of our values;

but in many cases, it seems superfluous. Therefore, even if it sometimes improves matters, it doesn't seem correct to call stability under reflection a general requirement for desires to count rationally. In particular, it doesn't seem true that the reason why Quinn's Radio Man doesn't have a reason to turn on car radios is the fact that he hasn't reflected enough on his behavioral pattern. Perhaps if he did reflect on it, his tentative preference would even confirm itself.

Nonetheless the Humean may insist that Radio Man only persists in his rather monotonous activity because he is insufficiently reflective. But this is not a point that counts in favor of low-brow conceptions of desire. To see this, we need to ask what explains the fact that reflection brings him closer to what he really wants. It seems that reflection has the function of making the agent abandon habitual behavioral patterns in favor of active deliberation about what it is in life that he values. It helps the young woman to think about the proposal in quiet because this prompts her to think about what, for her, a good life consists in. Rather than mechanically giving in to impulses, she frets over what she values. If this is so, however, then the situation is no longer one in which an agent has a desire conceived as a purely functional state. What may start off as a mere behavioral tendency turns, through Gauthier-style reflection, into an endorsement or rejection of proposed action. In particular, her reflection on her preference makes it more stable and reliable, but it also adds an evaluative aspect to her attitude. This explains why reflection makes a preference a better basis for determining the agent's values: that the agents gives the question more thought is evidence that she is actively making an evaluative judgment. It is also likely that a period of reflection helps bring the preference more in line with other goals. In the young woman's case, the work done by thinking things over is to weigh different value against each other and to find out whether or not accepting the proposal would cohere with these values.

Similar things must be true of the Radio Man. If it is true that he doesn't reflect on his choices, this does indicate, as the Humean proposal says, that he doesn't have a reason to act as he does. However, it doesn't indicate this because of a principle that rules out unreflective functional states as the source of reasons. Instead, what explains that the agent doesn't have a reason is that he doesn't in any way see the actions he performs as good or desirable. The fact that the Radio Man does not reflect on his preference is only a symptom of the more important fact that he lacks a high-brow view of the action. If he judged the actions to be good in some way or other, this would be enough to support the idea that the attitude gives the agent a reason. But if he fails to take an evaluative stance, his

low-brow functional state is powerless to make sense of his action.

What prevents Radio Man's strange functional state from generating reasons, then, is not its spontaneous, unconsidered nature but the fact that it is a low-brow state. We can conclude that the three proposals considered either do not achieve their goal of filtering out weird desires as "irrational" or collapse into a high-brow conception of desire.[47]

2.6 High and low

Humeans hold that low-brow states, in particular behavioral dispositional, are capable of rationalizing our behavior.[48] The arguments rehearsed in the last two sections suggest that the claim cannot be upheld. It may be said, however, that we have not been interpreting the Humean claim charitably enough. Along these lines, the Humean can mount a final defense of his position. He agrees he is committed to the sufficiency of desires-as-dispositions for reasons, but he insists that desires should be seen in the context of a more complex theory of agency than the one we have been considering. Specifically, he protests that we have neglected to take into account the entire *instrumental* dimension of desires. Pleading that the picture of the agent is too simplistic, he promises that once we take into account a fuller portrayal of the agent in action, the idea that our reasons are provided by low-brow states is no longer mysterious.

To see how such a defense could run in detail, consider again the two broad views of agency we have been working with. One side of the contrast, High Brow, takes intentional action to involve regarding its object under the guise of the good, which requires the agent to take an evaluative stance on the action's outcome. On the other side of the contrast, Low Brow sees acting on a desire as the actualization of a behavioral disposition. At this point, the defender of Low Brow may protest that this characterization of the Humean view makes the intentional action of an agent seem too much like animal behavior caused by an appetitive inclination. The fact that a Humean agent, like a fish, does not (necessarily) regard his doings as good does not imply that he should be seen as similar to a fish as regards mental complexity.[49] Thus Railton appeals to our

[47]For a penetrating discussion of Gauthier's proposals, see Brandom (2001b).

[48]From here on out, I will disregard the phenomenological conception to focus on dispositionalism as the Humean's best option.

[49]The term "Humean agent", like the term "Humeanism", is not intended to imply that Hume endorsed a picture along these lines. "Humean agent" is a way to refer to the kind of individual that does the things the Humean theory expects it to, thinks the thoughts the

intuitive sense that there is

> a great deal of psychic distance between a fish that swims to the surface "because it is hungry" and a child who responds to our question "Why did you come downstairs?" with the answer "I'm hungry". (Railton 1997: 306)

We shouldn't assimilate the desire of a Humean agent such as a child, which in Railton's view doesn't have the ability to make evaluative judgments, to an animal's mere functional state. We can reject the evaluative component of agency without going so far as to identify desires with states similar to animal appetites. The idea is that there is room for more sophisticated low-brow conception.

According to Railton, the typical Humean agent occupies a middle ground between having a full-blown evaluatively loaded desire and having a mere functional state. It is not, however, immediately obvious how such a sophisticated version of the low-brow conception of agency would look like. In which sense are an agent's desires more than mere functional states, especially in terms of their power to rationalize action? If a sophisticated version of the low-brow view is to work, there must be some way to enrich the low-brow conception to explain the psychic distance between Humean agents and pre-rational creatures without full access to agency. The idea is not to replace the dispositionalist conception of desires with another view but to reconsider it in a different light.

According to defenders of the low-brow conception, the way Humean agents interact with the world is significantly more complex than the picture of pre-rational agency suggests, which is reflected in the agent's capacities. For Railton, Humean individuals "exemplify agency even though they do not by their nature 'aim at' the good" (Railton 1997: 305). Railton takes up this task by noting biographical details of a Humean agent:

> Humean individuals engage in both theoretical and practical reasoning. They inquire into causes and effects; form beliefs about the conductiveness of means to ends; take into account the relative strengths and independence of desires; acquire habits; form intentions to act; and formulate and respond to rules and sanctions. Their conduct therefore can, it is claimed, be given fully fledged intentional, rational-agent explanations. Why-questions about their conduct can often

Humean theory predicts, and so on, and it is a well-established term in the literature.

> be answered correctly by citing *their reasons* for behaving as they do, and these will include: how they represented the situation, what their goals were, how they weighed their various ends, how they adjusted means and ends, and so on. (Railton 1997: 305)

What makes a creature a Humean individual — and thus, if Low Brow is correct, an agent — is his possession of a whole array of interrelated practical capacities. He is able to cope with the fact that he may have several desires that cannot be satisfied at the same time. In order to navigate his way through the world, he employs beliefs about the world, and in particular empirical beliefs about how to effect a change. To do so, he assigns different desires different strengths. His means-end calculations cause him to form intentions and to make decisions.

On Railton's view, unlike a pre-rational creature, a Humean individual is perfectly capable of acting on reasons in the full sense. To illustrate the difference, he gives the example of a girl and her younger brother before a weekend trip with their parents to the beach:

> [The] two children have been begging all week to go to the shore. Both, however, dislike long summer car rides. When the weekend comes one child absolutely refuses to get into the car. "But we're going to the beach, which you love!" "I don't care. I don't want to ride in this stuffy old car. I hate it! I won't do it." He has to be carried bodily to the car and buckled in, thrashing. Once in the car, he still refuses to be jollied along. "It's *your* fault I'm in this stuffy old car! I told you I hate it." The second child confines her thrashing to loud complaints. "Not another car ride! Last time I felt sick the whole time!" But when the time comes to leave she climbs into her seat of her own accord, waiting sulkily to be buckled in. On her face is a look that says "Okay, I'll ride in the car, but don't expect me to like it." (Railton 1997: 306)[50]

In this scenario, the children both have a desire to go to the beach, although their reactions are different. Let us accept with Railton that, despite their desire, neither of the children can be said to judge that going to the beach would be a good thing. Maybe they're not yet capable of such evaluations in the full sense. In any event, the attitude of either child toward the prospect of going to the shore is a low-brow attitude. What

[50]Railton attributes the example to Jean Hampton.

is more, the children share a dislike for the required means to going to the shore: they hate enduring a long ride in a car. What, then, is the difference between the two children? One child — let us say: the younger brother — entirely refuses to enter the car at his own will, i.e. to take the required means towards an end he desires. His two desires — going to the beach and staying out of the car — are in conflict, and he lacks the capacity to weigh his own desires. The more salient desire of avoiding the car ride gets the upper hand, but seemingly without the involvement of the agent. The child's dislike of the back seat pushes it to throw a tantrum.

This is a sign that the younger child doesn't fully exemplify agency. In particular, he lacks the ability to control his own desires and to harmonize them in the case of conflict. By contrast, the older sister is more mature in this regard. Like her brother, she doesn't like to enter the car, but she grudgingly complies because she is able to see the car ride as a means to the end she looks forward to. Crucially, she makes use of her ability to adjust means to ends. She doesn't let herself be controlled by what desire happens to be felt the most strongly at any given time but instead exerts active agential control over her own inclinations. She is able to weigh her greater long-term interest against the prospect of a minor but imminent displeasure. And she is able to adapt means to ends even if the means happen to be unpleasant. For Railton, she has the qualities constitutive of a Humean agent, qualities which her brother lacks.

Railton emphasizes that seeing the older child as in charge of her own desires doesn't force us to attribute to her positive evaluations of the states of affairs she wants. As he describes the scenario, she does not make any judgments, positive or otherwise, about the day at the beach. Nonetheless, her abilities as a Humean agent allow her to treat going to the beach as an end. For Railton, her case shows the difference between mere desires, which can be attributed even to the brother, and what we may call desires-as-ends. Like her brother, the older child is influenced by her desires or inclinations. But in a Humean agent desires "function as end-setting" (Railton 1997: 307). On Railton's view, what separates Humean agents from sub-Humeans behavers, then, is the ability to adopt and pursue ends as such.

For Railton, the addition of means-end oversight makes a difference to the rationalizing power of desires. While we have argued that the low-brow conception of desires is no basis for the view that desires provide us with reasons, Railton replies that this is true only for immature proto-agents. With respect to the brother, he concedes, we cannot say that his desires gives him a reason to go to the beach or to wait out the car

ride, or any reason at all. Because of his lack of practical competence, his desires do not set ends and therefore fail to rationalize his behavior. But to Railton it is a mistake to equate the Humean position with the view that assimilates intentional agents to the brother. An enlightened low-brow conception takes a view of Humean agents as closer to his sister, who has the ability, among other things, to rank desires and adjust means to ends. She does not just chase inclinations but pursues ends. For Railton, this means that we can appeal to her desires in intentional or "rational agent" explanations as things which rationalize her behavior. On this view, the inclinations of Humean agents have reason-giving power.

Railton is right to point to the various abilities associated with practical reasoning as prerequisites of responding to a consideration as a reason. However, he also holds that these abilities do not require acting under the guise of the good. Her further capacities notwithstanding, the girl in his example is acting on desires which, conceived as functional states, do not aim at the good. Mere dispositions, we have established, are insufficient to make an action reasonable. It seems that the addition of agential capacities alone doesn't change the fact that the girl's inclinations do not have to matter rationally to her. We may ask what about these abilities it is that makes the difference between a pre-rational behaver and a Humean agent. The answer must be that only a rational agent treats the things she desires as ends. But it is hard to see that one can treat something as one's end without regarding it in some way as good.

What additional resources does graduating to a full Humean agent bring to the table? The Humean may answer that what makes the difference is the fact that the Humean individual engages in instrumental reasoning. The Humean seems to rely on the idea that the fact that the instrumental principle holds gives the desire an additional significance. But that amounts to the idea that participation in instrumental rationality somehow manages to elevate a mere functional state into something with normative significance. It is hard to see how the instrumental principle could possibly achieve this feat all on its own. After all, means-end rationality is often said to be merely capable of transferring normative significance from the means to the end. But a transfer is not what we are envisaging at present. On the present proposal, the role assigned to instrumental thought is that of endowing a state which was previously normatively innocent with a normative standing. It is deeply mysterious how this could be instrumental rationality in operation.

Instrumental rationality requires you to take the means necessary to attain you ends. The story the Humean story would need to tell is that, as a Humean agent, by developing a desire-as-disposition, you become

subject to hypothetical imperatives. So on this line of thinking desires generate imperatives of the form:

> If you want to φ and believe that ψ'ing is required for achieving φ, then you ought to ψ.

The idea, then, must be that by way of generating imperatives, i.e. requirements, the desires become ends. However, this is unconvincing. In order to have force for an agent, a hypothetical imperative itself presupposes that an agent is already committed to pursuing an end — it cannot make it the case that the agent has the end. But surely the agent's being committed to a project is founded on a positive evaluative attitude towards the project. Practical imperatives matter rationally only to an agent aiming at the good.

Finally, the instrumental principle can make a difference to the agent only if the agent is in fact being guided by the principle in his reasoning. Railton's Humean individual engages in means-end arguments, true, but he falls short of being guided by the principle that he ought to take the means to his acknowledged ends. On Smith's dispositional theory, if he desires to φ, he has the behavioral pattern to do ψ insofar as he believes that φ'ing requires ψ'ing. Suppose he wants to φ and believes that ψ'ing is necessary to do so. Then either he really does ψ, or he fails to do so. If the latter, we are having difficulties describing the situation properly. The fact that he does not take the means does not indicate that he is making a mistake rationally. Instead it seems to suggest that the agent didn't really want to do φ in the first place. For if to desire φ is, *inter alia*, to take the required means, then failing to do the latter casts doubt on the former. The low-brow proposal does not allow for the true possibility of violating the instrumental principle.

Conversely, if the agent does take the means, this is a consequence of his behavioral disposition to do what coincides with the dictates of instrumental rationality. However this is what, given his disposition, he was going to do anyway. With the dispositional interpretation of the instrumental principle, to do differently was not really an option. It follows that in taking the means, he is not being guided rationally by a principle.[51]

[51]For an extended and difficult argument to this effect, see Korsgaard (2008: 32–46). According to Korsgaard, empiricist accounts of agency cannot account for the fact that the instrumental principle not only motivates but functions "as a requirement or a guide" (Korsgaard 2008: 52). On Korsgaard's view, if the Humean gives a dispositional account of instrumental reason, it turns out to be strictly speaking impossible for the agent to refuse to take the means *M* to the end *E*, for if the agent does not do *M*, it follows that he has not really accepted *E* as an end.

To conclude, Railton is right to call attention to instrumental reasoning as a crucial component that allows us to see an agent as acting for reasons rather than merely following inclinations. However, he neglects the way in which the instrumental principle is itself a way of spelling out what it means for an agent to see something as good or to treat it as his end. If this much is true, Low Brow cannot help itself to the notion of the instrumental principle in explaining why a low-brow agent's inclinations can justify his behavior because genuine instrumental rationality itself is in reach only for agents aiming at the good. Railton's contention that, in an individual with purely low-brow resources, desires may nonetheless act as end-setting cannot be upheld.[52]

[52]For an answer to the question how we should understand the instrumental principle, see chapter 6.

Chapter 3

High-brow commitments

3.1 Kripke's rule-following paradox

As the arguments of the last chapter have shown, any low-brow conception of agency faces serious difficulties. We have considered two Humean attempts to clear away the difficulties: demarcating a class of "irrational" desires and appealing to the instrumental reasoning of a Humean individual. Neither of the strategies has succeeded in defending the reason-giving power of low-brow dispositional states, which led us to reject the Humean account of action in general. Ultimately, however, the arguments are persuasive only if an alternative, high-brow theory can be shown to be viable. In order to do so, we first need a clearer idea of what such a theory entails. High Brow holds that in acting for a reason, an agent values the goal of his activity or regards it as, in some sense, good. Relatedly, if we are to understand intentional action generally as implying a want, we need to take a high-brow view of the intentional state of desire as well.[1] On a high-brow view, an agent who wants to perform an

[1]This formulation is conditional: we need to understand desire in a High Brow way *if* (as in Davidson's belief-desire model) we tie action to wanting. On the classical view, desire is the principal intentional state in the practical realm, and all intentional actions are accompanied by a desire. However, I will reject the classical view below in favor of a Sellarsian alternative (§4). On my view, intention — not desire — is the quintessential practical state. If we conceive intention rather than desire as the state that accompanies action, it is not, strictly speaking, necessary to argue for a High Brow conception of desire, since what is most important is a High Brow conception of intention. On the other hand, much of the discussion has been conducted in terms of desires rather than intentions. Therefore, in order to be able to engage with the literature, my argument in the following sections (§1–3) will

action is someone who regards the action or its expected outcome in a positive light. When we desire, we aim at the good.

The first question raised by this proposal is what it means for a state to aim at the good. This chapter will be concerned primarily with this question. Having been critical of the dispositional approach to desire, we are particularly interested in the way in which a state that aims at the good is anything more than a mere dispositional state. Investigating this question will first require some beating about the neighboring bushes. Kripke's discussion of the so-called rule-following paradox contains an important critique of dispositionalism about meaning, which, as we find, applies similarly to dispositional accounts of desire (§1). The discussion leads to the general thesis that semantic and mental content are normative (§2). It follows that desires, as contentful states, are rule-governed states. Two common ways of distinguishing desires from beliefs — through the notions of constitutive aim and direction of fit — are considered and interpreted in line with this conclusion (§3). "Practical commitment" is introduced as a notion to replace that of desire, as it better reflects the normative nature of the intentional state that accompanies action (§4). Finally, we address the question where this notion leaves us relative to the Humean doctrine (§5).

In order to move beyond and improve upon the dispositional theory of desires criticized in the previous chapter, we should start by pinpointing its difficulties. We will take our cue from Saul Kripke's seminal criticism of another dispositional theory, viz. the theory that what a person means by his words is constituted by facts about his behavioral propensities. By way of setting the stage, let us start by describing the paradox he lays out at the beginning of his *Wittgenstein on Rule and Private Language*, a reading and interpretation of what has become known as Wittgenstein's rule-following remarks.[2]

continue to target the classical question of how we should understand desires. The caveat is that, ultimately, desires will be seen to be less important than the closely related states of intention or practical commitment. We will drop the working assumption that desire is the crucial practical state in §4.

[2]Rather than portraying it as an exact representation of Wittgenstein's own negative and positive views, Kripke offers his interpretation of the paradox as "Wittgenstein's argument as it struck Kripke, as it presented a problem for him" (Kripke 1982: 5). Likewise, here we will not delve into the numerous interesting, but difficult, questions of Wittgenstein hermeneutics. Instead, our focus will be on Kripke's presentation of the important rule-following difficulties which, everyone should agree, are genuinely Wittgensteinian at least in their general direction, and in particular on Kripke's ingenuous treatment of dispositionalism.

Kripke asks us to imagine a competent speaker of the English language who knows the use of the word "plus" or the symbol "+" but who, although he has performed a large number of additions in the past, has never made the particular calculation "68 + 57" before. Suppose that all additions he has performed before operated on numbers smaller than 57. Now the person performs the calculation and, not surprisingly, obtains the answer "125". At this point, however, a skeptic intervenes and questions the person's confidence that he is giving the right answer. He brings up an alternative possibility: "Perhaps given how you used the term "plus''in the past, the answer you *intended* for this calculation is "5''rather than "125''." What the skeptic suggests is that the speaker's certainty about the calculation may not be justified. As the skeptic points out, the subject never gave himself explicit instructions as to what the answer to *this particular* arithmetical task should be, and, as we have stipulated, he has never attempted the task before. So what the subject must do is to apply the same rule which he previously meant by the expression "+" and which he has applied many times before. However, the skeptic argues that the subject may have used the expression "+" to denote, not plus, but quus, a made-up function which yields the result of addition for numbers smaller than 57, but 5 otherwise. If this skeptic is right, the correct answer to the question would have been, not 125, but 5.

Although the skeptic's hypothesis is wholly implausible, it turns out to be hard to refute. The challenge raised by the skeptic is the following: "What evidence can you adduce that shows that what you meant by "+'' in the past was plus rather than quus?" If the skeptic's hypothesis — that the person always intended "+" to mean quus — is as absurd as it sounds, there should be some facts about the person's past usage of the word the speaker can appeal to refute the claim. Kripke writes that "[o]rdinarily I suppose that in computing '68 + 57' as I do, I do not simply make an unjustified leap in the dark. I follow directions I previously gave myself that uniquely determine that in this new instance I should say '125'" (Kripke 1982: 10). We are usually certain about our usage, but in the light of the skeptical question, the application of as straightforward a rule as addition suddenly seems to be "an unjustified stab in the dark" (Kripke 1982: 17).

In order to show that the subject is justified in answering "125" we may point to the fact that he grasps the rule of addition. But that only raises the question what grasping the rule amounts to. What facts determine that the rule the subject has in mind is addition rather than quaddition? In the course of his discussion, Kripke considers a number of theories about what the fact that the agent means plus rather than quus consists in, theories that purport to show that the right answer to the question,

as the subject intended it, is "125". The first suggestion is to invoke facts about the history of the agent's usage of the word. In the past, he has used addition many times. However, as we have stipulated, prior usage never included the addition "68 + 57". As a result, all previous instances of addition are just as compatible with the hypothesis that what he meant by the word was "quus" as they are with the hypothesis that what he meant was "plus". If the directions the person gave himself are nothing more than a finite number of previous applications, then for all we know, he may have meant quaddition all along. The person's psychological history doesn't seem to settle what he meant by the word.

Kripke spends much effort discussing a dispositionalist answer to issue raised by the skeptic, which is also arguably the most promising naturalist reply available. Our interest in what follows is in Kripke's criticism of this approach.[3] According to this answer, there are facts about the subject which determine what he means by the word "+": facts about the subject's dispositions. We can understand this as a way to escape the limitations of the obviously inadequate idea that my previous additions settle what answer I should give now. Because the history of my previous usage is limited to a few thousand instances, we can easily find an application of the rule that hasn't been covered by it. On the other hand, dispositions can manifest themselves in a greater number of ways, so the dispositionalist answer holds the promise of explaining why I mean addition even in novel applications.

According to a dispositionalist analysis, to mean addition by "+" is simply to be disposed to reply with the sum of two numbers when asked "What is $x + y$?" That the subject means addition, then, implies that the subject is disposed to reply "125" to the question "68 + 57", "200" to the question "100 + 100", and so on. On the other hand, if the subject meant quaddition by the expression, he would tend to reply "5" to the same question. Even if the actual answers given in the past do not allow us to tell whether the subject meant addition or quaddition, it may still be true that the subject, had he been asked the question, would have replied "125".

Kripke's riposte to the dispositionalist defense is to point out that the skeptic challenged us to name a fact about the subject that we can appeal

[3] A discussion of the skeptical dimension of Kripke's (and Wittgenstein's) paradox is outside the scope of this chapter. Kripke goes on to consider a number of defenses against the skeptic, including a phenomenological account and a Platonist account of meaning, all of which he finds incapable of meeting the skeptical challenge. After dismissing these suggestions, Kripke proposes, in chapter 3 of his book, what he calls a skeptical solution to his paradox. In the text I focus on the difficulties of a naturalistic or reductionist solution and assume that we can live with the conclusion that no descriptive or naturalistically unobjectionable facts can serve the role of meaning-facts.

to as a *justification* of the answer to the mathematical question. What the skeptic questioned was our confidence that it was really addition that the subject meant all along. The answer seemed merely a "stab in the dark". But it does not make sense to react to such a challenge by pointing to a disposition. If the dispositionalist says that "125" is what the subject tends to reply and that earlier the subject would have given the same response if he had been asked, this misses the point of the skeptical question:

> How does any of this indicate that — now *or* in the past — '125' was an answer *justified* in terms of instructions I gave myself, rather than a mere jack-in-the-box unjustified and arbitrary response? (Kripke 1982: 23)

As Kripke points out, the trouble with the dispositional view is this: it doesn't "*tell* me what I ought to do in each new instance" (Kripke 1982: 24). Instead, it simply seems to rubber-stamp the agent's behavior, whatever it may be. This, however, cannot be what the meaning fact consists in. In particular, the dispositionalist theory seems to entail that all our behavior is automatically, and incredibly, correct according to the rule:

> A candidate for what constitutes the state of my meaning one function, rather than another, by a given function sign, ought to be such that, whatever in fact I (am disposed to) do, there is a unique thing that I *should* do. Is not the dispositional view simply an equation of performance and correctness?

Kripke points to Wittgenstein's view that if, concerning a subject, "whatever is going to seem right to me is right,", that indicates that something is amiss. For Wittgenstein, "that only means that here we can't talk about right" (Wittgenstein 1958: §258). If a theory implies that, with respect to some rule, any performance is, by definition, automatically right, then nothing is. The difficulty becomes apparent once we notice the possibility of *systematic* mistakes. It is not unusual for a subject to have dispositions to make mistakes, even in areas of their competence. A student may have the disposition, under certain circumstances, to give an answer that deviates from the one found in the addition table. The normal thing to say would be that, in giving deviating answers, the subject is making mistakes. But according to the dispositionalist, we can read off the function that a person means from his dispositions. Thus the dispositionalist seems to be committed to the absurd conclusion that, strictly speaking, we cannot make systematic computational errors. The answer the agent

would have given becomes identical to the answer the agent *should* have given.

Here is Kripke's summary of his discussion of the dispositionalist account:

> Suppose I do mean addition by '+'. What is the relation of this supposition to the question how I will respond to the problem '68 + 57'? The dispositionalist gives a *descriptive* account of this relation: if '+'meant addition, then I will answer '125'. But this is not the proper account of the relation, which is *normative*, not descriptive. (Kripke 1982: 37)

He goes on to explain the crucial contrast between normative and descriptive accounts:

> The point is *not* that, if I meant addition by '+', I *will* answer '125', but that, if I intend to accord with my past meaning of '+', I *should* answer '125'. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be *disposed* to respond as I *should*, but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is *normative*, not *descriptive*. (Kripke 1982: 37)

The core difficulty of the dispositional view, then, is that it is a descriptive rather than normative account of the relation between meaning and future action. The problem calls for a different approach to the general question of how we can use a term with understanding.

Kripke's criticism of dispositionalism captures a profound truth. The important point, for our purposes, is that it has implications for a correct theory of desire and thus for our theory of reasons. Kripke himself draws attention to this relationship in an extensive footnote.[4] There Kripke notes that Russell's dispositional theory of desires in terms of behavior-cycles, which we have discussed earlier, was a target of Wittgenstein's criticism during the development of the views which were later published in the *Philosophical Investigations*.[5] Wittgenstein's thought about meaning took one of its early forms in his criticism of Russell's behaviorist conception of the relation of desire to its object. Drawing out the implications, Kripke writes:

[4] Kripke (1982: 25, n. 19).

[5] For the discussion of Russell's view, see §2.4.

> Clearly the sceptic, by proposing his bizarre interpretation of what I previously meant, can get bizarre results as to what (in the present) does, or does not, satisfy my past desires or expectations, or what constitutes obedience to an order I gave. (Kripke 1982: 25, n. 19)

and adds:

> Russell's theory parallels the dispositional theory of meaning in the text by giving a causal dispositional account of desire. Just as the dispositional theory holds that the value I meant '+'to have for two particular arguments m and n is, by definition, the answer I would give if queried about 'm+n', so Russell characterizes the thing I desire as the thing which, were I to get it, would quiet my 'searching' activity. (Kripke 1982: 25 n. 19)

Here Kripke suggests a basic problem for dispositionalist accounts of desires. In Russell's view, a desire is defined in terms of the thing which tends to bring quiescence to my searching behavior. It follows that the object of a desire is the thing which if obtained would cause the desire to disappear. In a rare (almost) direct criticism of another philosopher's view, Wittgenstein points out the consequences of Russell's account. He writes in the *Philosophical Investigations*:

> Saying "I should like an apple" does not mean: I believe an apple will quell my feeling of nonsatisfaction. *This* proposition is not an expression of a wish but of nonsatisfaction. (Wittgenstein 1958: §440)

Again, in the *Philosophical Remarks*, he writes:

> I believe Russell's theory amounts to the following: if I give someone an order and I am happy with what he then does, then he has carried out my order.

and goes on, in parentheses:

> If I wanted to eat an apple, and someone punched me in the stomach, taking away my appetite, then it is this punch that I originally wanted. (Wittgenstein 1975: §22)

On the dispositionalist theory of desire, the object of a desire is determined, as it were, after the fact. If it turns out that a punch in my stomach will bring an end to my desire, we must conclude that its object was not, as I thought, the apple but the punch. Clearly this consequence is absurd: the object is not determined in such a way as to make such a switch possible. The problem is one of intentionality. As Wittgenstein points out, Russell's theory portrays the relation between desire and action as an "external relation", whereas what we need to do is to see the relation as "internal".[6] The object of the desire — my taking a bite of the apple — is already contained in the desire, conceptually, even if it doesn't exist yet, seeing as I don't have the apple yet.

The following section argues more carefully for a non-descriptive theory of desires and gives a sense to the idea that desire is related internally, rather than externally, to the action done from it. For now, let us restate the Kripke-Wittgenstein point about desires. There are correct and incorrect ways of reacting to a desire. Correctness or incorrectness here is always relative to the content of the state in question. We do not have a reason to (try to) satisfy all our desires, and some of our desires are such that we ought not to pursue them at all. But the desire imposes internal standards of correctness with respect to its conditions of satisfaction. And as long as the desire is in place, so are the standards. This suggests that Kripke's point with respect to the mental state of meaning something by a term carries over to the state of desire. Wittgenstein's polemic against Russell can be translated into Kripke's idiom: the relation between the desire and the action which follows from it is normative, not descriptive.

3.2 The normativity of desires

If, as Kripke argues, a person's meaning a concept by a linguistic expression is not reducible to dispositions, then, if the above suggestion is correct, the same must be true of desires. To see this, we need to (i) state and refine the general idea of the normativity of semantic content, (ii) extend it to the content of intentional states and (iii) ascertain that the idea applies to desires as well as to other intentional states.

i) The normativity of meaning: Kripke's criticism of dispositionalism allows us to draw a positive conclusion about what it is to use a linguistic expression with understanding. As we saw, Kripke holds that the relation of meaning to future action is normative: The availability of alterna-

[6] Wittgenstein (1975: §21).

tive skeptical hypotheses shows that facts about a speaker's past usage or his dispositions do not exhaustively determine the speaker's meaning. Therefore, the fact that a speaker means something by a certain expression cannot be reduced to the fact that the speaker has used the expression in a certain way in the past or facts about his hypothetical behavior. What, then, is it for the speaker to mean something by the term? The picture suggested by Kripke's discussion is different from the reductive theories he considers.[7] On this picture, for a person to mean something by an expression is for him to be guided by a linguistic rule which governs the use of the expression. The picture is closely related to one proposed in more explicit terms by Wilfrid Sellars. Here is Sellars's statement of the core thesis:

Language-rule thesis The linguistic meaning of a word is entirely constituted by the rules of its use.[8]

The high-brow theory of agency presented in this chapter and the next is largely based on a Sellarsian understanding of linguistic meaning, mental content and following. or being guided by, a rule. To make progress on the main topic, then, it will be necessary to introduce some core elements of Sellars's doctrine, including the language-rule thesis. The immediate consequence of this thesis is the Normativity of Content. If what a person means by a word is a matter of rules, then statements about meaning are, or imply, statements about how the person ought to behave in various circumstances. This implication has often been seen as worrying. According to a long-standing doctrine, we cannot derive *ought* from *is*. It follows that, if meaning statement imply oughts, we cannot derive what

[7]What does Kripke's (or Kripke's Wittgenstein's) own solution consist of? In part 3 of his book, Kripke ascribes to Wittgenstein an agreement with the skeptic, viz. that "there is no fact as to whether I mean plus or quus" (Kripke 1982: 71). There is no "straight solution" to the skeptical paradox, then, but according to Kripke, there is a "skeptical solution" to the problem. According to this solution, a speaker considered in isolation cannot be understood as meaning anything by his words. But the situation is different once we understand the person as a part of a linguistic community. If a person learns to add — if he adds reliably —, he is admitted to the community, which shares a common practice of adding. The practice involves mutual criticism and correction. Meaning something by a term is understood in terms of the attitudes of other members of the community. This story shares many features with the account in Sellars (2007e). For a discussion, see Brandom (1994: ch. 1). See also §4.1 below.

[8]See Sellars (1980b). It is important to emphasize that Sellars identifies meaning with its use in the sense of "functional role", not with its "instrumental use". The latter is taken to be crucial by a rather different view, which has been called "agent semantics" (Rosenberg 1974: ch. 2). Sellars emphatically rejects the contention of agent semanticists such as Grice or Austin that what a speaker means by a term is a matter of his communicative intentions, i.e. of what he is trying to convey by using the word. See also O'Shea (2007: 199, n. 7).

a person means from a purely descriptive picture of the world. However, this is not the place to worry about problems deriving from a scientific naturalism.[9] Instead, we should proceed towards a better understanding of Sellars's thesis, which should help us draw the right lessons from Kripke's argument.[10] Sellars writes programmatically:

> (a) linguistic activity is, in a primary sense, conceptual behavior; (b) linguistic activity is through and through rule-governed. (Sellars 2007c: 61)

Starting with the second programmatic claim, to say that language is essentially rule-governed is to say that using language is something that speakers *do*. That is, it is to lay stress on the pragmatic aspect of speaking. Our use of languages consists of doings that are evaluable as appropriate or inappropriate, correct or incorrect, with respect to a certain standard. If the basic unit of thought is judgment, the paradigmatic use of language is in basic indicative assertions.[11] Thus in asserting an indicative sentence, a speaker does something that can be assessed as correct or incorrect. It is certainly possible to know the meaning of a term and use it wrongly, yet knowing the meaning entails grasping its role of usage at least to a certain extent. Take for instance a subject's assertion: "This vase is white." The appropriateness of this assertion depends, among other things, on the appropriateness of the use of the predicate "white" in this context. Meaningful expressions have conditions of correct use.[12] Again, as the

[9]The epithet "scientific" matters here, because normative facts are only problematic from the perspective of the view that what is natural is determined exclusively by the ontology of physics or similarly austere natural sciences. Some writers, notably McDowell (1996), have proposed a more liberal naturalism which makes room for *ought* as well as for *is*. See §1.4.

[10]Sellars is not so much a direct contributor to the debate started by Kripke as a writer that independently developed many of the same Wittgenstein-derived themes. Kripke's 1982 book did more than any other piece in the literature to bring the central importance of Wittgenstein's rule-following considerations to the forefront of philosophical consciousness. The ensuing discussion, some of which is collected in Miller and Wright (2002), and in particular the attempts to defend the dispositional approach to meaning, are insightful. However, a number of important distinctions have been neglected in much of the literature. Writing some three decades before Kripke's participation in the debate, Sellars offered a forceful defense of these distinctions (see in particular Sellars (1980b), Sellars (2007e), Sellars (2007d) and Sellars (2007c)). In doing so, he anticipated many of the difficulties encountered later by Kripke and others. In the text, I will make use of the Sellarsian distinctions between different types of rules, acts and actions, obeying rules and conforming to rules. The hope is that these additional tools contribute to making the rule-governed nature of meaning less mysterious. The most direct connection of Sellars's writings on rules and the Kripke-derived debate is found in Brandom's perspicuous treatment of the normativity of content thesis (Brandom 1994: ch. 1).

[11]Cf. Brandom (1994).

[12]Cf. Boghossian (2002: 148).

word "white" means *white*, we can suppose that it applies — i.e. correctly applies — to all white things and to none that are not white. That is, it is correct, or appropriate, to assert statements predicating whiteness of these things. Meaning statements such as "*X* means *y* by *z*" entail a large number of normative truths of this kind. As a consequence, by coming to mean something by a word, we acquire both obligations and liberties. For example, if I mean *white* by "white", I am obligated not to use the word to describe things I know are not white. By using the word, then, I take on a responsibility, but I also acquire an authority. Meaning *white* by the term "white" entitles me to claim "*X* is the color of pure snow" if I already claim "*X* is white".

For Kripke the relation between meaning and future action is not descriptive. When we are indicating what a person means to say by an expression, what we are specifying is not what linguistic behavior the person will, most likely, exhibit: we are fixing what linguistic behavior is appropriate, whether or not the person will likely exhibit it or not. Whether or not the agent conforms, or tends to conform, to the relevant linguistic rules is not material to the question what he says by using the expression. What matters are the semantic rules that constitute correct use of the expression.

ii) If semantic content is normative, then so is the content of intentional states: The normativity of meaning infects the psychological realm. On a plausible assumption, the idea that meaning is essentially normative also means that mental content is essentially normative. Here's Allan Gibbard's formulation of the Normativity of Content:

> The thesis that "mental content is normative" is this: that when I attribute mental content – when I say, for instance, that Ebenezer is thinking that he has lost his keys – I'm somehow speaking oughts. (Gibbard 2003: 85)

Just as ascriptions of meaning are characterized by normativity, so are ascriptions of intentional states. According to the Normativity of Content, for a state to have intentional content is for the owner of that state to be governed by various rules. This means that when we say that someone has, for instance, a belief, we say that he has a license to do some things and a responsibility to do others. However, at least one writer has suggested that the arguments put forward against the dispositional theory of meaning do not apply to mental content.[13] There are two possible grounds for doubt here. First, the extension of meaning skepticism to

[13]Thus Colin McGinn writes: "The issue of normativeness, the crucial issue for Kripke,

mental content rests on the assumption that mental content is, to some degree, analogous to meaning. However, it seems plausible that when we think, our mental activity involves some type of expressions or symbols that act as bearers of significance. This requires not a language of thought, perhaps, but the existence of events or mental acts that are parallel to the meanings of our assertions and their component expressions.[14]

Second, one may accept that the analogy between talk and thought exists but at the same time doubt the idea that mental meaning could be subject to the same kind of skepticism as linguistic meaning. But if we admit that mental doings analogous to linguistic acts are involved in thought, it is natural to infer that these episodes are open to rivaling interpretations or misunderstandings in the same way as their counterparts in public linguistic assertions. If these episodes are to be taken as significant, they must have meanings, and with respect to these meanings we can ask the same questions we asked about the meaning of public utterances; in particular, they must possess conditions of *correct* use.

It is hard to see how one could consistently hold that the meanings of a person's words are, in Sellars's phrase, fraught with ought while at the same time denying that our intentional states are as well.[15] To support

has no clear content in application to the language of thought: what does it mean to ask whether my current employment of a word in my language of thought (i.e. the exercise of a particular concept) is *correct* in the light of my earlier inner employment of that word? What kind of linguistic mistake is envisaged here?" (McGinn 1984: 147).

[14]The idea need not be that when we think, we think in Mentalese in the sense of a language with a syntactical and semantical structure resembling that of our own, or even with any analogue to the syntax of natural languages. Cf. Boghossian (2002: 144–6).

[15]Writing before the discussion begun by Kripke, Sellars, though he does not use the phrase "normativity of content", clearly is an early champion of this thesis. Thus he likens the attempt to reduce intentional states to purely natural occurrences or dispositions to similar — in his view failed — attempts in ethics when he writes that

> the idea that epistemic facts can be analyzed without remainder — even "in principle" — into non-epistemic facts, whether phenomenological or behavioral, public or private, with no matter how lavish a sprinkling of subjunctives and hypotheticals is, I believe, a radical mistake — a mistake of a piece with the so-called "naturalistic fallacy" in ethics. (Sellars 1963a: §5, 131)

In another well-known passage, Sellars talks about attributions of knowledge:

> [I]n characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says. (Sellars 1963a: §36, 169)

Sellars writes about "epistemic facts" and attributions of knowledge, but as McDowell (2009b: 6–8) has pointed out, we should not understand Sellars to be concerned only with epistemology in a narrow sense. Rather, his topic is intentionality in general, the objective purport of thoughts, beliefs and statements about the world. Hence "epistemic" facts should

this intuition, we need to turn to Sellars's first programmatic claim cited above, the claim that "linguistic activity is, in a primary sense, conceptual behavior". Sellars argues that we can understand contentful mental episodes on the model of assertions. The analogy between thought and talk finds expression in Sellars's view in the form of the doctrine he calls Verbal Behaviorism, or for short VB.[16] The core idea behind VB may be put as follows: although it may not be literally true that thinking is a kind of speaking, it is literally true that sometimes speaking is a kind of thinking. Thus consider the intentional states or episodes of believing. For Sellars, one way for a speaker to believe that snow is white is for the speaker to say aloud, in a public language, that snow is white. Not only that, but Sellars insists that "believing *p*" has as its *primary* sense saying aloud that *p*.

Sellars's view starts by focusing on a very basic type of language use. Sometimes we simply spontaneously utter a sentence, without giving the matter a lot of thought. For instance, when Ebenezer, on his way to work, suddenly says "I have lost my keys", he doesn't think beforehand what he intends to achieve with his utterance. Perhaps, alone in his car, he is speaking without an audience, so he doesn't take an audience into account. Now, as Sellars points out, we can understand this utterance itself as a kind of thinking. It is not just, as the classical Cartesian theory goes, a linguistic act caused by a thought but literally itself a piece of "conceptual behavior". We usually associate conceptual behavior with inner, silent goings-on, but we can easily understand Ebenezer's blurting out that he has lost his keys as a non-silent thinking. Thus for Sellars to "think out loud" that *p* is "'candidly and spontaneously uttering"p" 'where the person [...] who utters 'p' is doing so *as one who knows the language to which 'p' belongs*" (Sellars 2007c: 68).

According to Sellars, who also calls the thinkings-out-loud "languagings", such spontaneous pieces of minded linguistic behavior are real and, even more strongly, they form the *primary* sense of what we mean by "thinking". This idea is puzzling at first. Sometimes we do not express our opinions openly or publicly and prefer to keep them to ourselves. To say this, of course, is an understatement — if we always thought out loud whatever popped into our minds, our lives would be much too chatty, not to

be understood simply as concept-involving facts. Insofar as a description of an episode as a "knowing" is to place it in the logical space of reason, so is describing an episode as one of believing. Placing an episode in that logical space, of course, is to provide a normative characterization of that state, rather than merely giving an "empirical description".

[16]"Verbal Behaviorism" may not be the happiest name for the doctrine. Sellars is not committed to a version of behaviorism in the traditional sense. For instance, he does not propose an eliminative account of intentional states.

mention the social difficulties deriving from openly expressing our candid opinion about our fellow speakers.[17] In quantitative terms, we keep our opinions to ourselves much more frequently than we voice our opinions candidly. The out-loud type of thinking only occurs when we are, in Sellars's words, in a "thinking-out-loud frame of mind", which of course we only rarely are. In which sense, then, can languagings be properly called the *primary* form of thinking? The answer is that they are primary conceptually, in the order of philosophical explanation.[18] Sellars asks us to understand regular silent thinking on the model of candid languagings. Accordingly, in his view the concept of a belief depends on the concept of a linguistic act. Roughly, VB conceives thinking as, primarily, thinking out loud that p and, secondarily, having the proximate disposition to think out loud, though not actually thinking out loud, that p.[19] Belief that p, in turn, can be understood approximately as a settled disposition to think that p.[20]

VB is controversial and Sellars himself has reservations concerning the view, calling it a "radically over-simplified".[21] Of course, the picture of

[17]Cf. also Ryle: "The trick of talking to oneself in silence is acquired neither quickly nor without effort; and it is a necessary condition of our acquiring it that we should have previously learned to talk intelligently aloud and have heard and understood other people doing so" (Ryle 1963: 28).

[18]Though not necessarily in the order of *causal* explanation.

[19]As Sellars writes:

But if *thinking-out-loud* is the primary concept pertaining to conceptual episodes, not every concept of a conceptual *episode* is the concept of a thinking-out-loud. There is, in the second place, the concept of a proximate — a 'tip of the tongue' — propensity to think-something-out-loud. Such propensities amount to subjunctive conditionals as "If Jones were in a thinking-out-loud frame of mind, he would think-out-loud: 'the bus didn't stop'" (Sellars 1980a: 9).

[20]Cf. Sellars (2007c: section IV). The qualifications "roughly" and "approximately" are necessary because Sellars himself rarely gives definitions without quickly adding caveats or even taking back what he has said by calling it, say, a mistake to take it as capturing the meaning of the term defined (e.g. Sellars 2007c: 65).

[21]See Sellars (1975: lecture II, §9). In which sense does Sellars view VB as an oversimplified conception of thought? One sense in which VB deviates from the truth (though there may be others) is in its behavioristic reduction of thoughts to mere dispositions to utter sentences. Sellars often contrasts VB with the classical theory of mental activity, which conceives thoughts as inner mental episodes (e.g. Sellars 1980a: 7). On the classical theory, it is by virtue of being caused by these inner episodes that our linguistic performances have meaning. As we have seen, Sellars agrees with the Verbal Behaviorist in rejecting this view. However, he disagrees with the claim that, unless they are actualized in speech, thoughts are entirely dispositional in nature. Ultimately, Sellars is happy to countenance mental acts as occurrent inner conceptual episodes on the Cartesian model. Thoughts depend on thinkings-out-loud in a different way, viz. in that the *concept* of a silent mental act is a derivative concept that depends in the order of knowing on the notion of an overt languaging. What is more, the intentionality of thoughts is thoroughly dependent on the intentionality of linguistic performances (Sellars 2007c: p. 79). In these important respects,

thinking drawn by the view is only an approximation, but a very helpful one. It helps to answer the question, raised above, whether the normativity of semantical content extends to intentional content as well.

Assuming the truth of VB, we can answer this question in the affirmative without hesitation. According to the language-rule thesis, thinkings-out-loud, which amount to pieces of linguistic behavior, are regulated by rules and imply standards of correctness and oughts. From the perspective of VB, thinkings-out-lout are also genuinely mental occurrences, pieces of conceptual behavior. For on the model we assume, silent thoughts are languagings *sans* thinking-out-loud frame of mind. The salient point of difference — that when we think without speaking aloud, we are keeping our thoughts to ourselves — cannot be of consequence to the way the content of the state is fixed. But if the normativity of content thesis applied to public thoughts though not to private episodes, we would be forced to say that the contents of the two types of state differed fundamentally. When I say candidly "This chair is red", the content of this public thought — what it means — is clearly a matter of the meaning of the component words as I have learned them. Given Sellars's conception of belief, the same thing must be true for a belief of the same content, because beliefs are explained as a tendency to think out loud. If languagings are fraught with ought then silent thoughts and beliefs must be as well.

iii) Like beliefs, desires imply oughts: Ascribing a belief to a subject implies numerous normative truths about the agent. Beliefs are the paradigmatic intentional state, corresponding, as they do, to indicative assertions of complete propositions. What is true for beliefs, it seems, must also be true for desires, which, like beliefs, essentially have propositional content.[22] When we ascribe desires to an agent, we are "speaking in oughts".[23] However, the idea that desires are analogous to beliefs in this way has been rejected by some writers. Paul Boghossian doubts that there are "oughts about desires in virtue of the mere fact that they are contentful states" (Boghossian 2008: 102). He explains:

> Suppose I say of Ebenezer that he *wants* that Howard Dean be the next president. In making this attribution, am I in any way speaking oughts? [...] To be sure, Ebenezer's desire has

VB is basically correct.

[22]The grammatical object of desire may also be a thing rather than a proposition. Someone may want a flashy new car or a better job. However, we can understand this type of desire in terms of propositional desire. "I want a Fiat" implies "I want to have a Fiat", where the state of affairs desired is my owning that car.

[23]Cf. Akeel Bilgrami's argument that desire as well as belief is an "internal oughts", i.e. a "fully normative state" (Bilgrami 2006: 212).

> conditions of satisfaction — it will be satisfied if and only if Dean is the next president. But, in and of itself, this doesn't translate either into a correctness fact or into an ought of any kind. Of course, Ebenezer may have this particular desire because he believes it to be a way of securing satisfaction of another of his desires, and so his desire may be said to be correct to the extent that his belief is true. But that would be entirely a matter of the correctness of the underlying belief; it wouldn't introduce a sense in which the desire itself may be subject to normative evaluation. (Boghossian 2008: 103)

Boghossian draws a stark contrast between beliefs and desires. Attributing to Ebenezer the belief that he has lost his keys carries the implication that by having the belief, Ebenezer makes himself subject to normative assessments because, as Boghossian puts it, there is a correctness fact associated with the belief. For Boghossian, it seems, this fact amounts to the supposition that Ebenezer's state is correct in the sense of corresponding to how things really are in the world. On this view, his belief is correct if, and only if, it is in fact true that he has lost his keys. Boghossian identifies the beliefs's correctness conditions with its truth conditions. Correctness conditions come with an associated norm, and the norm in question is to believe what is true. On the other hand, Boghossian holds that there is no analogous norm governing Ebenezer's desire, as desires do not come with conditions of truth, being incapable of being either true or false.[24] As a candidate for conditions of correctness of desires truth conditions can be ruled out.

In the passage cited, Boghossian concedes that desires can be correct or incorrect, but on his view they are so evaluable only in an improper sense. He is appealing to the Humean idea, discussed earlier, that desires based on mistaken instrumental beliefs can be called irrational, though not *qua* desires: their incorrectness is inherited from the beliefs from which they are derived.[25] The idea is that correctness or incorrectness must be a matter of a state corresponding to, or failing to correspond to, a state of affairs. On such an understanding of correctness, it is no wonder that only beliefs and the portion of a desire derived from means-end beliefs allow assessment as to their correctness.

However, *pace* Boghossian, the fact that desires are not truth-apt does not entail that correctness considerations are inapplicable to desires. The

[24] Although of course it can be true, or false, that a person has a certain desire, that desire itself cannot be true or false.

[25] See §2.5.

reason is that the sorts of normative evaluation of states we commonly engage in are more diverse than he allows. Consider two ways in which we can criticize someone's intentional state:

- We can criticize a state for failing to conform to a *vertical* norm. In the case of belief, the criticism can be founded on the claim that a belief isn't true. A vertical norm pertains to the relation between a propositional attitude and the extramental reality.
- But intentional states are subject to *horizontal* norms as well. Horizontal norms pertain to the "intralinguistic" (or content-to-content) links, i.e. to links between propositional attitudes. To criticize a state along the horizontal dimension is to question the extent to which it is supported by reasons or to challenge its justifications.[26]

The second type of criticism can be illustrated by examples:

1. You assume that she is rich just because he is a doctor. But that doesn't follow. You ought not to believe that — the medical profession doesn't entail wealth.
2. Why don't you believe that copper conducts electricity? As you know, copper is a metal, so you ought to agree that it is an electrical conductor.
3. If you think that Denmark has a king, then you ought not to believe that it is a republic. A country cannot both be a republic and a monarchy.

These are genuine cases of normative assessment of belief, none of which amounts simply to pointing out that the belief in question is mistaken. Instead what we do is to point out that the states are not well supported by evidence, are incompatible with other states or follow from other states. In particular we criticize states based on their relations to other states. If we only think of truth or falsity, we disregard a whole dimension of our practices of evaluating belief as successful or unsuccessful, correct or incorrect. Beliefs are not just evaluated in the vertical dimension but in the horizontal direction as well. In praise or criticism, we regard a belief in relation not just to the vertical norm of truth, but to the norms of justification as well.

[26]For the distinction between vertical and horizontal norms, see Zangwill (1998: 194).

Broadly speaking, the horizontal norms that govern beliefs are inferential norms, norms that govern transitions from one state to another. They are rules that instruct the subject, who is in a given prior state S_1, to move to another state S_2. Because they have conceptual content, beliefs are subject to conceptual norms. The Sellarsian lesson we should draw from Kripke's observation that the relation between meaning and future action is normative is a broadly inferentialist understanding of intentional states: conceptual content is determined, essentially, through the rules by which uses of states with that content are governed.[27] The normativity of content is a reflection of the fact that conceptual content is inferentially articulated.[28]

When Boghossian doubts that there are correctness facts pertaining to desire, he objects to the idea that desire is governed by vertical norms. However, he does not take into account the existence of horizontal norms. If what we have just said is right, these inferential rules are crucial for the understanding of all contentful states. Just as belief is governed by rules relating to the inferential transitions it ought to take part in, so is desire. Desires are contentful states which imply, and are implied by, other contentful states. Some conceptual transactions involving desires are mandatory, whereas others are permitted or proscribed. Without their involvement in horizontal rules, desires would not be meaningful.

In contrasting desires with belief, then, Boghossian overlooks the important dimensions in which desires are subject to normative evaluation. The horizontal dimension is more than sufficient to buttress the idea that the normativity of content thesis is true for desires as well as for beliefs. We may still ask whether the correctness of desire is *also* a matter of vertical norms, a standard of correctness playing a role analogous to that of truth

[27]Inferentialism understands a state's conceptual content in terms of its inferential role. It is useful to distinguish, with Brandom, between a narrow and wide sense of "inferential articulation". On the narrow understanding, the inferential role is defined as the entirety of intralinguistic moves (where "linguistic" includes thinking as well as speaking) allowed, and disallowed, by the conceptual content, i.e. with the proprieties of inference properly speaking. The wide understanding also counts language-world transitions, i.e. moves that are not strictly speaking inferences such as observational judgments and volitions, among the moves whose appropriateness is determined by the conceptual content.

Armed with this contrast, we can further distinguish between weak inferentialism, strong inferentialism and hyperinferentialism (cf. Brandom 2000: 28). Weak inferentialism holds that one factor that determines the conceptual content of a state is its inferential relations to other states, although there may be other factors as well. Strong inferentialism holds that the conceptual content of a state is determined purely by the state's inferential relations widely conceived, whereas hyperinferentialism identifies conceptual content with inferential relations narrowly conceived. With Brandom, I accept strong inferentialism (see also Brandom 2007: 656–660).

[28]See Brandom (2000: Introduction).

for belief.[29] Yet we can already see that desire attributions imply numerous oughts in a way which does not depend on an answer to this further question.

3.3 The direction of fit of desires

The goal of this chapter is a better understanding of the intentional state of desire, the quintessential practical state of mind. We are inching closer and closer to such an understanding. The previous section introduced, as a topic, the contrast between belief and desire. As a final step in completing a high-brow account of desire, we need to ask what it is that distinguishes these two types of state. The question is pressing if we want to preserve a sense in which, even on a high-brow view, the belief-desire model of intentional action has a kernel of truth.[30] Classically, there have been two answers to this question:

a) The type of an intentional state is fixed by the state's direction of fit. Beliefs have a direction of fit such that the state must fit the world, whereas desires have a direction of fit such that the world must fit the state.

b) Whether a given intentional state is a desire or a belief is determined by the state's constitutive aim. While beliefs aim at truth, desires aim at the good.

As we have emphasized, High Brow is built on the idea (b) that to desire is to regard something as good, but as we have seen in the previous section, there are worries that desire may not have a constitutive goal or norm in the same way as belief. Let us therefore, following suggestion (a), start with Michael Smith's invocation of the notion of direction of fit in the service of an argument for Humeanism. The argument's goal is to buttress the Humean idea that agency requires both desire and belief. Before canvassing the argument, we should note that Smith does not take his argument to establish the truth of the claim that *normative* reasons are composed of belief-desire pairs.[31] Our focus is on the genuinely Humean idea that only desires truly give us normative reasons. Nonetheless, Smith's argument is interesting for our purposes because, so long as

[29] See the following section.

[30] See §2.2.

[31] Smith calls his theory of normative reasons an anti-Humean theory (1994: ch. 5).

normative reasons and motivating reasons are connected, our argument also touches on the assumption that motivating reasons necessarily involve desires. Furthermore, Smith's well-known argument throws light on the metaphor of direction of fit. His use of the metaphor will be our primary focus.

Smith argues that for an agent to have a motivating reason *just is* for him to have a desire. For him, this follows from three simple steps. Here is Smith's teleological argument:[32]

> What is it for someone to have a motivating reason? The Humean replies as follows. We understand what it is for someone to have a motivating reason at a time by thinking of him as, *inter alia*, having a goal at that time; the '*alia*' here includes having a conception of the means to attain the goal. But what kind of state is the having of a goal? It is a state with which *direction of fit*? Clearly, the having of a goal is a state *with which the world must fit*, rather than *vice versa*. Thus having a goal is being in a state with the direction of fit of a desire. But since all there is to being a desire is being a state with the appropriate direction of fit, it follows that having a goal just *is* desiring. (Smith 1987: 54)

A person who acts on a reason has a motivating reason. For Smith, ascribing a motivating reason to that person implies ascribing to him a goal. Having a goal is being in an intentional state. There are two types of intentional state, those having the mind-to-world direction of fit and those having the world-to-mind direction of fit. Choosing between these two options, having a goal must be for the agent to be in a state with world-to-mind direction of fit, which according to Smith simply amounts to having a desire. Hence, motivating reasons are desires.

Should we accept this miniature argument that motivating reasons must be desires? The answer depends on what we mean by "desire". If we understand desire as a high-brow state, the argument is unobjectionable. However, Smith himself defends a low-brow conception of desire. With its reliance on the idea of a direction of fit, his teleological argument seems to support the low-brow view. Smith sets up the condition of having a goal and then brings his antecedently developed dispositional theory of desire to bear. Reversing the argument, we can understand his remarks as establishing a job description for the state we identify as hav-

[32]The expression is due to Wallace (2006a).

ing a motivating reason. Whatever a motivating reason is, it should consist in having a goal and thus be such that the world should fit the state, rather than *vice versa*.

In supporting his identification of a complex disposition as the relevant state, Smith relies heavily on the notion of a direction of fit. The expression "direction of fit", of course, is only a metaphor.[33] To use it, we must provide a paraphrase of the idea behind it. What is it for an intentional state to have mind-to-world rather than world-to-mind direction of fit, or *vice versa*? In general, the idea is that the distinction sorts linguistic acts and intentional states into two categories. Let us start with Elizabeth Anscombe's characterization of the distinction in her book *Intention*, which served to introduce or revive the topic for contemporary philosophy of action.[34] The basic question is how we can distinguish, in the most basic terms, between beliefs (or belief-like states) and desires (or desire-like states). In her *Intention*, Anscombe writes:

> In some cases the facts are, so to speak, impugned for not being in accordance with the words, rather than *vice versa*. This is sometimes when I change my mind; but another case occurs when e.g. I write something other than I think I'm writing: as Theophrastus says (*Magna Moralia*, 1189b 22), the mistake here is one of performance, not of judgment. (Anscombe 2000: §2, 4–5)

Here Anscombe is thinking of a person expressing the thought that he will do something, perhaps by saying "I will write my name on this sheet of paper". That such a linguistic act may have two very different functions can be seen from the case where the person does not afterward write his name. On the one hand, the expression may be interpreted as a prediction about future behavior. In this case, there has been a mistake: the person misjudged his own behavior. Here it is appropriate to speak of a "mistake of judgment", the case where the words are "impugned for not being in accordance with the facts". On the other hand, the expression may be an expression of intention, and the person may write a different word on the paper than he intends, perhaps by writing "Kim" instead of "Jim". This, too, constitutes a mistake, but the mistake is of a different kind. The person hasn't made a false judgment of fact but has failed to realize an intention. Anscombe calls this a "mistake of performance". Here it is not

[33] As Smith points out himself (Smith 1987: 51).

[34] Anscombe did not herself use the term "direction of fit". As L. Humberstone (1992) points out, the term was introduced by Austin (1953). John Searle helped popularize the expression; see in particular Searle (1983).

the words which are at fault; it is the facts that are impugned for not being in accordance with the words.

Picking up the same theme again later in her book, Anscombe provides a more illuminating example:

> Imagine a man going round a town with a shopping list in his hand. Now it is clear that the relation of this list to the things he actually buys is one and the same whether his wife gave him the list or it is his own list; and that there is a different relation when a list is made by a detective following him about [...] What then is the [...] relation to what happens, in the order and the intention, which is not shared by the record? It is precisely this: if the list and the things the man actually buys do not agree, and if this and this alone constitutes a *mistake*, then the mistake is not in the list but in the man's performance [...] whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record. (Anscombe 2000: §32, 52)

Here Anscombe illustrates the two complementary directions of fit by two different roles a hand-written list can play depending on its interpretation. In both cases, the object of the list is the man's purchases. Perhaps the list reads "Smith buys sugar, flour and potatoes." However, the relation of the list to the items bought depends on the context of the creation of the list. If the list is meant as a detective's report of the items bought, it represents a truth about the things bought, or it does so if successful. By contrast if the list is intended as a shopping list, it does not represent but mandate that certain things be bought; its object is something to be made true. Again the difference becomes apparent when there is a mismatch between the shopping cart and the list. In each case there is a mistake, but the mistakes are of different types. Whereas in the first, mind-to-world case, the mistake lies in the detective's report or judgment, in the second, world-to-mind case, the mistake lies in the man's performance.

Anscombe's immediate topic in these passages is a linguistic expression such as an inscription on a piece of paper or a hand-written list. But clearly the same contrast exists between types of intentional states. On the one hand, desires, intentions, hopes or wishes all contain mandates of what is to be done. They differ in the way they mandate this: while an intention has immediate practical relevance, a wish may be far-fetched or avowedly completely impractical and is not thought of as something in

the power of the agent. Beliefs, assumptions, hypotheses and imaginings, on the other hand, are representations of how the world in fact is.

Now compare Smith's gloss of the direction-of-fit metaphor:

> a dispositional conception of desires enables us to cash in non-metaphorical terms, and therefore in turn finds support from, the metaphorical characterization of beliefs and desires in terms of their different directions of fit. For the difference between beliefs and desires in terms of direction of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that p and a desire that p on a perception with the content that not p: a belief that p tends to go out of existence in the presence of a perception with the content that not p, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that p. Thus, we might say, attributions of beliefs and desires require that different kinds of counterfactuals are true of the subject to whom they are attributed. We might say that this is what a difference in their direction of fit is. (Smith 1987: 115)

Smith explains the two complementary directions of fit in terms of a counterfactual question: What would become of the state in question if the subject perceived or otherwise realized that $\neg p$? Belief and desire differ in this respect. While in these hypothetical circumstances, the belief that p would go out of existence, in the same circumstances the desire that p would endure. The tendency to disappear when joined to the perception that p is what makes belief a state with mind-to-world direction of fit, while the tendency to endure despite the perception is what gives desire the world-to-mind direction of fit. Here again we are reminded of Russell's theory of desire, according to which a desire is a state that ceases to exist when the behavioral cycle comes to an end but endures while the searching behavior is continuing.[35] As we have already noted, Smith follows Russell in defending a dispositional theory of desire and beliefs. It is no wonder, then, that he also employs dispositional terms in cashing in the distinction of directions of fit.

In this respect Smith's theory differs fundamentally from Anscombe's. Both Smith and Anscombe explain the notion by pointing to the case

[35] See §2.4.

of a mismatch between the state's content (p) and (the perception of) a contradicting reality ($\neg p$). Anscombe emphasizes the normative aspect of the mismatch. The vocabulary she uses is normatively loaded. The facts or words are "impugned", so that, we might imagine, either of the two is blamed for the mismatch, and there is a *mistake*, either in the performance or in the judgment. In Anscombe's description, the shopping list and the report differ in the way in which it is appropriate to react when a mismatch is noticed — the former mandates going back to the grocery store, the second might warrant hiring a better detective.

By contrast, Smith explains the directions of fit metaphor by describing in neutral terms how the agent's intentional states *would* change in a counterfactual situation. His account is an attempt to purge the metaphor of its normative aspect. But by doing so, the notion loses an essential element. To see this, consider how Smith's theory fares with two examples based on his own suggestion of seeing the states relative to the perception that p:

1. *Belief*: Joe has the firm belief that his pants are aquamarine. At some point, he sees that they are the same color as something he knows is turquoise, but he doesn't let himself be influenced by the observation. Despite his realization, he does not drop his belief that his pants are aquamarine.

2. *Desire*: Jill desires to eat healthy. After a stressful day, she indulges in a greasy burger with fries. As she realizes that what she is eating is not healthy at all, she decides that she doesn't in fact want to do so anyway. Her observation causes her to drop her desire.[36]

In each of the two examples, the subject reacts to the perception that p in a way we would not expect her to under the assumption that she is rational. Now on Smith's view, a state has mind-to-world direction of fit just in case it tends to disappear when conjoined with the realization that $\neg p$ and world-to-mind direction otherwise. However, because his belief that his pants are aquamarine is so deeply entrenched, Joe's belief endures despite his realization that they have a different color. But then Smith must conclude that the direction of Joe's state is not, after all, mind-to-world but rather its inverse. Again, Smith holds that a state has world-to-mind direction if it tends to endure when conjoined with the realization that $\neg p$, whereas it has world-to-mind direction of fit if it tends to disappear when conjoined with that realizations. Jill's desire to eat healthy, however, happens to be so constituted that it goes away when she becomes

[36]For further discussion of this kind of upstream reasoning, see chapter 6.

aware that she is not eating something healthy. Smith is forced to say that her state is not in fact desire-like but belief-like.

But the verdicts about Joe and Jill are both clearly inadequate. The mere fact that the subject reacts against our expectations does not magically change the orientation of the state towards the world. In his defense, Smith may point out that the directions of fit imply dispositions only with the caveat of a *ceteris paribus* clause which allows for exceptions. On the other hand, as we have previously seen, a subject may have a disposition to make systematic mistakes.[37] The deviating reactions need not be one-off exceptions: Joe may have a chronic mental hangup about his pants and Jill about fast food. Another defense would modify the definition of the direction to add a clause explicitly allowing for cases like Joe's or Jill's:

> An agent's state S with content p has mind-to-world direction of fit if it tends to go out of existence when S perceives that $\neg p$ unless the agent is irrational.

This change would allow the view to accommodate the counter-examples. But it would do so at the cost of reintroducing the normative element that was banished in the dispositional formulation, for to say that someone tends to do something unless he is irrational is merely a different way of saying that he *ought rationally* to do so.[38] The direction of fit metaphor is inherently normative: desires or intentions are states with which the world should fit, whereas beliefs or predictions are states which should fit the world.[39]

We have seen that, if we are going to deploy the idea of a direction of fit to distinguish desire from belief, we should understand the notion in normative terms. Turning now to answer (b) distinguished above, Bernard Williams has proposed that the characteristic feature of belief is that it aims at the truth.[40] Correspondingly, we can understand desire in terms of its constitutive aim. In David Velleman's words, an "enterprise at which we can succeed or fail", such as believing or desiring, is

[37] See §2.

[38] In other words, removing the normative aspect from the direction of fit metaphor is a futile endeavor. To adapt a Sellarsian phrase, we can pitchfork normativity out of the door, but it may well return by the window.

[39] In a later article, Smith proposes a more sophisticated account of differences in direction of fit. On this account, beliefs and desires differ with respect to their counterfactual betting behavior (Smith 1998). However, the new account is still a dispositional account and thus subject to the same objection as in the text.

[40] Williams (1973).

one that "must have an object against which success or failure can be measured" (Velleman 2000b: 176). On the common conception championed by Williams, belief constitutively aims at truth, whereas desire constitutively aims at the good.

The proposal that desires aim at the good is sound. Intentional action — having a motivating reason — requires aiming at the good, and if we are to hold on to the idea that a motivating reason requires a desire-like state, then desire must aim at the good as well. However, Velleman argues that the proposal falls short since desire needs an object against which its success can be measured and, in his view, goodness cannot be such an object. Now if the aim of desire is the good, it doesn't seem to be related to it in the same way as belief is related to truth. Velleman distinguishes between the formal object of an enterprise and its substantive object. As he explains,

> The formal object of an enterprise is a goal stated solely in terms of, or in terms that depend on, the very concept of being the object of that enterprise. [...] Any enterprise that has a formal object must have a substantive object as well — that is, a goal that is not stated solely in terms that depend on the concept of being the object of that enterprise. (Velleman 2000b: 176)

As an example, he mentions the concept of a hunt. To say that a hunt constitutively aims at a quarry is merely to specify its formal object. There would be something essential missing from the concept if all we could say about its object was that it was the quarry — if, that is, no specification of the substantive object was forthcoming. According to Velleman, there can be an enterprise of hunting only if there is a particular object that transcends the activity in question.

Now for Velleman, to say that belief aims at truth is to specify its substantive standard of success, rather than merely a formal object.[41] On his view, to say that in believing we care about truth is more than to say that in believing we care about whatever it is that we pursue in believing; it is to provide a particular standard against which to measure the success of the endeavor. For Velleman, truth makes for an unproblematic substantive standard of success for belief, but we cannot say the same about the alleged standard of desire. If we can say that the constitutive

[41]Velleman writes: "The object of theoretical reasoning is to arrive at true belief; and since true belief needn't be defined in terms of success in theoretical reasoning, it constitutes a substantive rather than formal standard of success" (Velleman 2000b: p. 180–1).

aim of desire is the good, then, according to him, we still haven't fixed a particular object yet. According to Velleman, then, practical reasoning constitutively aims at the good, if it does, only in the formal sense in which a hunt constitutively aims at the quarry. If, as it surely is, figuring out what to desire is a non-vacuous activity, desire must have another, substantive standard.

Velleman's view involves a striking disanalogy between theoretical reasoning — forming beliefs — and practical reasoning — forming desires.[42] What makes a piece of theoretical reasoning, with the conclusion that something is true, a good instance of doxastic deliberation is a concern for what is true. In the case of practical reasoning, on the other hand, we need to look elsewhere for a substantive goal. Goodness can itself only be specified in terms of the activity of practical reasoning. If the good is just the best thing to do, the specification of the good is simply what results from a successful, appropriate process of practical reasoning. Alternatively, we can give a particular characterization of the good. Doing so requires making a particular normative judgment about what it is we seek in seeking the good. But we were looking for a way of distinguishing desires from beliefs on a conceptual basis, and surely a substantive, tendentious conception of the good can hardly be part of the concept of a desire.

In terms of the distinction introduced in the previous section, Velleman's criticism is that goodness does not function as a vertical norm of assessment in the way truth does. However, the idea that truth is a substantive norm is hardly uncontroversial. The idea that truth can function as a yardstick for belief seems to presuppose an understanding of truth as correspondence with reality. On such a view, ontologically speaking, beliefs are true in virtue of truthmakers, facts which we can appeal to as furnishing an independent point of reference. However, on many other conceptions of truth — such as deflationist or pragmatist views — such an independent yardstick is not available.[43] As an example, deflationist views see the concept of truth simply as a grammatical device that allows disquotation and making certain sorts of generalization which would not be possible without it. But if we take a disquotational view on the predicate "is good" as well as on the predicate "is true", the purported stark contrast between practical and doxastic deliberation disappears. The deflationists holds that no substantive concept of truth is available, and that none is needed, meaning that Velleman's argument that goodness com-

[42]Velleman later names the aim of being in conscious control of one's behavior as what is constitutive of practical reasoning (Velleman 2000b: 193).

[43]Cf. Horwich (1998), James (1907).

pares unfavorably to truth fails.

Here we need to remember that both belief and desire are governed essentially by horizontal norms which exist in the form of inferential rules. Velleman may be right that desire does not have a substantive goal, but he is wrong to assume that it needs one. He is thinking of vertical norms only, but the inferential norms governing desire give practical reasoning a point. The inferential norms, of course, are norms of reasoning, so the constitutive standards are defined in terms of the activity that is ruled by the standards. But it simply is not true that this makes the activity vacuous or pointless. On deflationist theories of truth the fact that we cannot state the aim of doxastic deliberation in terms not involving theoretical reasoning or inferring does not entail that theoretical arguments are empty. Similarly, in order to show that practical reasoning has content, we need not understand the aim of practical reasoning, as it were, from outside the practice of deliberation. Since we know these practices to be robust, we have little reason to suspect that the lack of an external aim renders them purposeless. Therefore it need not be seen as a difficulty that desire aims at the good only as a formal object.

The way desires are embedded in a network of inferential norms also gives us a way to distinguish them from beliefs and to illuminate the notion of direction of fit. What is characteristic of desire, or other states with the same direction of fit, is not its various dispositional relations but its normative relations — what the agent ought to say, think or do, i.e. both what he is permitted to do and what he is prohibited from doing. This leads us back to our gloss of the notion of directions of fit, which, as we have stated, should be construed in a normative way. In Smith's view, a subject's belief that p is a state that tends to go out of existence when the subject perceives that $\neg p$. Instead we should say that there is a horizontal inferential norm asking you to move from "I perceive that $\neg p$" to dropping the belief that p; but there is no such norm that asks you to drop your desire that p in the same circumstances. This corresponds to Smith's idea that a subject's desire that p should persist despite the perception that $\neg p$. Additionally, there is a requirement that if you have a desire that p, you should also develop the intention to φ, provided you believe that φ'ing helps to bring about p.[44]

[44]We should retain Anscombe's normative understanding of the direction of fit metaphor. Smith defends a functionalist theory of desires that explains them in terms of their causal role in a system of propositional attitudes. On this view, it is constitutive of desire that it plays a certain causal role in such a system. The upshot of the present section is that although we can retain the idea of an account of desire in terms of its functional role, we should prefer *normative* functionalism to the more common dispositional or causal functionalism (cf. Zangwill 1998). According to normative functionalism, the essential charac-

The distinction between belief and desire, then, is built on the entirely different inferential profiles of these states. In deviating from the rules, Joe and Jill are being irrational, but their intentional states are not switching their orientation to the world. The rules that govern these states — inferential in the wide or narrow sense — are rational rules. To desire is to evaluate an object or state of affairs as good in some way. The standard of success is internal to our rational practices, but the inferential rules are robust, so there is no need to worry that the enterprise of practical reasoning or intending would be vacuous.

3.4 Practical commitments

After making the case against Humeanism in the last chapter, this chapter has been devoted to developing a positive high-brow conception of agency. Progress has been slow and laborious, but in the process we have collected a number of threads that now allow us to formulate what it is for intentional action to aim at the good. It is time to pull these threads together.

As a result of the preceding discussion, we have understood the claim that the constitutive aim of *desire* is the good in terms of its involvement in a network of inferential norms. It may be said that this terminology is misleading. Calling a state as we have described it a desire means reconstruing a term that has been understood as low-brow in the philosophy of action as a high-brow state. As a result, the word "desire" carries distinctly Humean associations that may impede comprehension. Rather than breaking established usage, we could react to this difficulty by simply stopping to talk about desires entirely. Citing the inadequacy of low-brow states to rationalize behavior, we may conclude that there is nothing to the Humean claim that desires provide us with reasons or that desire is crucially involved in intentional agency.

However, as we have emphasized, the belief-desire model of action, the idea that there is at least a representing and mandating component of

teristic of desire — and of belief — lies in its conceptual content, a content which consists of proprieties of inference.

Causal functionalism emerged in the 1970s with the work of Armstrong, Putnam and Lewis. However, it may be argued that Sellars is one of the forefathers of functionalism. In essays such as Sellars (1963a) or Sellars (2007d), he advocates a conception of mental states where the identity of a state depends on its role in a system of relations. Sellars developed his view, which arguably can count as a normative functionalism, before the introduction of causal functionalism and may have influenced early writers on the topic. Lewis (2002) explicitly cites Sellars (1963a) as an influence.

action, has some merit. Doing away with the concept of desire entirely would leave a theoretic void in the practical realm. Although we have seen that mere functional states or qualitative feels do not play the role Humeans take them to play, there is room for a more adequate notion of a state to replace the notion of desire as the occupant of that role. A good word for this quintessential practical state is "practical commitment".[45] In what follows, I will conceive intentionally doing φ as constitutively involving a practical commitment, either to φ'ing itself or to some goal ψ whose attainment is (believed) to be advanced by φ'ing.[46]

In the course of the present chapter, we have already been implicitly creating a profile of the notion of a practical commitment. The notion must fulfill four desiderata:

1. Practical commitments are high-brow in the sense of aiming at the good. Being practically committed to p implies looking favorably on actions that have p as likely consequence.

2. The fact that practical commitments aim at the good is a matter of their subject being constrained by a profile of inferential norms characteristic of the world-to-mind direction of fit.

3. Owing to the way their contents are fixed by inferential norms, practical commitments are normative states, i.e. states with essentially normative content. As contentful states, they are, as Kripke's remarks have shown, fraught with ought. As a consequence, they cannot be reduced to mere dispositions.

4. As on the traditional belief-desire model, intentional action involves a pair of intentional states with complementary directions of fit, i.e. both a practical commitment and a doxastic commitment.

The notion of a doxastic commitment is the counterpart to a practical

[45]Bilgrami reaches a similar conclusion when he writes that "[d]esires too are commitments. If I desire something, say, that I should help the poor, then I am committed to doing various things, such as, say, giving money to charity or joining the communist party" (Bilgrami 2006: 215). See also Bilgrami (2004: 128).

[46]Often replacing a philosophical term is possible only partially. When engaging with a philosophical tradition, it is necessary to use the language of that tradition. For this reason, I will continue to use the word "desire" when delineating the position of a philosopher from the Humean tradition, and often understand it in a high-brow way. Likewise, unless the context makes it clear that the topic is low-brow states, talk of desires in the preceding sections should be interpreted as referring to practical commitments.

commitment, just as traditionally belief is the counterpart of desire.[47] Whereas practical commitments are commitments to a certain project, ultimately to acting in a particular way, a doxastic commitment is the commitment to a claim being true. The elementary, most straightforward way to undertake the doxastic commitment that p is to assert "p". A paradigmatic way to undertake a doxastic commitment is in response to a perceptual situation, in which you take some propositional content to be true, thereby committing yourself to its truth. Similarly, the elementary way to undertake the practical commitment to φ is to assert "I shall φ". A paradigmatic way to react to the undertaking of such a commitment is to respond by acting in such a way as to make it true that one φ's, by doing φ.[48]

That practical commitments are essentially normative states is part of what the word "commitment" is intended to signal clearly. When you adopt a normative state, you bind yourself to a project or claim. In this regard, practical commitments are similar to promises. It is helpful to understand commitments in contractual terms, but the analogy may be thought not to be apt. Making a commitment is unlike promising in some respects. In making a promise, the promisor is binding himself at some point in the future to render a certain service to a promisee. Necessarily, making a promise requires a predetermined audience to which one becomes accountable when the promise is broken. When the presidential candidate gives a speech promising to lower the taxes, the promise is made to his audience, the electorate. If, as President, she later fails to make good of the promise, the voters will hold her responsible for her inaction. The discursive commitments we are envisaging, by contrast, are not directed at a particular audience. There is no defined group of people to whom the person undertaking the commitment owes an explanation if things don't go as planned.

It would be closer to the truth to say that undertaking a commitment is making a promise to oneself. However, it isn't exactly clear what making a promise to oneself entails. Promises come with sanctions, but if one has only promised oneself, then one can easily evade the sanctions by absolving oneself of the promise. What negative consequences could there be to breaking a promise to oneself? Such a vow seems conceptu-

[47] Practical and doxastic commitments are the species of discursive commitment at the center of Brandom's scorekeeping model. For the introduction of this pair of notions, see Brandom (1994: 233). Interestingly, Brandom accords the notion of a doxastic commitment explanatory priority insofar as both practical and doxastic commitments are states for which reasons can be given and giving reasons for discursive commitments requires the ability to make assertions, i.e. expressing doxastic commitments.

[48] Cf. Brandom (1994: 8).

ally to require a split personality where promiser and promisee are kept separated, and it is not easy to see how such a separation would function for something so basic as a practical commitment.[49] Unlike breaking a promise, failing to meet one's practical commitment does not necessarily impact another speaker. Nor are there necessarily sanctions associated with it. Nonetheless, we clearly have an idea of what being committed to a certain *claim* amounts to in the case of doxastic commitments.

Committing oneself to making a propositional content true is to bind oneself by rules. The rules in question are conceptual rules. The idea is easily illustrated for doxastic commitments. For instance, if I believe, i.e. if I am committed to the claim, that dogs are mammals, then I bind myself by an inferential rule that mandates that if I believe that Fido is a dog, then I ought also to believe that Fido is a mammal, or that if I believe that Pluto is a dog, then I ought also to believe that Pluto is a mammal. Making an inference from a sentence of the first sort to the second is something I am obligated to do, given my commitment to the genus-species relationship between mammals and dogs.

The idea carries over to the practical case. Suppose I intend to marry the richest woman in town. If I discover that of all the unmarried women in town Mary is the richest — and otherwise eligible — then I ought to intend to marry her. By undertaking the original long-standing practical commitment about my future spouse, I have committed myself to making various inferences of this kind. As long as a belief remains in force, it exerts its normative influence, giving permissions to, and imposing obligations on, the speaker, and the same is true for practical commitments. Similarly a practical commitment remains in force while it exists. It is true that it is possible to withdraw a practical commitment without necessarily opening oneself to sanctions from others, but that doesn't imply that having a commitment, so long as one does not withdraw it, is normatively ineffective. We can change our minds about what to do but unless we do, the conceptual pressure a practical commitment exerts is genuine.[50]

[49]Cf. Davidson (2001b: 90). For a discussion of the disanalogies between promise and commitment, see Brandom (1994: 262–6).

[50]Korsgaard comes to a similar conclusion when she explains her Kantian conception of the instrumental principle. According to her, the central notion is that of "willing an end". To will an end is not the same as pursuing the end or as being disposed to act on the end. Willing the end is committing yourself to "take the means to this end". She writes:

> Committing to taking the means is what makes a difference between willing an end and merely wishing for it or wanting it or thinking that it would be nice if it were realized. (Korsgaard 2008: 65)

On Korsgaard's view, willing an end is an "essentially first-personal and normative act"

Finally, although we have introduced practical commitment as a substitute for desire, it is more accurate to say that it corresponds to the notion of intention in the traditional topology of mental states. Committing oneself to doing φ is to form the intention to do so. While everyone agrees that belief is the crucial state in the theoretical sphere, writers in the literature are split in two groups over the question whether desire or intention is the principal state in the practical realm. Philosophers in the first group identify desire as the crucial practical ingredient of intentional action.[51] Some writers have argued that we can entirely dispense with the notion of intention in favor of talking about desires, holding that talk about intentions can be reduced to talk about desires.[52] The second, minority view, on the other hand, takes intention to be the more important notion, or the more basic in the order of explanation. For adherents of this view, any person who acts necessarily has an intention that accompanies it, and the intention is causally or non-causally responsible for the action.[53] On this view, while talk about desires has its place, it is not the crucial ingredient in action. This view holds that intention, not desire, is the counterpart of belief.

As we are identifying practical commitments with intentions rather than desires, we are situated firmly in the latter, minority camp. On my view, the difference between desire and intention is roughly this. To have a desire that *p* is to take a positive view on the state of affairs *p*. To have such an attitude also entails looking favorably on actions that have *p* as a likely consequence but it doesn't necessarily entail preparedness to take those actions, even if one knows that they are likely to promote the action one desires. Someone who wants to have a vacation in Capri and knows that working a second job will help him realize this wish may nonetheless not be willing to take the job. Having a desire does not yet entail treating something as a goal, and not pursuing the goal is not irrational.

(Korsgaard 2008: 57). In describing the act, she employs Kantian terminology relating to laws and autonomy. The act of willing is to be conceived as the act of giving oneself a law, or as governing oneself. Willings, as opposed to mere dispositional desires, are constitutive of having a will and thus of agency. As she writes, willing an action requires one to "consciously pick up the reins" (Korsgaard 2008: 59).

[51]This large group notably includes defenders of the belief-desire model of action, including Davidson (2001a).

[52]In earlier writings, Davidson holds such a view. See Davidson (2001a: 8) for his view that "intention" is a syncategorematic word. Later in his career, Davidson changed his mind on this issue and came to hold that intentions — in particular future intentions — are indispensable, even if in his framework the notion of desire remains primary. Cf. Davidson (2001b).

[53]Important defenders of the priority of intention include Anscombe (2000), Sellars (1966) and Brandom (1994: ch. 4).

By contrast, an intention does entail such a preparedness.[54] If I intend to make it the case that p and I know that φ'ing would cause p to be true, then I thereby intend φ'ing — or, at least, I ought to have this intention. Desires, but not intentions, are optional in the sense that I can remain indifferent to my own desires without incurring irrationality, but I cannot in the same way remain rationally unmoved by my intentions.[55]

3.5 Psychologism and Humeanism reconsidered

To summarize our results, acting for a reason needs to be understood as aiming at the good, i.e. as the expression of a practical commitment which itself is subject to rational norms. What does this mean for our overarching inquiry into the nature of reasons for action? Let us begin with the question raised, in chapter 1, whether reasons are psychological states. We saw earlier that an adequate theory of reasons needs to show an acceptable relation between normative and motivating reasons.[56] Noticing the distinction between mental acts and their contents, we concluded that an adequate theory identifies motivating reasons with contents rather than mental acts. When you act for a reason, you act because of what you believe in the sense of the thing believ*ed* rather than the attitude of believ*ing*. The believing itself comes into view as a reason only when it is also a believ*ed*, as in what we called opaque cases.

On the present theory, a motivating reason is a practical or doxastic commitment. Like the word "belief", the word "commitment" admits an -ing/-ed disambiguation, although unlike the former, it doesn't wear this feature on its sleeve. Thus when we speak of someone's commitment, this may refer to the action or propositional content he is committed to, as

[54]Brandom suggests that different conative states may be more or less overtly committive. He writes that "desires and intentions are essentially, and not just accidentally, things appropriately assessed as to their success, in the sense of their fulfillment" and adds that "[e]ven less overtly committive intentional states such as conjectures, hopes, and wishes are contentful only insofar as they settle how things ought, according to them, to be" (Brandom 2001a: 588).

[55]Another way to distinguish desire from intention would be to say that desire, but not necessarily intention, implies that achieving the result desired is (potentially) mixed with an element of pleasure or enjoyment. Sellars, who takes intentions as the primary practical state, defines desires derivatively as "roughly, relatively long term dispositions to have thoughts of the form: "(Other things being equal) I shall do A (or bring it about that-p)'' where it is presupposed that doing A or contemplating the truth of 'p' is something one would enjoy" (Sellars 1973: 206). For a more long-winded discussion, see in particular Sellars (1966).

[56]See §1.4.

when we say that two people are committed to the same project. On the other hand, using the same word we may also refer to the person's act of undertaking a commitment, i.e. of endorsing a project or propositional content. If what we said earlier was right, to say that a persons's motivating reason is his commitment, this must mean that the reason is the content the person is committed to *as* something to which he is committed. In other words, the theory proposed involves a version of content-psychologism.

We can briefly illustrate the contrast with the help of the distinction, due to Brandom, between normative attitude and normative status.[57] Adopting a normative attitude towards a speaker in Brandom's sense is to take another speaker to have a certain normative status. Taking normative attitude is to assess the other speaker as to what he is committed or entitled to. A normative status, by contrast, is essentially something that is assigned by another speaker. Whereas normative statuses are the playing chips in what Brandom calls the game of giving and asking for reasons, acts of adopting normative attitudes constitute the moves in the game. Along this dimension, reasons for action are normative statuses rather than attitudes: it is something we essentially take other agents, as well as ourselves, to have. The distinction adds a social element to content-psychologism: having a reason is having a normative status which can be assigned by other speakers.

Turning to the main topic of this chapter, we have seen that Humeanism consists of two claims:

1. Desire-based reasons: Desires, and only desires, provide us with reasons.

2. Instrumentalism: The only way in which reason can be practical is in the form of means-end reasoning.

As we have pointed out, the truth of the first thesis depends on what exactly one understands by the term "desire". Humeans understand it along low-brow lines, chiefly as dispositions, leading to:

Sufficiency Having a desire to φ is sufficient for having a *prima facie* normative reason to φ.

On a low-brow conception, however, we have seen that it is not true that desires are the source of our reasons. The search has led us to a more

[57] Cf. Brandom (1994: 165–6).

adequate conception of the fundamental notion of the practical realm: practical commitments. On a low-brow conception *Sufficiency* turns out to be false, but what if, in the place of desires, we substitute the High Brow notion of practical commitments?

Sufficiency C Being practically committed to φ'ing is sufficient for having a *prima facie* normative reason to φ.

Our chief reason for denying the original sufficiency claim was that a person with a weird functional state would turn out, according to *Sufficiency*, to have a reason. We rejected this as absurd because Radio Man had only a weird disposition, which does not by itself justify his conduct. Based on the new formulation, we cannot derive this conclusion. According to *Sufficiency C*, having a practical commitment is sufficient for having a reason. But, as we portrayed him, Radio Man precisely is not committed to his actions or identified with them as an agent. We rejected the idea that he has a reason to turn on radios as absurd because he only had a weird functional state to show for it.

We argued that Radio Man's pure functional state is arbitrary and seemingly *ad hoc*; that it was not sufficiently integrated rationally; and that the action is virtually unintelligible because his desire is a mere element of his natural psychology. Things are different with a person who has formed a *bona fide* intention to turn on radios. Let us describe the situation in more detail. In the scenario envisaged, the person, when he is approaching a new stereo, is not just following the raw mindless goal of turning on the radio, he is evaluating this goal positively. How exactly should we understand this evaluation? Generally speaking, he must be regarding the activity as good in some sense or other, but the specifics depend on how we describe the case further. There is something the person takes to be a reason for operating the radios. Here are a number of possibilities:

1. The man derives enjoyment from turning on radios. In his experience, doing this has proved pleasant to him. Perhaps we find the fact that he enjoys the activity inexplicable, and perhaps he himself has no idea what about the action he likes. However, human nature being as it is, our predilections are often of the hard-to-explain type. There need not be a deep reason for his action. Similarly, it may be that for whatever inscrutable psychological reasons he is uneasy when not turning on radios on a regular basis. The goal of avoiding an unpleasant feeling is as good a motive as any.

2. The man is curious or values activities "outside the mainstream". We actually often act on similar reasons, as when we say: "I'll have the garlic flavor because I've never had garlic-flavored ice-cream before." or "Let's take the other route to the supermarket today, just because we can."

3. The man is simply used to doing it and he values following a routine or habit (Wittgenstein, when asked what he would like for dinner, reportedly said he did not care what he was served for dinner so long as it was the same thing every day). We may not agree that this is a good reason. However, whatever may be objected against such a justification doesn't stop people from reasoning in this way: "This is what I've always done; why change now?" Although this may not in fact be a good reason, the agent may regard it as such, and that is enough to make the action intelligible.[58]

Although diverse, these possibilities have something in common: they point out a way to justify or rationalize the action. Such a justification need not be particularly convincing. What is more, the degree to which these considerations ostensibly justify the action may be small. But it does portray the action in some way as making sense, in however minimal a sense. If we are to understand the Radio Man's action as an expression of his agency, permitting the use of intentional, i.e. rational-agent, explanation, there must be a shred of justification which shows something about the action that, in the eyes of the agent, makes sense. Of course it may be hard, or impossible, to discover this aspect from the outside. The agent may conceal his reasons. It is not part of this position that all reasons are open to public scrutiny. Even the agent himself may find it difficult accurately to assess his own reasoning after the fact or, in a psychologically complicated case involving inner obstacles, even at the time of his action. However, that does not mean that no such desirability characteristic exists. Unless such an aspect can, at least in principle, be found, the behavior is not amenable to intentional explanation.[59]

Any of the three possibilities suffice to support the idea that, when he is

[58]Here is how Thomas Scanlon describes the action of getting a cool drink when he is thirsty: "First, there is the unpleasant sensation of dryness in my mouth and throat. Also, there is the thought that a cool drink would relieve this sensation and, in general, feel good. I take this consideration, that drinking would feel good, to count in favor of drinking, and I am on the lookup for some cool drink." Scanlon concludes that "[i]t is this future good — the pleasure to be obtained by drinking — that makes it worth my while to look for water [...] The motivational work seems to be done by my taking this future pleasure to count in favor of drinking" (Scanlon 1998: 38).

[59]See §4.2 for Milton's Satan as a particularly difficult example.

operating stereos, the person is, in our sense, aiming at the good. On the proposed view, this is to say that his intention to turn on radios commits him practically to doing so. Although such an intention may be a long-term commitment, it need not be. While some practical commitments are held for a prolonged period, others may result from a spontaneous decision. Nor is it part of the view that a commitment must be the result of careful reflection, as suggested by Gauthier's conception of considered preferences.[60] While a practical commitment can be the outcome of a well thought-out argument, it can equally be arrived at in a haphazard way, as when one happens upon a good idea. Or even a bad idea — there is much room for irrational commitments, commitments that one would do better, from a rational standpoint, not to have.

Though practical commitments are sometimes — often, even — irrational, they are rational states in another sense. As contentful, rule-governed states, they are integrated in a system of inferential relationships. As Nick Zangwill puts it, "a propositional attitude is a node in a complex, crisscross network of rational relations" (Zangwill 1998: 197), and intentions or practical commitments aren't any different. Because of their rational integration, practical commitments do not strike us as *arbitrary* or *ad hoc* in the way the original Radio Man's weird functional state does. Practical commitments are by nature something the agent stands behind, as the agent has made up his mind that performing the action constitutes a good for him. The commitments are, in this sense, part of his identity as a person. As an expression of his will, they have normative authority. Thus whereas a mere functional state is insufficient to make sense of an action, a practical commitment has a stronger claim to succeed in rationalizing an action.

This is not to say that we must necessarily agree with an agent that the actions to which he has committed himself are worthwhile or the things he ought to be doing.[61] This would lead to the absurd conclusion that judging that one has a reason to do something automatically makes it true that one has such a reason. Like all judgments, an evaluative judgment can be mistaken.

It remains doubtful, in any case, that *Sufficiency C* is true as it stands. Counter-examples are not far to seek. For one thing, does an agent have a normative reason to φ if the path by which he reached the decision to φ started from a mistaken factual premise? Here, at least, it seems far from clear that it would correct to attribute to the agent a normative reason

[60] See §2.5.

[61] See the discussion of the bootstrapping problem in §6.1.

rather than a merely supposed reason.[62] Furthermore, we should also be wary of attributing a normative reasons — even a *prima facie* normative reason — to someone who acts out a morally vicious intention. Although the agent clearly has a motivating reason, he acts although he doesn't have a good reason to do so — or so we would be tempted to speak. Question about specifically moral reasons cannot be settled here. The point is merely that our practice of attributing normative reasons is compicated. Even in its adjusted formulation, then, the Humean claim that having a conative state suffices for having a normative reason is open to formidable challenges.[63]

Even if *Sufficiency C* should turn out to be true, however, that should not be seen as giving aid and comfort to defenders of Humeanism. In its adjusted form, it is not a particularly informative thesis since it does not impose any substantive constraints on reasons. Having a practical commitment is *defined*, through its rational links, in terms of its role in reasoning, so it is part of the meaning of the term "practical commitment" that practical commitments rationalize behavior. The result is a high-brow position that is no longer recognizable as a form of Humeanism. Practical commitments are the source of reasons only in the weakest possible sense. Practical commitments, as we have introduced them, are defined in terms of the inference licenses they involve. These licenses include *inter alia* the permission to transition from having the commitment to do φ to *actually doing* φ, or to doing something else which promotes the attainment of φ. Here, then, we can confer to the idea of having a reason to do φ the sense of having the license to form the intention to do φ and to realize this intention. Having a practical commitment is directly related to having a motivating reason. But someone's being practically committed does not *explain* having a reason any more than having a reason explains being practically committed; the two concepts are interdependent. Instead, what has been said supports the suggestion, mentioned earlier, that hav-

[62]For a discussion of this question see chapter 5. The view proposed there is not a version of Bernard Williams's existence internalism in any recognizable form. For a critique of existence internalism, which takes reasons to depend on motivational states in another sense, see §§5.2–3. As I argue in these sections, we can ascribe intentions and reasons consequentially, which decouples reasons from our avowed intentions.

[63]While Humeans hold that unmotivated desires are beyond the reach of rational criticism, High Brow makes no such assumption about intentions, so even if one has an intention, it may still itself be irrational. A *prima facie* reason generated from an irrational intention is unsound. What is more, the idea that aberrant intentions may produce further reasons through instrumental reasoning may be worrying. But the reasons generated by an aim are only as good as the aim they are derived from, which in the case of aberrant intentions is not very good. For further discussion of the bootstrapping problem, see §6.1 and §6.5.

ing a reason for action is an irreducible notion.[64] This result should be encouragement to defenders of the view that the notion of a reason for action is the elementary, fundamental notion in the philosophy of action.[65] On the present view, practical commitments cannot be separated from reasons for action, or *vice versa*. The concept of a reason for action is undefinable except in terms which presuppose the same notion.

[64] See §1.1.

[65] See e.g. Scanlon (1998: 17–8), Raz (1975).

Chapter 4

Defending High Brow

4.1 Is High Brow incoherent?

In the previous chapter, we have seen the contours of a high-brow conception of intentional agency. According to this theory, practical commitments aim at the good, and intentional action involves a practical commitment and thus an evaluative element. We have spelled out the notion of a practical commitment in terms of rational, inferential rules. Clearly, in order to count as being subject to rational, inferential rules, one needs to possess conceptual capacities. High Brow implies, then, that the capacity to act for a reason requires the possession of a faculty of rationality and that actualizing the capacity for intentional action involves the deployment of concepts.

Some philosophers find such a picture of agency appealing, whereas others are wary of High Brow conceptions of acting for a reason. On the one hand, we can find versions of a high-brow conception of agency throughout the history of philosophy, in writers ranging from Plato to Kant.[1] The idea is summed up in the Thomist thesis: "Quidquid appetitur, appetitur sub specie boni." We desire objects under the guise of the good. The long history of the view indicates there is a presumption in favor of its correctness. As Railton puts it,

[1] See Plato's *Meno* (77a–78a). Kant notes with approval an "old formular of the schools": "nihil appetimus, nisi sub ratione boni; nihil aversamur, nisi sub ratione mali" (2003: AA 59). See also Tenenbaum (2003).

> High Brow is a view with excellent pedigree, tracing its ancestry back to ancient Greece. According to High Brow, just as belief necessarily "aims at" the True, action necessarily "aims at" the Good. (Railton 1997: 302)[2]

A few paragraphs after delivering this praise, however, he cautions,

> High Brow is, however, highbrow. Many philosophers, in my experience, are not. (Railton 1997: 304)

According to a number of philosophers, there are difficulties with the idea that intentional action requires the involvement of conceptual capacities. They argue that as a theory of acting intentionally in general High Brow is too demanding. This chapter addresses two broad issues of this type:

1. *Is High Brow incoherent?* If you act intentionally, you have a practical commitment, and having a practical commitment means being subject to, and having one's conduct regulated by, conceptual rules. However, this thesis faces a fundamental objection. If acting in accordance with a rule is itself something that must be done intentionally, then acting on a practical commitment seems to involve another action which again requires a practical commitment. But then a vicious regress looms. According to the objection, it follows either that intentional action cannot occur *sub specie boni* or that our account of acting under the guise of the good in terms of conceptual rules must be faulty.

2. *Is High Brow too demanding?* According to High Brow, acting intentionally requires you to conceive of what you're doing as something good or worthwhile. But some creatures who are in principle incapable of making evaluative judgments still are often said to act on desires. Moreover, even mature rational agents sometimes act spontaneously or thoughtlessly, and when they do, they do not seem to engage in evaluative judgments. Accordingly, the objection runs, it cannot be true in general that intentional action must occur *sub specie boni*.

I begin the chapter by discussing the first set of worries about the viability of High Brow (§1); the remainder of the chapter is devoted to variants

[2]See §3.3 for the important caveat that the good is only a formal object. See also §6 below.

of the second worry. The next section considers counter-examples due to Michael Stocker against the principle that the good, and only the good, attracts us (§2). Next I turn to the problem that not all intentional actions are conceptually reflective (§3). After some general remarks about the structure of practical reasoning (§4), I make an attempt at solving the problem of unreflective agency while nonetheless maintaining High Brow (§5). Finally, I address the perennial question whether non-rational creatures, too, are capable of acting for reasons (§6).

Acting for a reason has been explained as aiming at the good, which in turn was explicated in terms of being subject to conceptual rules. This creates a puzzle for High Brow which may be described as follows. Start with the intuitive difference between obeying a rule in a strict sense and merely acting in conformity with the rule. Take McDowell's concise example of "someone following a marked trail, who at a crossing of paths goes to the right in response to a signpost pointing that way." (McDowell 2009a: 129) We can think of a signpost as a kind of rule, since, like a rule, it gives us instructions about what to do. The signpost's shape and inscription directs the agent the right. If, as McDowell describes him, the agent's taking the path on the right is a response to this rule, surely in his action he is obeying a rule. By contrast, the agent may walk to the right without understanding the sign or even completely at random. In that case, although the action would conform to the rule, the agent would not be obeying the rule. Genuinely obeying a rule involves more than mere conformity to a regularity.

Applying this distinction to our topic, we have explicated acting for a reason as a performance governed by at least one inferential rule.[3] One may reason as follows. Suppose I am intentionally φ'ing as a way of acting on my intention I to φ, and this performance is enjoined by the rule of inference R. My φ'ing, then, must be in "reaction", in McDowell's phrase, to this rule. Further, it clearly is not enough that my performance should merely conform to R. The performance must occur because of the rule, so more than merely coinciding with what R prescribes, it seems that my doing φ must be an instance of obeying R. The puzzle becomes evident when we add that, apparently, obeying the rule R is something one needs to do intentionally. In other words, it seems that in order to count as obeying R, I must have another intention I_2 to follow R. Now according to the proposal, I_2 must be rule-governed. It follows that the act, my obeying I, must be a reaction to a further rule R_2. Once again,

[3] I will simplfy things by supposing that only one inferential rule is involved. In reality, intentional states are governed by a mutidue of rules.

rather than merely doing what R_2 enjoins by accident, I must be obeying R_2 intentionally. But at this point it is clear that we are embarking on an infinite regress. For if doing what R_2 instructs me to do is an intentional action, I must have another intention I_3 to follow the rule, and so on.

According to the puzzle, our high-brow account, combined with a seemingly plausible account of what a reaction to a rule amounts to, implies that doing φ for a reason entails having an infinite number of intentions and following an infinite number of rules. But this is an absurd consequence. The objection concludes that High Brow must be fundamentally incoherent or that at least our way of spelling out the theory must be faulty.

As prerequisites to solving this puzzle, we need to put in place a number of Sellarsian distinctions. So we first need to consider a related problem raised by our presentation of Sellars's philosophy of language.[4] Recall that for Sellars, the linguistic meaning of a word is constituted by the rules of its use.[5] To use a word, say, the adjective "red" in the assertion "This is red", is to produce a performance that is subject to various meaning rules. In an early essay, not long after introducing his conception of meaning, Sellars quickly notices that it faces an objector who asks:

> Are you not confronted by a dilemma? For surely the rules for a linguistic system are themselves linguistic phenomena. Therefore either you must hold that they, in turn, are rule-governed, or else admit that at least one linguistic structure exists which is not "rule-governed" in your sense. You can scarcely be prepared to adopt the latter course. If you take the former, you are committed, surely, to an infinity of rules, meta-rules, meta-meta-rules, etc. (Sellars 1980b: 142, n. 5)

In this essay, Sellars hasn't worked out a satisfactory answer to this objector yet, but we can find an answer in his later seminal essay *Some Reflections on Language Games*. There, Sellars presents a restatement of the problem. He writes that if, as the language-rule thesis holds, knowing the meaning of a term is to know the rules of use that govern that term, then learning a new language must amount to learning the rules that govern the expressions of that language. Now suppose that a person learns the meaning of "red", i.e. acquires the concept of red. The thesis entails the existence of a rule of the form

[4] In §3.2.

[5] Sellars (1980b: 142).

> If you are in circumstances *C*, you ought to do *A*.

The natural form of a rule is a conditional sentence. Now if learning the use of "red" requires learning the rule, it seems that learning that requires learning the truth expressed by the statement. What sentences would that be in the specific case of "red"? Plausibly, among those sentences should be the rule:

> You ought to utter the words "This is red" only if the object you are pointing at is red.

This rule is arguably partly constitutive of the meaning of the word "red". But in order to understand the rule in question, you need to understand the meaning of its constituent words. It follows that to learn the rule, you already need to know what "red" means: you must already possess the concept of red. But learning the concept is what we were trying to explain in the first place. If acquiring a concept requires learning a rule, learning the rule cannot itself require already familiarity with the concept in question. More generally, in order to learn a language one already needs to know a metalanguage in which the rules for that language are formulated. But unless we want to assume that the knowledge of such a metalanguage could be innate, this starts a vicious regress: learning a language requires knowledge of a metalanguage, which in turn requires knowledge of a meta-metalanguage, and so on.

Sellars locates the problem in the way our assumption is formulated. We assumed that learning the concept *red* is coming to obey the rules of the term "red". For him, the first step in the right direction is to take note, as we have, of the two ways in which a person may relate to a rule. An agent who has the habit of doing *A* in *C* — in the example, saying "this is red" only when the object he points at really is red — conforms to the rule. All that is needed to conform to a rule is to do *A* in the required circumstances with some regularity. By contrast, obeying a rule is more demanding. As Sellars writes,

> whereas *obeying* rules involves using the language in which the rules are formulated, *conforming* to rules does not, so that whereas the thesis put in terms of *obeying* rules leads to a vicious regress, it ceases to do so once the above substitution is made. (Sellars 2007e: 29)

The suggestion is that for mastering the concept *red*, obeying the corresponding rule is not required, conformance to the rule being sufficient.

Only if the agent is obeying the rule does he need to have a rule-expression in his mind. The suggestion is initially appealing. Following Wittgenstein, Sellars likens speaking a language, with its assertions and other performances, to playing a game such as chess, with a number of available moves. Chess is governed by rules which must be learned in order to play it, but as Sellars points out, being a competent player of chess does not entail the ability to explicitly formulate the rules of the game. Thus Sellars develops the proposal:

> playing these games is a matter of *doing A when the circumstances are C, doing A′ when the circumstances are C′*, etc., and [...] the ability to formulate and obey the rule, although it may be a necessary condition of playing "in a critical and self-conscious manner" cannot be essential to playing *tout court*. (Sellars 2007e: 29)

Although seeing playing a game as mere conformance to its rules is attractive and contains an element of truth, the idea is ultimately unworkable. The problem is that this conception of playing a game portrays the rules of the game as "externally related" to the game. Sellars notes, but does not fully endorse, the following objection-cum-suggestion:

> surely one is not making a move in a game (however uncritically and unselfconsciously) unless one is making it *as a move in the game*, and does this not involve that the game be somehow "present to mind" in each move? (Sellars 2007e: 30)

In Sellars's view, the first half of this suggestion is true whereas the second half, the rhetorical question, misleads. The key to his understanding of rules lies in seeing how the first can be accepted without the second. Sellars explains the first and true part further when he notes with approval the idea that:

> learning a game involves learning to do what one does *because doing these things is making moves in the game* (Sellars 2007e: 32)

The trouble with a person who *merely* conforms to a rule is that although he goes through all the right moves, he doesn't make the moves *because of the rules of the game*, i.e. because they are the right rules. In relation to

the rules, his moves are accidental. But this runs counter to our intuition that the rule should be present in the activity, in a sense that makes it true to say that the agent acts because of the rules.

The rhetorical question in the second part of the quote offers a natural-sounding interpretation of the relation between the rules and the moves. On this proposal, the only way for one to make moves in the game because of the rules of the game is "in virtue of the fact that one made one's moves *in the light of* these demands and permissions, reasoned one's moves in terms of their place in the game as a whole" (Sellars 2007e: 32). But if this is not a viable option because it leads to the regress of meta-languages, the assumption that the agent must have the rules in mind in order for them to be present in the moves in order for them to be relevantly explanatory of the moves must be false. Thus for Sellars it is possible for the behavior of the subject to be governed by rules even though they are not explicitly present in the subject's mind. The fact that the agent does not, in doing what he does, explicitly intend to follow the rule-formulations does not entail that the behavior is accidental with respect to the rules:

> there can surely be an unintended relation of an act to a system of acts, which is nevertheless a necessary relation — a relation of such a kind that it is appropriate to say that the act occurred because of the place of that kind of act in the system. (Sellars 2007e: 32)

For Sellars, the distinction between the complex phenomenon of obeying a rule — implying an explicit intention — and the undemanding phenomenon of merely conforming to the rule — implying a lack of necessity — is a false dichotomy: there is a middle ground between mere conformity and obedience to a rule. For Sellars, what we need is the notion of pattern-governed behavior: behavior that happens because of the rule, in the appropriate sense, without the involvement of an explicit intention to follow the rule.[6]

Even if we can see pattern-governed behavior as a middle ground between mere regularity and explicit rule-following, for Sellars it is crucial for our understanding of language. The key to solving the problem of a subject's learning a language, then, is understanding pattern-governed behavior, which according to Sellars is

[6]Sellars (2007e: 34).

> the concept of behavior which exhibits a pattern, not because it is brought about with the intention that it exhibit this pattern, but because the propensity to emit behavior of the pattern has been selectively reinforced, and the propensity to emit behavior which does not conform to the pattern selectively extinguished. (Sellars 2007d: 87)

An agent's pattern-governed behavior, whose pattern is shaped by a system of linguistic rules, is in accordance with the rules of the system. In other words, like a creature that merely conforms to a set of rules, the agent is disposed to do what the rules mandate without thereby deploying an intention to do so. The difference from *mere* conformity is not in what the agent does but in how he has acquired the relevant propensities. Here the contrast to rule-obeying behavior is useful. A teacher may teach a chess novice the rules of chess by giving explicit instructions: "You ought to move the rook in a straight line, you are not allowed to castle after having moved the king, and so on". Other than giving instructions, however, it is also possible to bring about the relevant behavior through selective reinforcement. The chess teacher may let his students play and encourage the moves that comply with the rules while criticizing moves that don't. As Ryle points out, in order for a player to know chess, what is important is what he does on the board.[7] If the lessons succeed, the student's behavior not just conforms to the rules of chess but actually becomes a piece of pattern-governed behavior.

In practice, learning chess involves a mixture of explicitly memorizing rules and simply acquiring habits through reinforcement and extinction that is not in propositionally explicit form. Of course, in chess training, selective reinforcement usually is verbal; and chess is a game that students learn after having learned a language. If we take Wittgenstein's metaphor of a language game seriously, then to learn one's first language is to learn a special sort of game, one can play without already being able to play any language-like games. Of course, a school-level child acquires its linguistic abilities partly through explicit instruction, e.g. in the form of grammatical prescriptions. But on Sellars's view, at its most basic level, the acquisition of a first language involves only learning through (friendly) drill or conditioning. The methods employed by the teachers, vastly different from their equivalents in chess, can be summed up in the expression "Mommies and daddies, frowns and smiles": if the child says things — makes moves in the language game — that are correct according to the rules governing the expressions, it is encouraged, while it receives

[7] Cf. Ryle's lucid description of learning chess in his Ryle (1963: ch. 1, 40–44).

a negative reaction when its moves are incorrect. Using a variety of techniques, the teachers get the children to exhibit good uniformities in their behavior. At the end of the process, the child's verbal behavior is pattern-governed in Sellars's sense. No child pauses to think of linguistic rules, let alone grammatical rules, in deciding what to say. Nonetheless, its behavior is, in a perfectly intelligible sense of the word, governed by those rules.

Two related objections to this conception arise naturally. First, how can you "follow" a rule — and be guided by the rule — without explicitly using that rule as a premise in your reasoning? Second, even if this is possible, in which way is the rule really present in the activity if it involves no explicit grasping of a proposition? The first problem can be put in the following way. It is often assumed that the concept of a rule is the concept of something that prescribes a certain course of action in a specified set of circumstances. A rule is, or is expressed by, a general *ought*-statement of the form:

Ought-to-do If you are in conditions *C*, you ought to do *A*.

What constitutes compliance with a rule? A subject that is subject to the rule conforms to it when he does *A* when he is actually in conditions *C*. On a popular conception, this means that the subject needs to become aware that he is in these circumstances and, as a consequence of this belief, decide to do *A*. This obviously requires the subject following the rule to have considerable conceptual capacities. He needs to be able to recognize that he is in the relevant conditions mentioned in the antecedent and he needs to know what constitutes compliance with the response mandated in the consequent. It is, to say the least, unlikely that we can presuppose these kinds of capacities in an agent when we are trying to explain what it is for the agent to be able to speak the language — when we are giving an account of what his conceptual capacities amount to in the first place. What is more, if we take one of the relevant rules to be the aforementioned:

> You ought to utter the words "This is red" only if the object you are pointing at is red.

then understanding the consequent requires that the agent already be in possession of the concept of redness, the concept whose mastery we are hoping to explain. If we must interpret rules along these lines, they cannot play the role Sellars gives them — the role of supplying a basic explanation of our ability to speak meaningfully.

As Sellars points out, however, this line of thought relies on an unwarranted assumption about rules. It assumes that semantic rules need to be, as Sellars calls them, rules of action. If a rule of action is involved,

> for the rule itself to play a role in bringing about the conformity of 'is' to 'ought', the agents in question must conceive of [action] A *as* what ought to be done in circumstances C. This requires that they have the concept of what it is for an action to be called for by a certain circumstance. (Sellars 2007c: 59)[8]

A rule of this type cannot form the basis of our mastery of language, so if there were only rules of action, the language-rule thesis would be unworkable. Fortunately, there is a second type of rules which has the following form:

Ought-to-be *X*s ought to be in state *F* whenever such and such is the case.[9]

Prescriptions of this type, which Sellars calls rules of criticism, specify not what the agent ought to do but in what state something — in our case a subject — ought to be.[10] For this reason, rules of criticism are also known as ought-to-be's, to contrast with rules of action or ought-to-do's. There are three differences between the two types of rules:

1. Whereas ought-to-do's, which require conceptually structured recognitional capacities on the part of its subject, ought-to-be's can, in principle, apply to inanimate objects as well as to persons. Here is an example:

   > Clock chimes ought to strike on the quarter hour.[11]

   Although the verb used ("strike") has the form of an action verb, this rule is an ought-to-be since it specifies what should be the case

[8] I have corrected a misprint in the Brandom-Scharp volume.

[9] Sellars (2007c: 59). The distinction between ought-to-be's and ought-to-do's is similar to LLoyd Humberstone's distinction between situational oughts and agent-implicating oughts (I. L. Humberstone 1971).

[10] Sellars's talk of "being in a state" should be understood in a wide sense that includes a creature's behavior as well as its psychological states. Bear in mind that Sellars's topic is learning and teaching of conceptual behavior. A creature's doings — raising an arm, say, or producing language-like sounds — are subject to ought-to-be's. The relevant contrast is between mere trained behavior, which is in principle available to non-agents, and intentional action, which involves more stringent cognitive requirements.

[11] I take this example and the following from Sellars (2007c: 59–61).

— the clock's striking — rather than what someone should intentionally do. Not being a person, of course, the clock cannot do anything intentionally. *A fortiori*, it does not have any conceptual capacities to recognize situations and adjust its behavior.

2. Ought-to-be's are more interesting for our purposes when applied to living creatures rather than things. However, even when applying to persons, an ought-to-be need to involve an awareness of the conditions of application. Thus the rule

   One ought to feel sympathy for bereaved people.

   is a rule of criticism that only applies to persons, but here, too, the predicate "feel sympathy" does not refer to an action. Like a clock's striking, it is not something that one can do intentionally or for a reason.[12] In Sellars's terminology, feeling sympathy is an act rather than an action. Actions are actualities which "appropriately can be spoken of as deliberate or impulsive, carefully or careless" (Sellars 1980a: 8). They done intentionally and for a reason. A feeling such as sympathy is the actualization of a potentiality, but it is "non-action", and although there may be reasons for performing an act, a subject does not perform an act with a reason in mind. Even when the subject of an ought-to-be is a person, the rule refers to acts rather than actions.

   The difference between action and acts is important in particular in the context of conceptual or linguistic activity, which for Sellars to a surprising extent consists of acts rather than actions. There are of course linguistic actions, including baptizing a ship, proclaiming victory or making a birthday call. In doing these things, we are hoping to achieve a certain effect. On the other hand, not all pieces of verbal behavior are properly classified as linguistic actions. Sellars emphasizes that what he calls "languagings" or "thinkings-out-loud" are not actions that we decide to perform as a result of practical reasoning.[13] Similarly, undertaking a commitment is a linguistic act, a doing, but not an intentional action. Importantly, while ought-to-do's pertain to actions, ought-to-be's can refer to mere acts, including linguistic acts, as well. As we are envisaging things,

[12]Feeling an emotion is not something one can do at will, though perhaps one could do something that helps bring it about that one feels sympathy. This would, however, not amount to intentionally feeling sympathy but changing one's own habit through self-criticism.

[13]See §3.2.

saying "this is red" in response to a red object should be thought of as a languaging and hence as an act rather than an action. Therefore the subject can produce the performance in response to a rule without forming an intention to do so.

3. In order to conform to the sympathy rule, the subject arguably needs to have the concept of someone bereaved. Yet it is important to note that having a concept of the circumstances is not necessary for ought-to-be's in general. Applied to animal training, we can imagine a rule

   > These rats ought to be in state *F* whenever *C*.

   Through conditioning, we could bring it about that an animal conforms to this ought-to-be rule. In order for such training to succeed, it would not be necessary for the animal to have a concept of circumstances *C*. Instead it would be sufficient if the animal had the ability to "respond differentially to cues emanating from C" (Sellars 2007c: 60). As a result of the training, the rat's behavior conforms to the rule in question. Similarly, there is nothing mysterious in assuming that a person could have his behavior governed by a rule of this sort without presupposing sophisticated conceptual capacities on his part.

Ought-to-be's are different from ought-to-do's in that they involve acts rather than intentional actions and that reliable differential responsive dispositions[14] suffice for having one's behavior governed by them. For Sellars, however, there is an important link between the two: ought-to-be's "point beyond themselves" (Sellars 2007c: 59). In particular, every rule of criticism points to a related rule of action. Thus,

> These rats ought to be in state *F* whenever *C*

points to

> One ought to bring it about that rats in state *C* are in state *F*,

and

> *S* ought to utter the words "This is red" only if the object he points at is red

[14] Brandom (1994) ch. 4.

points to

> One ought to bring it about that *S* utters the words "This is red" only if the object he points at is red.

Bringing it about that someone exhibits behavior in accordance with an ought-to-be is something one does intentionally, and to attain such a goal, one needs to engage in practical reasoning, determining the best means for realizing the pedagogic end. But crucially, these rules of actions are directed, not at the subject of the behavior in question — the person learning the meaning of "red" — but to the one attempting to make his behavior conform to the ought-to-be rule — to the language trainer.

We are now in a position to answer the objections. An ought-to-be is less demanding than an ought-to-do because unlike the latter, it requires no explicit process of reasoning taking account of a rule statement. The solution to the puzzle is to conceive of the indispensable ground-floor rules that govern our use of language as ought-to-be's rather than ought-to-do's.[15] As to the sense in which these rules are present in the rule-governed activity, the difference between pattern-governed behavior and mere rule-conformity lies in the history of the acquisition of the patterns in question. We understand linguistic behavior as rule-governed in the sense of being pattern-governed because the pattern was instilled in the subject by its trainers when it learned the language. We can now explain this in greater detail. In learning a first language, a child comes to conform to ought-to-be rules. It acquires the uniformities of behavior it acquires because its trainers selectively reinforce correct behavior and extinguish incorrect behavior. The child itself does not mentally come into contact with expressions of the rules in question, but the trainers do. In performing their task, the trainers consciously think of the rules of criticism they wish to bring their students to conform to. When they perform the educational tasks they do, i.e. when they selectively smile and frown, they do so because these encouragements are ways of following the pedagogic rules of action ("one ought to bring it about…"), the rules which the ground-level rules of criticism point to.

The two distinctions, between act and action and between trainer and trainee, enable us to see what is true and what is misleading about the idea that if moves in a game are made "as moves in the game", they must

[15] I use the word "ground-floor rules" because, as Sellars points out, some language rules are ought-to-do rules (Sellars 2007c: 61). The important point is that, for the reasons outlined in the text, *all* linguistic rules could not possibly be ought-to-do's.

involve explicit awareness of a rule-statement. It is true that the statement of the ought-to-be has a function in the explanation of the rule-governed behavior, but that function does not amount to the subject's explicitly gasping a proposition while producing his performance. Instead, it is the language-teacher who while teaching that rule explicitly has the statement in mind, reasoning practically on its basis. When a competent subject says "This is red", his move is not accidental: he makes a move because of the ought-to-be rule that governs the expression. But the significance of this "because" is not that the rule occurs in his reasoning in any way. It means, more indirectly, that the performance is part of a system of behavior shaped by a pattern which was acquired through the actions of a trainer who, at last, explicitly envisaged the rule-statement.

Returning to the original question, the same set of distinctions shows that High Brow does not involve itself in an infinite regress. When a person intentionally raises his arm in order to φ, he is thereby expressing his intention or practical commitment, for example to hail a cab. It is true that his raising his arm counts as expressing the commitment only if in doing so he is being guided by the rules governing the intention. But the objection incorrectly assumes that the only way his activity could be guided by rules is for him to envisage the rules explicitly. It overlooks the possibility that his expression of his commitment is a piece of pattern-governed behavior. It is appropriate to say that he is doing what he does *because of the rules*, not because he is explicit thinking of any rule-expressions — he isn't —, but because he has acquired the behavior-pattern corresponding to the rule through the conditioning of a teacher who did explicitly envisage the rules.

Relatedly, the agent's forming his intention should not be thought of as another intentional action, which would give rise to a series of further intentions. Instead it is a mental act which is not preceded by any further intention.[16] Again, the agent's letting himself be guided by a conceptual role is not itself an action that should send us looking for a further intention. Acknowledging a practical commitment does not amount to consciously following any instructions. While the rules shape the subject's behavior, as rules of criticism they stay in the background. Yet it is perfectly legitimate to say that practical commitments are through and through rule-governed. The agent is not "following a rule", an ought-to-do, in the sense of having a corresponding intention, but his conceptual

[16]Of course, if the act of forming an intention, or undertaking a commitment, is an *act* rather than an action, something that can only be the subject of ought-to-be's, what is envisaged in the intention is an *action* rather than an act, something that is called for by ought-to-do's.

behavior is nonetheless guided by the rule, an ought-to-be. Undertaking a commitment is making a move in the game *as a move in the game*, with its demands and permissions, because those demands and permissions played an active role while the behavior was acquired. According to the high-brow thesis, that intentions or practical commitments aim at the good is constituted by their being subject to a network of ought-to-be rules. Therefore, performing an intentional action does not generate an infinite series of intentions or rules, and we have no reason to suspect that High Brow is incoherent.

4.2 Must we act under the guise of the good?

Some philosophers think that High Brow is too high-brow or overly intellectual. The worry that the view makes unrealistic demands on the agent encompasses two separable objections:

1. Someone may object that in stipulating that intentional agents necessarily act under the guise of the good, we exclude non-rational animals and human infants from membership in the class of intentional agents. According to the objection, we regularly ascribe intentional actions to those creatures and High Brow cannot account for these ascriptions.

2. Granting that only mature rational beings are capable of intentional agency, High Brow may still be thought dubious even with respect to creatures who can be assumed to have rational capacities. On this second objection, it is true that our actions are sometimes expressions of, or the upshot of, evaluative judgments, but in these cases we act reflectively and with a high degree of self-awareness. However, not all our actions are accompanied by an assessment of the action as something good or desirable. In particular when an agent acts without deliberating in advance or reflecting on the merits of the action, it is inappropriate to ascribe to him a deployment of the concept of goodness.

Deferring discussion of the first, more radical complaint until §6, I will begin addressing the second by considering an argument against the Guise of the Good by Michael Stocker. On the high-brow view, the Guise of the Good is a conceptual truth, but Stocker disagrees. He holds not only that the claim that "the good attracts" cannot be conceptually true but also that there are counter-examples that show the generalization to be mistaken.

According to Stocker, the view that we act *sub specie boni* exemplifies an overly optimistic view of human psychology. Of course it is no part of any High Brow position that everything every agent does is good. It is only that committing themselves to an action, an agent takes its outcome to be in some way desirable, and it is of course always possible that he may be wrong.

Nonetheless, the argument runs, the Guise of the Good makes the unwarranted assumption of a straightforward relationship between evaluation and motivation.[17] Thus Stocker argues that it is entirely possible to remain unattracted by a perceived good. He describes a man who used to be active in politics because he cared about the suffering of the members of his community. After years of public service he has grown disenchanted with the nature of politics. Although he still remembers his past devotion to the needs of his community he is now bitter and has stopped caring altogether about the well-being of these people. He still sees what is good for them but his bitterness has made him indifferent to their good. Because the man has become disaffected, a psychic state Stocker calls a "malady of the spirit", he is no longer motivated by his judgments about what would be good for the persons he used to care for. In such a spirit, "one sees all the good to be won or saved and one lacks the will, interest,

[17]Hursthouse (1991) argues against the Guise of the Good by providing a "recalcitrant set of counterexamples". For instance, she points out that an action may be "explained by a way of love, affection, or tenderness — kissing or lightly touching in passing, seizing and tossing up in the air, rumpling the hair of, or generally messing up the person or animal one loves; talking to her photograph as one passes, kissing it" (Hursthouse 1991: 58). This action, she argues, is in no way rationalized by a belief-desire pair in the classical sense. In particular, although the action is an expression of an emotion, the agent is not purposefully expressing an emotion. Hursthouse calls actions of this type "arational" rather than "irrational": they are done actions for which no reasons can be found.

It is clear that, if the examples are real, they cannot count as the agent acting *sub specie boni*. As I see it, however, Hursthouse's cases, though interesting, do not constitute a fundamental objection to the Guise of the Good thesis. According to the Guise of the God, *intentional action* aims at the good. We have identified intentional action with acting for a reason. The fact that an act done out of pure affection is done for no reason at all, then, does not show (as Stocker urges) that intentional action may occur without an evaluative element. Rather, what the examples suggest — and what is surely true — is that we need to make room for the fact that not everything we do is an expression of intentional agency.

Hursthouse objects that the "significant class of actions" she points to are intentional actions because they are not unintentional (Hursthouse 1991: 64). But unlike her, I do not think it is implausible to say that there is a class of human doings that are neither intentional nor unintentional, given that calling an action unintentional usually implies that it is intentional under another description. What is more, although her cases are certainly significant, it is also true that they are at the periphery of purposeful human behavior. Rather than building a theory of the behavior of rational beings on a "lowest common denominator" which is present in arational doings as well as in intentional doings in the narrow sense, we should focus our attention on paradigmatic cases. See also §6 below.

desire or strength" (Stocker 1979: 744).

For Stocker, the examples shows that "something can be good and one can believe it to be good without being in a mood or having an interest or energy structure which inclines one to seek or even desire it" (Stocker 1979: 745). Stocker emphasizes that a mood of this kind can intervene between an evaluative judgment and attraction, leading to a more complex relation between evaluation and motivation. This phenomenon, commonly called *accidie*, impedes our ability to be moved by normative judgments. Stocker is right to underline the possibility of *accidie*. It is a big leap, however, from this observation to the claim that this reveals believing in the Guise of the Good to be optimistic or "unjustifiable" (Stocker 1979: 749). It is true that the Guise of the Good posits a close relation between regarding an action as good and acting accordingly. But it does not entail that all our judgments about the good necessarily become effective in action. Because this would be an implausibly strong judgment-internalism about action, we should instead defend a weaker, more defensible version of judgment-internalism according to which we are motivated by our evaluative judgment only on condition of being rational.[18]

According to the Guise of the Good, cases of *accidie* should be understood as instances of irrationality. Does this mean that examples like the bitter politician pose a problem for high-brow conceptions of agency? It would certainly pose such a problem if High Brow had no way of coping with the phenomenon of weak-willed agents. However, there is little reason to think that this is so. As Stocker rightly says, the influence of the intermediary psychic states on our motivation is complex. Weakness of the will is a murky topic with many subtleties which a complete account would have to address in detail. As such a detailed account is clearly outside the scope of our current topic, however, an outline of how a high-brow conception can explain motivational inefficiencies in agents will have to suffice. Thus note that the politician is unusual in that, although he sees that a state of affairs — helping the members of his community — is good, this positive evaluation does not influence his actions. In which sense is that outcome good? Clearly fulfilling the needs of the community is good for those people, but if that was all that could be said, there would be nothing unsurprising about the man's indifference. There is nothing surprising about a person being unattracted to what is good relative to a certain group — perhaps good for the Bulgarian Association of Pharmacists. If the case is to present an instance of *accidie*, we need to assume that the politician not only regards the actions in question as good for the community but also, for that very reason, considers the actions good *tout*

[18] See §1.5.

court. If he makes an all-out judgment that helping the needy of his city is good, not relative to this or that group or interest, but *simpliciter*, then we would expect him to let himself be influenced by this thought.

What the psychic state of disillusion gets in the middle of, then, is the politician's ability to respond appropriately to some of his all-out evaluative judgments. His appropriate response would be to move from the judgment about the needy to forming an intention. In which way is such a response appropriate? The view we propose conceives all-out evaluative judgments as practical commitments. Commitments, as we have insisted, are normative rather than purely dispositional states. The commitment to help the inhabitants of the town is explained, not in terms of the reactions the state usually produces but in terms of the reaction the state calls for — in terms of *oughts*. In this instance, one of the normative consequences of the commitment is to form the relevant intention. Doing so is what, given his commitments, it would be correct, rationally speaking, for the person to do.[19]

How do we explain the fact that, although the evaluative judgment requires producing an intention, this doesn't happen in the politician's case? Here it is useful to consider two different ways of relating to a practical commitment.[20] On the one hand, we can attribute, from a third-person perspective, commitments to another speaker. We do this with doxastic commitments, as when I attribute to someone the commitment that X melts at 1084℃ based on my attribution to the same person of the commitment that X is made out of copper. In attributing the further commitment, we rely on the word "should" or "ought". Thus we may say, "If you think that the pipe is made out of copper, you should also believe that it melts at 1084℃." Similar relations exist with respect to practical reasoning. I may attribute to an agent the commitment to use a wet blanket based on his commitment to kill the fire: "If you intend to kill the fire, you should also intend to throw a wet blanket on the flames." When we make attributions based on what an agent should do according to his own commitments, we make use of practical reasoning but take up the detached attitude of an observer. On the other hand, a different way to deal with commitments is to take them up in first-person deliberation. Here, too, we engage in practical reasoning, but we do so with a view

[19]Brandom, who understands akrasia as the undertaking of incompatible commitments, notes that it is "one of the cardinal strengths of the deontic scorekeeping approach to intentional states in terms of normative statuses that there is nothing conceptually mysterious about the possibility of such incompatible commitments. Difficulties in coherently understanding akratic action and endorsement of incompatible beliefs arise from exclusive emphasis on a causal-functional model of intentional states" (Brandom 1994: 270).

[20]Here I am following (Brandom 1994: 269–271) closely.

to determining which course of action to take. Whereas in third-person attributions of commitments, we derive the conclusion that the agent we are talking about *should* do φ, first-person practical reasoning results in a conclusion of the form "I shall do φ", thereby constituting an acknowledgment of the commitment rather than a mere attribution.

The striking contrast of attributing and acknowledging commitments depends on a distinction between two social roles. Whereas attributions of commitments to another result, at best, in advice for the agent, acknowledgment of practically relevant commitments typically result in the performance of the action in question. Thus when from the social perspective of an assessor we conclude "*X* should φ", on the basis of attributions of prior commitments, our mode of thought, although it deals with practical matters, is distinctly detached or theoretical. A practical conclusion from the deliberative point of view, on the other hand, is practical not just in terms of its subject matter but also in its issue. We have a reliable disposition to respond to acknowledging a *hic-et-nunc* practical commitment by producing the performance in question.[21]

Notice that we can attribute commitments in a third-person fashion not just to others but also to ourselves. If we conclude a line of reasoning in this way, the "should" that appears in the conclusion can also apply to oneself. If I treat my own beliefs in a detached way, I can conclude that, given my beliefs and collateral commitments, I should believe that the pipe melts at 1084℃. Similarly if I treat my own practical commitments in a detached way, I can conclude that I should, all things considered, throw a wet blanket on the fire. The important thing is that doing so does not automatically entail forming the practically efficacious intention to do so. The upshot of my reasoning about the pipe is my attribution to myself of a doxastic commitment, but this need not mean that I acknowledge the commitment. Similarly the upshot of my reasoning about the fire is my attribution to myself of a practical commitment, but I need not thereby draw all the required conclusions. In particular, I need not acknowledge the commitment *behaviorally* by responding to it by performing the appropriate action — throwing a wet blanket on the fire.

Cases of *accidie* seem puzzling because they involve a peculiar non-responsiveness to positive evaluation. On the model presented, this is because "should" conclusions do not always reliably give rise to "shall" conclusions. The non-responsiveness is a matter of self-attributions of practical commitments coming out of sync with acknowledgments of those

[21]Brandom writes that "intentions are causes, for in the properly trained agent, acknowledgments of practical commitments reliably causally elicit performances" (Brandom 1994: 261).

commitments. Reasoning about one's own commitments *only* in a detached third-person manner robs the commitments of their direct relation to motivation. In particular, "I should" differs from "I shall" in terms of its noninferential significance, i.e. in the way it does not license or call for a nonlinguistic response.[22] But it is clear that if an agent judges all-out that he should do φ that he is thereby committed to doing φ, whether he likes it or not. Although it is psychologically understandable how there may not be a practical realization of the "should" evaluation, it reveals a rational fault in the agent: he fails to draw out the consequences of the commitments he has. Self-attribution of a practical commitment does not in every case trigger the reliable differential disposition to act, but that does not mean that there is no rational requirement to do so. The appeal to Stocker's maladies of the spirit explains the failure to form an intention psychologically but doesn't justify it.

These are only the contours of an account of *accidie*. But they allow us to see in principle that an account that understands evaluative judgments in terms of practical commitments is at least in as good a position as a rival dispositional account to account for cases where an agent is not attracted by what he himself regards as good. So Stocker's example that the good need not attract does not pose a difficulty for the Guise of the Good. However, Stocker's argument against high-brow conceptions has a second part, which is more challenging to our thesis. He argues not only that the good may fail to attract but also that *not only* the good attracts. In particular, he attempts to show that, in our darker hours, we may have an appetite for the bad as well as for the good. He gives the example of a person consciously making bad decisions as to his nutrition:

> Given certain moods, interest structures, energy levels, and the like — e.g. my having ceased caring about my well-being — what I want is this food, even though, perhaps even because, I realize it is the wrong amount, the wrong sort, ... i.e., bad for me. (Stocker 1979: 747)

Now it should not come as a surprise that we choose to do things even though they have some negative aspect or other and even though we know that they have it. Many if not all the things we do have not only desirable features but downsides as well. Driving to work is convenient but damages the environment; sleeping in is pleasurable but could cost you your job; eating organic meat is healthy but expensive — we are used

[22] Cf. Brandom (1994: 270).

to weighing pros and cons. What is more challenging is the radical claim that we can find something attractive, not despite its downsides, but precisely because it is bad, or at least seen as bad. According to Stocker, given certain adverse conditions, we can be attracted by the bad *qua* bad; and he concludes that, at least as a general thesis, the claim that we necessarily act *sub specie boni* must be mistaken.

Suppose an agent acts on a desire to eat a large amount of unhealthy food. It is certainly true that he may do so despite his realization that doing so is bad for him. But is it really the bad *qua* bad that attracts him? The structure of Stocker's example is that the agent acts on the desire to φ and the fact that φ'ing is bad is the purported object of the agent's desire. Now Stocker holds that such cases leave the defender of the Guise of the Good with two equally unattractive options:[23]

1. The agent judges that φ'ing *qua* doing harm is itself good.

2. The agent does φ only in order to attain some other goal ψ, which itself is desirable.

Stocker thinks that the first option is "too implausible" (Stocker 1979: 748). On the other hand, the second option does not cover all cases of "desiring the bad". While Stocker concedes that in some cases, doing harm can be a means to attaining a further goal, he insists that sometimes, when we are affected by a particularly bad mood, there need be nothing further that justifies the harm done. Perhaps this is true in certain modes of self-directed disgust. In such cases, the proper object of the desire just is the damage, so that he is doing something bad precisely because it is bad.

Of the two strategies Stocker suggests, the second is not particularly promising. There are certainly many cases where we do harm, to others or to ourselves, in order to achieve a further goal. But I agree with Stocker that we have no reason to assume that harming others or ourselves can never itself be our purpose. As he rightly points out, when we help another person, we often do so without ulterior motives. If I help my friend, the goal is not necessarily to deepen the friendship but may simply be to do what I can to help him in the spirit of friendship. Similarly, if in a mood of disgust or spite I do damage to someone, there need be no further motive involved — my desire may simply be for the other person's ruin. Yet if this is possible with respect to another person, a mood

[23] Cf. Stocker (1979: 748).

of self-hatred may equally prompt a desire to harm myself, without any expectation of further pleasure or relief.

Stocker is right, then, to reject the second defense, but I think he doesn't take the first defense seriously enough when he dismisses as implausible the idea that in the relevant cases the bad is a believed good. In my view, when someone genuinely desires an action as one of harming someone, he thereby regards doing harm as a good thing. Stocker find this absurd, presumably because he regards such a desire as in itself inconsistent. Thus it may be thought that evaluating positively a bad action such as causing damage is self-defeating. However, it is easy to see that the ascription

Jim treats hurting *X* as good* precisely because it is not good**.

need not be a self-contradiction. In the sentence, "good" is used in two different ways. In its second occurrence, "good**" refers to a substantive conception of the good. Such a substantive conception incorporates particular values. Although it is certainly true that doing physical damage to a person is bad, this assessment constitutes a substantive value judgment. In its first occurrence, however, the word refers to a purely formal notion.[24] Regarding something as good* is to take it as worth pursuing or desirable, as something that provides structures to the action by fixing its end. In particular, this notion of the good informs our processes of practical reasoning. But regarding something as good* need not entail its goodness** in any substantial sense of the word.

There are many substantial conceptions of the good to choose from. Such a good** may be whatever is morally required, the egoistical notion of satisfaction of the agent's own needs or the famililist idea of promoting the well-being of one's close family or clan. From this variety, it is clear that the claim that agents act under the guise of the good cannot be intended to posit a necessary connection between intentional action and the good** in any substantial sense. To think this would be to assume, absurdly, that particular non-trivial normative judgments are part of the very concept of intentional action. So an agent may treat harming someone else, an action he himself deems bad**, as good*. Nor is this thought restricted to harming others. We have little reason to assume that one could not treat harming oneself has positive or desirable, even if this is not part of any good system of values. Thus when someone desires eating an unhealthy amount of food precisely because it is harmful, this attitude towards the

[24]Compare the discussion of the formal notion of "good" in §3.3.

action as bad** in a substantive interpretation does not preclude him from also regarding it as good* in the formal sense. It is possible that the agent is committed to overeating and he takes the means necessary for achieving this goal, however misguided.[25]

If agents who are influenced by their psychic structure to intentionally aim at doing something substantially bad do not count against the Guise of the Good, then neither do more far-reaching instances of perverse desire. We can even concede the possibility in principle of an agent with radically inverted priorities. A particularly vivid illustration is due to Anscombe, who alludes to the case of Satan from Milton's epic poem *Paradise Lost*.[26] In the poem, Satan declares:

> So farewell hope, and with hope farewell fear,
> Farewell remorse: all good to me is lost;
> Evil, be thou my good (Milton 2005: IV, 108–110)

Disappointed with God and divine goodness and troubled with self-doubt and despair, Satan decides that from this point on, the bad will be his good. But when he does so, what he rejects is not the entire project of pursuing a formally good thing, the project of pursuing ends — to do so would be to renounce agency altogether. Instead he turns away in disgust from the conventional divine or human conceptions of the good — a substantial conception which includes what is morally right rather than "evil". His announcement says as much. Can we genuinely understand such a decision? Perhaps we can if, with Anscombe, we speculate about Satan's reasons for renouncing the traditional conception of the good: he may be doing so as a way of freeing himself from the slavery of God's laws or simply out of curiosity about how it is like to act unencumbered by moral codes.[27] Whether we can truly be said to know the reasons that move Milton's Satan will depend to some extent on the availability of such auxiliary accounts. Nonetheless, even if we are left with the bare observation that he is taking the fact that something is on all conventional accounts good as a reason to despise it, we have enough to go on to see at least in principle how even he, perversely, is acting *sub specie boni*.

[25] A similar response to Stocker's objection is suggested briefly in Tenenbaum (2003).

[26] Anscombe (2000: §39, 75). Velleman (2000a) cites the case of Satan as a counterexample to the Guise of the Good.

[27] Anscombe (2000: §39, 75).

4.3 Unreflective agency

High Brow need not be troubled by Stocker's alleged counter-examples. However, there is a different argument, due to David Velleman, that purports to threaten the idea that we always act under the guise of the good. Like Stocker, Velleman thinks that the Guise of the Good is too demanding, but unlike him, he doesn't think that the view is overly optimistic. He thinks it is overly reflective.[28] Velleman holds that, although High Brow may apply to some agents, some of the time, it doesn't portray the intentional agent in its full generality:

> The agent portrayed in much philosophy of action is, let's face it, a square. He does nothing intentionally unless he regards it or its consequences as desirable. The reason is that he acts intentionally only when he act out of a desire for some anticipated outcome; and in desiring that outcome, he must regard it as having some value. All of his intentional actions are therefore directed at outcomes regarded *sub specie boni*: under the guise of the good. (Velleman 2000a: 99)

Velleman takes issue with this assumption:

> Surely, so general a capacity as agency cannot entail so narrow a cast of mind. Our moral psychology has characterized, not the generic agent, but a particular species of agent, and a particularly bland species of agent, at that. (Velleman 2000a: 99)

So what exactly is wrong with the Guise of the Good? On the high-brow view, wanting something amounts to making an evaluative judgment about it, but for Velleman it is not plausible that the capacity to desire requires the capacity to perform evaluative judgments. On behalf of the high-brow view, he invokes a "qualified formulation", proposed by Davidson, that "the natural expression" of desire is "evaluative in form" (Davidson 2001b: 86). Thus my desire to stay dry is an attitude that is naturally expressed by statements such as "It is desirable to stay dry" or "I ought to stay dry". This affords the high-brow theorist a way of saying that intentional action requires evaluative judgments without being committed to the idea that every action is accompanied by an exercise of

[28] For a similar argument against the Guise of the Good, see Setiya (2007: 59–68).

evaluative capacities involving concepts such as "ought" or "good". On Velleman's suggestion, these concepts are a natural way to express the attitude, but they do not necessarily form part of the judgments themselves. On this view, one can be motivated by a desire without mentally grasping the corresponding evaluative proposition. But this raises a different problem. For Velleman the view that intentional action involves evaluative judgment adheres to what he calls the story of rational guidance, which says that "acting for a reason entails being influenced by the force of a mentally grasped justification of one's action" (Velleman 2000a: 104). The problem is this:

> According to Davidson's qualified formulation, however, a proposition that's essential to the justification of the action — namely, the proposition that the action's expected consequences are desirable — is merely a proposition that would naturally be used to express the agent's desire. And the agent can be moved by his desire without being able to express it or grasping the proposition with which it would naturally be expressed. He can therefore satisfy Davidson's story of motivation without having mentally accessed anything that justifies his action. (Velleman 2000a: 104–5)

In other words, this suggestion allows for the possibility of agents acting on a desire without being in the right way in contact with the proposition which in fact carries the justificatory force. The evaluative proposition of the form "… is good" only exists, as it were, in the background. But for Velleman, the agent acted in the light of this evaluation only if, at the time of acting, he grasped the proposition.

Thus with respect to the relation between intentional φ'ing and the evaluative judgment that φ'ing would be good, Velleman sees the High Brow theorist impaled on the horns of a dilemma:

1. The defender of High Brow can argue that φ'ing intentionally doesn't require mental contact with the evaluative proposition (that φ'ing is good). The action is caused by an inclination, but it is rationalized by the evaluative proposition. However, Velleman argues, then it is hard to see how the action counts as an expression of the proposition that justifies it. For how can the agent be guided by the proposition if he doesn't have it before his mind? According to the objection, the justifying proposition cannot play the crucial role it does unless the agent has access to it.

2. On the other hand, the High Brow theorist can argue instead that in φ'ing intentionally, you are necessarily entering in a mental rapport with an evaluative proposition. Because the proposition that justifies your conduct is mentally present, we can see how the evaluation is also what guides your action. However, this implies that every intentional action must be accompanied by an exercise of advanced conceptual capacities. That is not a realistic psychological picture. Making judgment about the goodness of things is on a higher conceptual level than object-level judgments. If you make such a normative judgment, you are reflective about what you do. But surely some actions are entirely unreflective.

How can the defender of the Guise of the Good extricate himself from this dilemma? First off, we should accept that Velleman's criticism of the second horn is warranted. To say that every intentional action is accompanied by the mental grasping of an evaluative proposition is to overstate the role of the intellect in the typical intentional agent. The second horn implies that in order for the agent to engage in something as unsophisticated as a simple intentional action, he must deploy high-level concepts. Velleman's official argument is that if we identify desires with evaluative judgments, we imply that only agents who have the conceptual wherewithal to make evaluative judgments can have desires.[29] The concept of the desirable or good belongs to a higher stratum of concepts operating on the level of reflection rather than the object level. Although small children competently use concepts on the object level — the concept of a fork, horse, etc. — they have yet to learn the use of the more sophisticated machinery of abstract discourse. As the current suggestion links intentional action to having the concept of what is desirable, a concept that small children arguably don't have, the suggestion entails, against expectations, that small children do not genuinely have desires.

The objection that creatures to whom we usually attribute agency lack the concept of goodness deserves to be taken seriously.[30] For the moment, however, note that this objection against the second horn assumes that if one kind of creature does not possess concept X even though he is capable of acting then X cannot be a presupposition of intentional action in a different kind of creature, either. Rather than explicitly arguing against this assumption, I will only note here that this assumption is not trivial before looking for another motive for the assumption. Fortunately, such a motive, which explains why even if we restrict our attention to mature

[29] Velleman (2000a: 104).

[30] I address this objection in §6.

rational agents we could think that acting on a desire cannot entail positive evaluation, is not far to seek. On the present suggestion, the outright identification of desiring with evaluative judgments means that acting on a desire is a matter of explicitly deploying an evaluative concept. This picture has us mentally token the concept of goodness or "ought" each time we act for a reason. No doubt sometimes this may happen, but on this high-brow suggestion every intentional action needs to conform to this model.

The trouble is that, with the concept of goodness belonging to a higher stratum of reflective concepts, the current suggestion seems to entail that *every* intentional action is conceptually reflective. The implication is that an intentional agent explicitly grasps a proposition or thought of the form "... is good". But this seems false simply as a claim about what is involved psychologically in acting for reasons. We do not seem to pause mentally on an evaluative step. Mere intentional action does not seem to mentally consult explicit maxims of action. These are sophisticated employments of our rational capacities, of the intellect, that need not be actively employed in action.

Velleman correctly suggests that accepting the second horn of the dilemma has implausible implications. However, his criticism of the first horn is faulty. By appealing to the story of rational guidance, he relies on a principle:

Guidance Principle If an agent acts on the consideration p, then he must be guided rationally by the positive evaluation of p.

Although the principle is sound, the way Velleman interprets it is problematic. As he understands it, the principle says that being rationally guided by p requires that you are explicitly aware of the proposition that p. If his interpretation is correct, then the only way to be guided by p is by being in a "mental rapport" with the proposition. On this interpretation, the first option — saying that doing φ intentionally without making an explicit evaluative judgment — is instantly ruled out. However, Velleman's interpretation of the Guidance Principle is not without alternatives. To see this, we need to reconsider what happens when we engage in practical reasoning.

In describing the high-brow view, Velleman refers to Davidson's view of intentional action. On Davidson's view, intentional action can be treated as if it were the conclusion of a practical argument.[31] The description

[31]See Davidson's description of seasoning a stew (Davidson 2001b: 85–85).

aptly captures the high-brow view. It is useful to understand the intentional explanation of an action as the exhibiting of a sample piece of practical reasoning which has as its conclusion the decision to perform the action. In our terminology, a practical commitment can be the result of a practical syllogism. When I open my umbrella on the street, a classical explanation in terms of belief and desire has this shape:

> *Desire*: I want to stay dry.
> *Belief*: I will stay dry only if I open my umbrella.

It is easy to see, however, that these propositions do not permit us to construct a valid practical inference because they are attributions on the mental states rather than their expressions. From the fact that I have a certain mental state, nothing interesting about what I should do follows. This naturally leads to the sequence:

> *Premise*: Staying dry is good.
> *Premise*: I will stay dry only if I open my umbrella.
> *Conclusion*: Thus, I shall open my umbrella.

The two premises of this argument correspond to the components of the belief-desire explanation. The second premise is a matter of empirical, causal fact, whereas the first premise expresses the pro-attitude or intention towards the goal. The attempt to construct a valid practical argument requires the introduction of normative vocabulary — in this case, "is good". We can understand someone coming to undertake the commitment to open his umbrella — and to acknowledge the commitment by doing so — as the result of a practical inference taking these starting points. The point now — and this is what I take to be a crucial motivation behind the anti-high-brow argument — is that it doesn't seem psychologically realistic to assume that, when we open an umbrella, we necessarily perform this argument or something like it. On the picture proposed, we pause mentally on the evaluative step. As we have pointed out, this premise incorporates normative vocabulary which is at home at a higher reflective level. The picture requires us to acknowledge mentally the goodness of what is being pursued in action. But as far as I can tell, we do not always make such an explicit acknowledgment, or even very often. We hesitate to say that an explicit employment of normative concepts is a necessary part of acting for a reason.

Note that there seems to be no analogous problem with the empirical premise. It is not surprising that an agent who opens his umbrella is

explicitly aware of the fact that he has to open his to umbrella unless he is prepared to get wet. Nor does it surprise us that the action is rationally guided by the empirical proposition in the form of an explicit premise. By contrast, we hesitate to accept that the agent is explicitly grasping the premise of the form "... is good" unless he is opening his umbrella with extraordinary reflectiveness. We aren't always thorough in this way. To echo Velleman, this model captures a specific, intellectual but rather bland agent who keeps explicit track of his commitments, but the model isn't wide enough to cover the generic agent, who is sometimes thoughtless or hasty, unreflective or rash.

But this doesn't mean that we have to give up the Guise of the Good. In my view, if we see things the right way, we can accept a version of what I called the first horn of the dilemma. To see this, we first need to re-examine the Guidance Principle. It is true that in order for an agent to truly act on the thought that p is good, she needs to be rationally guided by that thought. On Velleman's view, the only way to be rationally guided by an assessment is to have a proposition with normative content present to one's mind. However, there are ways of satisfying the guidance condition other than entertaining a proposition.

In fact, Velleman's argument is based on a widespread assumption that Ryle calls the intellectualist legend. This philosophical picture sees the primary exercise of minds in specifically intellectual operations of theorizing. To be rational, on this view, is to be able to recognize truths. Ryle argues against "the general assertion that all intelligent performance requires to be prefaced by the consideration of appropriate propositions" (Ryle 1963: 30). For Ryle, it is wrong to assume that every intelligent operation needs to be the work of the intellect, a further mental act of theorizing about a truth that accompanies the operation. But an operation can be properly called intelligent or rational without assuming the existence of any such further "internal process of avowing to [oneself] certain propositions about what is to be done" (Ryle 1963: 30). It is not true that contact to the proposition is required in order for the agent to "execute his performance in conformance with those dictates", as one might think, mistakenly, that a chef needs to recite his recipe inwardly so that he prepare his meal skillfully. But the cook's skill doesn't lie in his knowledge of intellectual truths but in rules or criteria that govern his performance but that he may nonetheless not be able to formulate.

4.4 Achilles and the tortoise

Let us now see if we can replace the intellectualist legend castigated by Ryle with a more adequate picture of rational guidance. This requires a deviation from the course of the argument to make a general point on the basic form of practical reasoning. I will continue the defense of the Guise of the Good in the following section. In his short paper "What the Tortoise said to Achilles", Lewis Carroll makes an important point about the nature of arguments by presenting a philosophical conundrum in the spirit of Zeno's paradoxes.[32] Caroll tells a tale of the levelheaded hero Achilles, who is challenged by the skeptical tortoise to perform a seemingly easy task: not, as in Zeno's original, to win a race but to prove that a simple syllogism, *modus ponens*, is good. In a first step, the tortoise asks the hero to write in his notebook the propositions:

(A) $p \rightarrow q$
(B) p
(Z) q

Now the tortoise points out that someone who isn't convinced that A is true or that B is true might yet accept the sequence of propositions — the inference — as valid. This person might express his view by saying that he rejects both A and B but accepts the hypothetical proposition that if A and B, then Z. Having made this observation, the tortoise further notes that the converse case is equally conceivable. A different person might accept A and B but reject the hypothetical proposition. The tortoise turns out to be a skeptic of this second sort.

According to the tortoise, the fact that the conditional can be rejected shows that he is not bound by logical necessity to accept the conclusion Z. Insisting on his skeptical view of the conclusion, he challenges Achilles to force him to accept the conclusion. The tortoise invites Achilles to write down the hypothetical as an additional premise:

(C) If A and B are true, then Z must be true.

After Achilles has added (C) to his notebook, the tortoise agrees to accept the proposition. But just when Achilles triumphantly declares that his task is over, the tortoise demurs: he still refuses to accept the conclusion. Achilles replies, "But if you accept A, B and C, then you must accept Z!"

[32] Carroll (1895).

The tortoise quickly points out that Achilles has just presented him with *another* conditional proposition. Repeating his earlier move, he says that he is willing to accept this proposition if Achilles agrees to add it to his list as:

(D) If A, B and C are true, then Z must be true.

Naive as he is, Achilles hopes that now the tortoise doesn't have a choice but to accept Z. Unsurprisingly, the hero's delight doesn't last long, for once the statement D is added to the sequence, the tortoise again refuses to concede that he is compelled to accept Z. As in Zeno's original, this starts an infinite regress. As Achilles's notebook is quickly filling up, it becomes apparent that convincing his skeptical interlocutor requires adding an infinitude of conditional propositions C, D, E and so forth, each one longer than its predecessor — an impossible task.

Although the tortoise appears to be cooperative, he keeps refusing to accept inferences as valid, so Achilles never reaches the point where he can get the tortoise to agree that the intended consequence follows from the premises. What conclusion should we draw from this observation? We might be tempted to begin to have doubts about the validity of reasoning in general. Achilles's heroic efforts fail to defend even the simplest *modus ponens* against the skeptical insistence. That the tortoise can achieve a stalemate, if not a victory, is a paradoxical outcome.

But the tortoise's success relies on a faulty picture of logical arguments. If Achilles accepts the terms of the game as laid down by the tortoise, he is bound to lose. The tortoise's claim of being cooperative is based on the fact that he is willing to grant the principles that Achilles appeals to as valid. For instance, he accepts that A and B together imply Z, subsequently he accepts that A, B and C together imply Z, and so forth. He does not, however, do this by conceding that the conclusion follows in practice. Instead he only agrees to add the principle to the list of valid propositions in the notebook. Written down as a proposition, the principle is added as an additional premise and takes the form of a conditional — "If A and B, then Z", "If A, B and C, then Z", etc. The assumption here is that adding a further proposition comes down to the same thing as accepting an inference as valid.

These two do not really amount to the same thing, however, and the tortoise exploits this fact in the paradox. Although he accepts the conditional proposition, he relies on the fact that moving from this conditional, along with the antecedent conditions, to the consequent is a further step

that can itself be challenged. It is true that this is a further contentious step. But this means that accepting a conditional and regarding the corresponding inference as valid is not exactly the same thing. The faulty intellectualist picture that makes Carroll's tale seem paradoxical takes it for granted that the only thing that matters in logic are propositions of the kind the tortoise accepts. The apparent paradox indicates that we should reject this picture.

A better picture sees a logical argument as a sequence of propositions, premises and conclusion. When we reason, when we make an inference, we transition from acknowledged commitment to the premises to acknowledged commitment to the conclusion. In making an inference, we regard the transition as a valid one. For example, if we reason:

> A: Argos is a dog.
> B: If Argos is a dog, then Argos is a mammal.
> Z: Thus Argos is a mammal.

we assume that the premises together imply the conclusion. Relying on this implication is related but not equivalent to adding as a further premise the conditional proposition (C) "If Argos is a dog and if Argos is a dog, then Argos is a mammal, then Argos is a mammal". Instead it amounts to being committed to treating the inference as good.

The difference is between, on the one hand, explicitly holding a proposition to be true and, on the other, treating an inference as valid in practice. For an inference to go through, both explicit truths and a valid inferential principle have to be present. It is not enough to accept the premises; we must also accept that the premises entail the conclusion. To do this, we necessarily rely on the kind of cooperation the tortoise stubbornly refuses: the willingness to appeal to a common practice of accepting certain inferences as good. Unless a propriety of inference implicit in practice is acknowledged, Achilles will not succeed, no matter how many explicit premises he adds. Every step of the way, the tortoise concedes another conditional proposition. This adds to the explicitness of the argument, but it never reaches a point where, *per impossibile*, all implicit aspects of an underlying practice has been removed. The assumption that removing all implicit treating-as-good is possible and even required to show that an argument goes through is part of the faulty intellectualist picture that we need to give up.

Any skeptical conclusion, then, that arguments never properly speaking go through or that it is futile to justify a piece of reasoning, is unwar-

ranted. Rather than agreeing to the tortoise's terms, Achilles should appeal to a shared rational practice of accepting certain inferential steps as primitively good. The tortoise did not blunder in challenging the hero's inferences. It is perfectly permissible to ask for further confirmation or support of an inferential relation. And it is possible appropriately to respond to such a query. Achilles's patient answers are a way of doing that. In responding to each of the challenges, Achilles makes more and more of his argument explicit. The problem is with the idea that codifying *all* steps in the form of premises is possible or that it is even desirable. After all, the argument consisting of A, B, C, D and Z is hardly more interesting or illuminating than the one he started out with.

The picture suggested by Carroll's story raises a number of points which will be important in what follows. *First*, what is the function of conditionals such as proposition B or C above in this picture? It should, of course, be pointed out that the logic of conditionals is intricate. It will not be possible to explain their role comprehensively here. However, as an approximation it is helpful to follow Ryle in thinking of conditionals as "inference-tickets".[33] The idea is that accepting the conditional is akin to having in one's pocket a train ticket that allows you to go from one place, where you already are, to another, where you are going. To spell out the analogy, the conditional is a license that permits the transition from the antecedent to consequent. Suppose you have the commitment that p in your commitment-box (or pocket). Accepting the conditional "p⟶q" then amounts to being entitled to put into your box the commitment that q.

Nonetheless, as before, we need to keep in mind that there is a difference between accepting the conditional proposition, which explicitly *says* that you are allowed to make the transition, and being actually prepared in practice to regard the transition as good. There still is a possible gap between the two. Achilles in the story adds another conditional, which makes explicit what was already (or should have been for an interlocutor willing to cooperate) implicit in practice before. The conditional, then, is a device for making explicit an implicit inference license.

Second, we noted that making an inference is to regard the transition from premises to conclusion as valid. But of course not all the transitions we treat as good really are good. Sometimes our mistake is just an oversight, sometime the defect runs deeper. Speakers do not always agree as to which inferences are good. Spotting faulty inferences is one of the ways in which we work out our disagreements with another speaker. If

[33]See Ryle (1971) and Ryle (1963: ch. 5). See also (Sellars 2007b).

there is disagreement, we can criticize the other by pointing out that a given inference is incompatible with another commitment or that it is unfounded. By doing so, the speaker can challenge the correctness of the inference. This is what the tortoise in effect is doing. There are different ways to respond to such a challenge, but one of the main ways is to make the propriety of the inference explicit in the form of a claim. We often defend an inference by formulating a conditional, and this is also what Achilles does.

Sometimes, as with the tortoise's nerve-racking questions, such a challenge can seem pointless. The tortoise's suspicious attitude towards the inference in question is unwarranted. But in other circumstances, a challenge of this type is perfectly legitimate. In these cases it is a way to effect conceptual change. A typical response is to appeal to a conditional which licenses the inference in question. Making the goodness of the inference explicit exposes it to public scrutiny. The challenger can continue the challenge by targeting the conditional. This can be done, for instance, by supplying counter-examples. Challenges of this kind help articulate the concepts in question better, which again is helpful in detecting interpersonal difference concerning the meaning of words. When we subject our commitments to challenges of this kind, we engage in what Sellars calls the Socratic method.[34]

Third, it might be found unusual that no appeal has yet been made to the notion of logical validity. The reason why this may seem unusual is a tendency to treat as valid only inferences which are logically valid. According to a long tradition, logical inferences are inferences that are valid in virtue of the form of the propositions only. An example of such an inference *modus ponens*: "p. p→q. Thus, q." It is characteristic of arguments of this type that we can easily tell that they are valid without knowing anything about the specific conceptual content lying behind the propositional constants. Thus we need not know what "p" and "q" stand for to recognize the argument as valid.

Arguments with *modus ponens* structure, then, are formally valid, but not all valid arguments are formally valid. Take as an example the argument

(1) Philadelphia is East of Pittsburgh.

(2) Thus, Pittsburgh is West of Philadelphia.

Clearly this inference is valid and we do not hesitate to treat it as such, but to know that it is one needs to know specifics about the concepts "East

[34]Sellars (1980b).

of" and "West of". Thus the validity is not just a feature of the form of the proposition. In the past, many have treated formally valid inferences as the only valid type of inference. According to these philosophers, we should understand material inferences like (1)-(2) as essentially elliptical.[35] If they are right, we need to complete the argument by supplying a further conditional premise that establishes a connection between (1) and (2). Thus the allegedly enthymematic (1)-(2) would require supplementation by

(3) If place X is East of place Y, then place Y is West of place X.

Adding this as a premise turns the earlier inference into a formally valid argument and the conjunction of the premises and the conclusion into a logical tautology. The tradition takes it that, unless we supply such a conditional premise, inferences of the (1)-(2) type are essentially incomplete. As a consequence, material inferences are treated as derivative compared to formal inferences.

Brandom calls this position the formalist approach to logic, which contrasts with his preferred materialist approach to logic.[36] The materialist view here has no connection to the doctrine of the same name which is a version of metaphysical monism. Nor are "material" inferences related, other than by name, to the "material" conditional.] The difference is one of explanatory priority. The materialist starts with proprieties of inference that are implicit in practice and goes on to explain formal validity in terms of material validity. The formalist, on the other hand, starts with taking as primitive what speakers explicitly say and believe, in the form of propositions, and proceeds to explain material inferences in terms of the truth of propositions. As was said, this works by expanding material inferences by adding explicit conditional propositions. Brandom follows Sellars in holding the formalist approach to be a dogma which we ought to reject.[37] In what follows, I will adopt the materialist approach. Comparing the conceptions is outside the scope of this chapter. Three points should be mentioned that favor the materialist view:

1. The story of Achilles and the tortoise shows that no matter how many premises we add to an argument, we still need to rely on an underlying implicit practice of regarding certain inferences as proper. Crucially, this is true also for formally valid inferences: we

[35]The expression "materially valid inference" is due to Sellars (see Sellars 2007b: 3ff).
[36]Brandom (1994: 97–107).
[37]Cf. Sellars (2007b).

need to admit certain inferences as primitively good in a shared practice. But if we admit primitive goodness on this level, it seems natural to extend this to inferences that have material validity, but not formal validity. Why deny that we regard the goodness of inferences of the form (1)-(2) as primitive goodness when we accept primitive goodness for formal inferences?

2. An important motivation for materialism that cannot be fully explored here is that formalism cannot support an inferentialist conception of conceptual content. The conceptual content of an intentional state is what the agent is committed to when he is in the state. This, in turn, is determined by, on the one hand, what would be a reason for having that commitment and, on the other, what having that commitment would be a reason for. The conceptual content of a state then is a function of the valid inferences that permit undertaking the commitment and the valid inferences that undertaking the commitment would permit. Now if the inferences in question were restricted to formally valid inferences, the conceptual content would be too narrow to support inferentialism. If we allow materially valid inferences to add to the conceptual content, however, the inferential relations become much richer. The notion of a non-enthymematic material inferences is essential if we do not want to preclude an inferentialist semantics.

3. Finally, we simply do, in everyday reasoning, treat material inferences as good without supposing that there are hidden premises lurking in the background. Thus it seems natural to go from (1) to (2) without a detour to (3).

This last point leads to a natural extension of the idea. The Pittsburgh-Philadelphia argument does not follow the model of the classical syllogism, in which the conclusion is inferred from exactly two premises. But at least since the advent of modern logic it has been clear that arguments are more varied than the ancient format of the syllogism allows. To name just one example, "¬(p and q). Thus ¬p or ¬q" is clearly a valid argument, yet it has only one premise. The fact that not all inferences are of the two-premise format was noticed already by Hume:

> As we can thus form a proposition, which contains only one idea, so we may exert our reason without employing more than two ideas, and without having recourse to a third to serve as a medium betwixt them. We infer a cause immediately from its effect; and this inference is not only a true

> species of reasoning, but the strongest of all others, and more convincing than when we interpose another idea to connect the two extremes. (Hume 1978: 1.3.8, 97 n. 1)

Hume's topic in this passage is the inference from cause to effect, which to him is the only means of achieving knowledge of the empirical world. According to Hume, an inference is just a type of habitual association of mental episodes: a habit of the mind to move from one set of perceptions to another. When the mind is presented with the cause, it immediately transitions, by an easy and natural movement, to the effect. We do not need the idea of a connection which mediates between the cause and the effect; we do not need knowledge of a law. This is reflected by the structure of causal reasoning. Hume points out that an inference which lacks an intermediate step — a statement of the law — is still a genuine sort of reasoning. The role assigned by a formalist approach to a conditional law statement is played in Hume's view by the habit produced by the constant conjunction of past instances in the mind. We can recognize this idea as a version of the view, which is the present topic, that in single-premise inferences, the connection between premise and conclusion is implicit in practice, rather than explicit in propositional form.[38]

Hume uses the word "inference" chiefly in the context of casual reasoning: reasoning about the effect of a cause. In our usage, the word has much broader application, which includes but is not restricted to the context of empirical or causal thought. Thus we can apply the idea of a single-premise inference to simple non-causal theoretical reasoning:

(4) Argos is a dog.

(5) Thus, Argos is a mammal.

We can accept that there is a materially valid transition moving from (4) to (5) which doesn't pass through a conditional such as "If x is a dog, then x is a mammal". The inference is not *logically valid*, but it is treated as materially valid by competent users of the concepts *dog* and *mammal*. The fact that all dogs are mammals doesn't figure in the premises of the argument. In other words, it remains implicit in the practice in which the speaker and his audience participate.

[38]It must be mentioned that in another regard, Hume does not have the right view of inferences. Thus he credits the imagination, an essentially arational faculty, with the habit of moving from cause to effect, and he argues that these inferences cannot in any way be proved valid in a rational way. Thus to him causal reasoning has very little to do with rationality. In my view, there is nothing wrong with saying that making inferences of this type are the doings of our reason.

It may be objected that despite what we said the short argument is still incomplete. But in the light of the preceding observations, we can see that this is either misleading or wrong. The inference is incomplete in the sense that the argument is seen to be correct only in combination with a further rule, viz. a rule according to which *inter alia* the transition from (4) to (5) is permissible. Still the objection is misleading, for *all* arguments are incomplete in this way. As the tortoise's insistence shows, we always need a rule of inference to supplement the premises; the validity of an argument never is purely a matter of the premises. Thus it is wrong that the single-premise argument is lacking in a way that regular syllogism-style arguments are not. The argument does not require the addition of a further conditional to be made valid, and adding the conditional does not make the inference fool-proof or independent of the underlying inferential practice.

Finally, it is important to note that just as in the theoretical domain, we can find valid single-premise arguments in the practical sphere. Here are two examples:

> It is 5 p.m..
> Thus, I shall have the 5 o'clock tea now.

and

> I shall clean up the kitchen Thus, I shall bring out the trash

It is natural to accept the transition as materially good without bringing in an explicit conditional premise, even if the inference doesn't possess logical validity. The rule that allows this transition remains implicit in the inferential practice. We can be entitled to moves of this kind without having in mind an explicitly conditional proposition.

4.5 The Guise of the Good defended

Returning to the objection raised by Velleman against the Guise of the Good, these reflections on the nature of reasoning help us guard against the intellectualist legend. Velleman's criticism of the Guise of the Good was that an intentional action must be guided by an evaluation if we are to see it as rationalized by it and that this in turn requires grasping a proposition involving "good" or "ought". But the argument assumes that

if an intentional action is guided by an evaluative proposition, the agent's reasoning must contain that proposition as an explicit premise. We can now see how the Guidance Principle can be satisfied without such an explicit "mental rapport". The reasoning can be guided by the evaluation even if the agent does not pause, even for a moment, to inwardly consider an evaluative premise.

To see this, return to the piece of rainy-day reasoning introduced above:

> Argument 1:
>
> *P1*: I will stay dry only if I open my umbrella.
> *C*: Thus I shall open my umbrella.

A logical formalist might object to this argument as incomplete and propose an expansion:

> Argument 2:
>
> *P1*: I will stay dry only if I open my umbrella.
> *P2*: Staying dry is good.
> *C*: Thus I shall open my umbrella.

What we have seen is that even now the tortoise might object and demand supplementation by a further premise:

> Argument 3:
>
> *P1*: I will stay dry only if I open my umbrella.
> *P2*: Staying dry is good.
> *P3*: If I will stay dry only if I open my umbrella and it is good to stay dry, then I should open my umbrella.
> *C*: Thus I shall open my umbrella.

We could possibly go on, but as a result of the previous section we no longer have the patience to humor the skeptical reptile. Despite the tortoise's protestations to the contrary, the practice of making inference does not involve the ability of making all relevant rules of inference explicit in the form of premises. In fact we have no such ability. A determined skeptic can always refuse consent to making the inferences required to connect the premises to the conclusion. We always rely on the readiness to accept certain inferences as valid. But we have also seen that there is no need to make all our inferential steps explicit. We implicitly accept the principle of inferences on which the moves rely as valid: they

are part of our inferential practice. Far from being reducible to a set of explicit premises, a shared practice of treating inferences as valid forms the necessary condition of using explicit premises in the first place.

It follows that Argument 2 is not in fact incomplete and does not require supplementation by a further premise *P3*. It is certainly possible to add this premise. But the fact that we have this possibility does not suggest that Argument 2 was in any way deficient. True, the propriety of the inference is not immune to doubt. If someone, like the tortoise, questions the validity of the inference, one good way to react is to supply *P3* as a further explicit premise. We can then move to the more explicit Argument 3. But in normal cases, the additional premise is little more than additional weight serving little purpose. It is perfectly legitimate to move from *P1* and *P2* to *C* without paying these complications any mind — and we do so most of the time.

Next, Argument 1 stands to Argument 2 in a similar relation as Argument 2 to Argument 3. By the same line of reasoning that shows Argument 2 to be perfectly acceptable without supplementation, we can see that Argument 1 need not be seen as an enthymeme. It is natural for the agent to move from *P1* to *C* without adducing any intermediary steps. Again, such an argument, though short, need not be seen as incomplete. It is even more obvious than before that it does not follow from this that the argument is immune to challenges. Quite the opposite — short arguments are apt to be questioned. A skeptical inquiry by another, or critical reflection by the speaker himself, evokes the response of moving to a more articulate practical inference by the introduction of an intermediary premise. Crucially, moving to Argument 2 by adding premise *P2* can be seen as a response to such a challenge.

This is not to say that there should never be explicit normative premises such as *P2* in practical arguments, nor that they are superfluous. The point is rather the limited one that these propositions can be seen as dispensable *as explicit premises in the argument*. The function of *P2* in Argument 2 is to make explicit the propriety of the inference, in Argument 1, from *P1* to *C*. If there is nothing in Argument 1 as explicitly stated to play the same function, that is because the same function is played by an ought-to-be rule governing the inference in the reasoning. The rule of inference allows going straight from the realization that only opening the umbrella will keep me dry to the decision to open the umbrella. Adding the evaluative thought that staying dry would be good is a way to make explicit what we implicitly rely on when we treat short practical inferences like Argument 1 as valid.

With regard to Argument 2, when the agent draws the conclusion from *P1* and *P2* to *C*, he implicitly takes that inference to be good. As we have seen, it is possible to articulate the inference further by codifying it in the form of *P3*. What is made explicit by formulating this premise, however, is already implicitly present as a rule of inference in Argument 2. *P3* has the function of codifying an inferential practice. The important point now is that adding the evaluative premise *P2* serves a similar purpose. This has a consequence for the status of the assessment. It is true that the positive assessment of staying dry is not present in the form of an explicit premise in the shorter argument. However, the assessment is still clearly relevant to the way the decision comes about. Though not operative as an explicit thought, it is what provides rational guidance to the agent in drawing the practical conclusion by being the inferential rule that governs the argument. By going through the short Argument 1, the agent implicitly commits himself to the goodness of the inference. The agent's regarding staying dry as good plays a vital part in his drawing the conclusion to open his umbrella. The evaluation is present, not in the form of a proposition explicitly grasped, but as an implicit inferential pattern.

Velleman's criticism of the Guise of the Good relies on the thought that the only way to be rationally guided by a judgment is through direct mental contact with the relevant evaluative proposition. Leaving the intellectual legend behind allows us to see that in some practical inferences, the argument leading to the decision is guided rationally by an evaluative attitude even though the agent's action is not preceded by an intellectual act of grasping the evaluative aspect. When our agent opens his umbrella after going through a short practical argument, we can say that he values staying dry and that he acts *because* of this evaluation. Rational guidance need not be intellectual: the rule of inference that governs the practical commitment is itself rational insofar as it is subject to criticism, including self-criticism by the agent himself.

Although we sometimes make our reasoning explicit, we are not always so thorough or careful. In Velleman's view a high-brow conception of agency necessarily portrays the agent as conceptually meticulous — and thus as bland. The above discussion explains how action guided by an assessment can be sometimes careful, sometimes unreflective. When we are acting as a matter of routine, our evaluative attitudes stay in the background and guide our conduct in the form of inferential principles. This need not signal a deficiency. Explicitly paying attention to our normative principles in a Socratic manner is occasionally a worthwhile goal but must surely remain something of a philosopher's dream as a gen-

eral maxim. What is more, some situations in life call for quick reactions rather than deep reflection. In those moments where we do strive for conceptual clarity, however, we have the option of turning implicit inferential patterns, governed by an ought-to-be rule, into explicit premises that we can use in a practical argument. In doing so we bring our practical commitments out into the light. Acknowledging a practical commitment opens it up to public examination and criticism, but it can also lead to our endorsing the principle in a more clear-headed way, a process that may give us greater confidence. Specifically it allows us to see the commitment in relation to whatever else we are committed to bringing about. A degree of conceptual clarity is a prerequisite of harmonizing the various goals and values that we care about and that contribute to our sense of self as a unified person.

If this account is true, Velleman's argument against the Guise of the Good has no bite. Out of the two options presented as a purported dilemma, we can accept the first. If we act on the evaluation that p is good, this need not imply that the thought that p is good is an explicit premise in an argument, because the evaluation can be present, in the background, as an ought-to-be inference rule. When an agent acts intentionally, he does so *sub specie boni* because the action is the expression of a practical commitment amounting to a value judgment, but he need not make this judgment explicitly. Finally, we have seen how the Guidance Principle can be maintained even without supposing that the agent grasps explicit propositions. We can conclude that no real dilemma exists.

4.6 Agency without rationality?

Although the arguments we considered so far take issue with the High Brow thesis that *all* our actions involve an evaluative stance, as we have interpreted them they grant the assumption that intentional agency is exemplified only by mature rational beings. However, this claim, about the demarcation of agency, is also controversial. To conclude this chapter, we will explore the sense in which agency, as a competence, requires capacities that are available only to concept-users.

Roughly, High Brow holds that doing φ intentionally requires being practically committed to φ'ing, which is a way of regarding φ'ing as good. It follows that having the ability to deploy normative concepts is a prerequisite for intentional agency. Even more, the point affects other terms as well. Call "agency-concepts" the handful of concepts only apt to describe

intentional agents. Intentional agents do things *intentionally*, *for a reason*, they are sometimes wrong, sometimes right *instrumentally*, i.e. about how to do what they have reason to do, they *choose* what to do and realize their *intentions*. Strictly speaking, according to High Brow, agency-concepts only apply to creatures whose conceptual repertoire contains normative concepts. Creatures without the concept of the good fall outside the proper range of application of agency-concepts.

However, it strikes many as natural to use some or all of these concepts in descriptions of the behavior of creatures that are not plausibly credited with anything so much as sophisticated normative concepts. In particular, the idea is that agency is not the reserve of mature human language-users but also with the reach of higher non-human mammals or prelinguistic human infants. This observation can be turned into an argument against High Brow. We have no qualms to describe, say, 13-month old children as doing something for a reason. But if the proposed theory of agency is correct, such an attribution constitutes an improper employment of the use "reason". However, such a classification may seem counterintuitive. Here is Velleman taking issue with the High Brow theorist or, in his nomenclature, the cognitivist:

> If the cognitivist means to characterize desire as an attitude toward an evaluative proposition, then he implies that the capacity to desire requires the possession of evaluative concepts. Yet a young child can want things long before it has acquired the concept of their being worth wanting, or desirable. Surely, the concept of desirability — of something's being a correct or fitting object of desire — is a concept that children need to be taught. And how would one teach this concept to a child if not by disciplining its antecedently existing desires? (Velleman 2000a: 104)

In particular, possession of normative concepts such as desirability or goodness is often regarded as a dubious prerequisite for agency. The reason is that we cannot ascribe the evaluation that φ is good to a creature unless it can also *say* of φ that it is good. Non-rational creatures lack the ability to deploy normative concepts in the emphatic sense in which language training is indispensable for concept-use. But then no creature could regard something as good and, as a consequence, act intentionally without having first learned a number of concepts. According to the objection, High Brow is committed to denying the dimension captured by agency-concepts to infants or higher animals, which are, after all, creatures with minds.

It is worth emphasizing that with respect to human children, the bar is not as high as it may seem. It may be thought that the Guise of the Good mandates, as qualification for eligibility in the class of agents, mastery of decidedly high-brow expressions such as "you ought to", "worthwhile" or "desirable", to say nothing of "goodness". Pre-school children do not typically express their approval by calling things "desirable" or "worthwhile". But we should not conclude that they steer clear of evaluative language. In particular, they often use richly descriptive terms such as "fun", "unfair" or "mean" to assess actions or people as good or bad. These thick normative concepts, which imply normative judgments, are in heavy use long before dry, academic-sounding normative vocabulary catches on.[39] Clearly, the ability to deploy normative concepts must not be exercised by using purely normative vocabulary but can take the form of using thick concepts as well. Interpreting High Brow accordingly will perhaps reduce the urgency of Velleman's argument that young children lack the concept of desirability.

Still many philosophers maintain that creatures entirely devoid of rationality are nonetheless capable of instantiating agency, citing common usage. There is a temptation to react to arguments of this type by conceding some of the ground to Low Brow. The proposal would be to accommodate the semantic intuition favoring non-rational agency by relaxing the criteria for ascribing normative concepts. Such a proponent would be convinced by the arguments of the preceding chapters that intentional action necessarily involve regarding φ or its outcome as good in some way. On this view, what introduced the trouble is the thesis that ties regarding φ as good to having acquired linguistic abilities. Sure, one can hardly be practically committed to a project without in principle being able to express practical commitments verbally. As states governed by language-like rules, practical commitments are part of a package one acquires when one learns a language. But although the proponent wants to remain faithful to High Brow, he is unhappy with the implication that pre-linguistic creatures are not agents. So although he accepts that acting for a reason is conceptually structured, and indeed requires normative concepts, he proposes to allow for different kinds of concept-use. On this proposal, although the highly developed concepts used by mature human agents are tied to their linguistic abilities, there are lower-grade concepts as well that are not necessarily derived from the abilities of language-users. From this perspective, mere animals or infants can be seen as possessing concepts, albeit unsophisticated ones.

[39]Bernard Williams describes as "thick concepts" concepts that combine an evaluative judgment with a significant descriptive content. Cf. Williams (1993: 155–6).

Though I think we should resist this temptation, the proposal does recognize the fact that pre-rational animals or infants resemble human beings in that they have minds. To see the proposal in comparison to the view we have been defending, notice that there are two aspects of mindedness that a philosophical theory may focus on. Sapience is the aspect of mind accessible only to creatures who partake in language and discursive thought. Many creatures who are not sapient are nonetheless undeniably sentient creatures that experience pain or pleasure, rely on sensory impressions and physical inclinations and react through purposeful conduct. The strategy pursued by the conciliatory proposal is to put sentience first in an account of what is distinctive of beings with minds. Sentient beings are perceptually aware of their surroundings and capable of purposefully effecting changes in their environments. On this strategy, we can ascribe to mere animals or infants a species of concepts whose place is exclusively in this domain. While the ability to use sophisticated concepts only comes with the acquisition of language, sentient beings deploy rudimentary concepts. Taking this path would permit subsuming mere animals under agency-concepts, albeit in a less sophisticated way.

This first strategy focuses on the ways in which pre-rational animals resemble rational animals. But we can also choose to emphasize the discontinuity between the rational and non-rational parts of the animal kingdom. On this second explanatory strategy, to attribute to something a conceptual structure is to speak of a phenomenon with a high degree of complexity. The level of complexity involved in concept-use is unattainable except by creatures well-versed in linguistic abilities. The strategy gives pride of place to sapience, rather than sentience, in our account of the mental. It follows that conceptual capacities are dependent on rationality and exclusive to users of language in a strong sense.[40] What is more, as aiming at the good requires making use of concepts, intentional agency is, as far as we know, the preserve of mature human beings.

To make sense of this duality of strategies, we should note that each approach has its place. While some philosophical difficulties call for an explanation in terms emphasizing sentience, others are better treated while paying attention to the discontinuities between discursive and non-discursive animals. The answer which strategy is to be preferred, then, depends on the philosophical purposes we are hoping to accomplish. That said, the purposes of understanding human agency dictate choosing the

[40]Although it is common to speak of the language of bees or whales, their type of communication is (as far as we know) not linguistic properly speaking. To count as a language, a practice must at least be learned and taught, have a certain grammatical structure and include ways of making temporal, modal and normative distinctions.

sapience-first approach. In other words, High Brow should not concede too much ground by relaxing the criteria for attributing concepts to subrational creatures. The reasons are threefold.

First, our topic, in the philosophy of action, is not the behavior of living organisms in general but specifically acting *for reasons*. If we are to understand the concept of a reason, we must take into account its inexorable connections to the concept of practical reason*ing*. On our view, practical reasoning itself is a matter of making inferences from premises to conclusions. But the ability to make inferences rests firmly on linguistic abilities. A speaker does not count as inferring q from p unless he is in principle able to couch in a public language the argument "p, thus q."

Moreover, although our nature, like that of our relatives in the animal kingdom, contains inclinations, the way we agents with a rational background act on these aspects of our nature is fundamentally different from the way animal behavior is prompted by inclinations. In a mature rational agent, an inclinations comes into view as a reason only insofar as the agent *endorses* it and endows it with normative force. This requires the ability to step back from one's predispositions in a way that depends on the ability to ask oneself whether or not it would be good to realize the inclination. In particular, in order to act on an inclination *as a reason* one needs the ability to frame propositions involving normative concepts. We have seen that the agent need not in each case mentally entertain such propositions. Yet even if the agent does not exercise the ability to reflect on his inclination, to count as seeing the consideration that prompts the inclination as a reason he must be presupposed to possess the relevant linguistic competences.

Finally, our purpose is to understand what it is *for rational beings like us* to act intentionally. If we choose an explanation of action in sentience-derived terms, we cannot take into account all the features that come into view only for a rational agent. Rationality is a package that involves many abilities besides the ability to act for reasons. A rational subject is also a moral subject. Rationality involves a sensitivity to reasons for belief as well as to reasons for action. Ultimately, the practical reasons of a rational being cannot be made sense of without taking into account these further aspects of rationality, aspects which we do not find in subrational forms of life. These dramatic differences are obscured if we try to put our theory into the Procrustean bed of a theory of agency that preserves the continuity with mere animals.

These are compelling reasons to bite the bullet and to live with the fact

that a strictly High Brow conception entails that mere animals and infants fall outside the extension of intentional agency. Some may still find this classification mysterious. Two considerations may help dispel this air. First, we should bear in mind that denying concept use and intentional agency to non-rational beings is at least partly a terminological decision.[41] It is matter of a choice to use "concept" to mean an inferentially articulated capacity with extensive connections to rationality and to language. As I have argued, this decision is well-founded. Yet this doesn't entail that we should ban the use of the word in another sense which is more in line with the continuity strategy delineated above. There is of course nothing wrong with a use of the word "concept" that classifies non-rational animals or infants as rudimentary concept-users. The point now is simply that a use of the word with this more generous meaning would not be helpful to gain understanding of what it is to act intentionally. That here a more exclusive sense of "deploying a concept" is preferred does not mean that philosophers who have included animals among bearers of conceptual capacities have been wrong — they may also be engaging in a different discussion.

Second, although our topic is what is special about the conduct of rational beings, we can choose to treat *rational* agency as a species of a genus that includes as another species, say, the agency of sentient non-rational mammals. Grappling with this problem, John McDowell, in his *Mind and World*, takes the line of defending the idea that, although we share a sentient nature with mere animals, only creatures endowed with conceptual capacities, the inhabitants of the logical space of reasons, are capable of responsiveness to reasons.[42] In more recent writings, McDowell might seem to have changed his mind about crediting mere animals with the ability to respond to reasons.[43] The change, however, is only terminological. It allows us hew closer to common usage by countenancing, for example, that when a mere animal flees from a predator, we say that its reason is the danger it is in. We can see the action as a response to "something that is in an obvious sense a reason for it" and we can "represent the behavior as intelligible in the light of a reason for it" (McDowell 2009a: 128).

Despite the appearances, McDowell is not changing the substance of his position. He insists that, although animals can respond to reason, they cannot respond to reasons *as reasons*. This requires the ability, which animals lack, of stepping back from one's inclinations to ask whether the

[41] Cf. Sellars (1980a: 12).
[42] See McDowell (1996), in particular lecture VI.
[43] Cf. McDowell (2009a: 128ff).

consideration that inclines one to do something really is a good reason for doing so. McDowell's distinction allows us to see animal behavior and the intentional action of rational beings as two specific forms that responsiveness to reasons can take. In this way, we can acknowledge the superficial similarity between agency in the full sense of the term and what we might call animal agency while at the same time insisting that, when applied to rational beings, intentional action has presuppositions that are inevitably out of reach for non-discursive creatures. In this way, we can accept that animal agency does not involve sophisticated conceptual capacities, while insisting that in our case, the case that we are interested in, regarding an object or state of affairs as good, with all the presuppositions this implies, is part of the idea of acting for a reason.

If there is still a residual mystery about how the action of rational beings differs from mere animals in such a way that only the former, but not the latter, are inhabitants of the logical space of reasons, this is unsurprising, as it is a hard problem. However, one step towards dispelling the mystery would be an account of what a mere animal needs to learn in order to become attuned to that logical space. As we have seen, to give such an account is partly to explain the acquisition of linguistic abilities, i.e. how the subject's behavior comes to be patterned by the various ought-to-be rules that structure language.[44] As a final note, an account of this type may also be what Velleman is overlooking in his complaint about young children in the passage cited at the beginning of this section. There, Velleman suggests that, on a High Brow view, a child could not possibly acquire the conceptual capacities required to have desires conceived as evaluative judgments because, at the beginning of the process, it would lack any desires the teacher could leverage in his training program. But clearly the young child already has inclinations and is already responding to reasons, although it doesn't yet count as being practically committed or as being responsive to reasons *as such*. If the child comes to conform to the various ought-to-be's constitutive of its language and eventually learns to playing the game self-critically, that is precisely what must happen in order for his responsiveness to reason to transform, eventually, into a full-blown responsiveness to reasons as such.

[44]See §1. The more specific question of how a mere behaver becomes an agent, i.e. a bearer of practical commitments, is addressed below (§5.1).

Chapter 5

Internalism

5.1 Judgment internalism

The previous three chapters have been concerned with the motivation, development and defense of the High Brow conception of agency. Along the way, we have introduced the notion of a practical commitment and a few other concepts and ideas that help explain intentional action. In the remaining chapters, we will apply these tools to a number of philosophical problems that have been discussed extensively in the literature of the philosophy of action of the past decades. The hope is that the tools developed in our defense of High Brow will help us make progress on these fronts.

This chapter begins this task by tackling two entirely different views that have both been called "internalism": judgment internalism, on the one hand, and, on the other, existence internalism (i.e. Bernard Williams's internal reasons conception).[1] The internalist idea is that there is a necessary connection between reasons and motivation, but this idea has been developed quite differently by various writers. On the one hand, existence internalists hold that one cannot have a moral reason without being in a state of motivation that is related to the reason in a certain way.[2] Having a reason, on this Humean view, implies a desire-like state. This

[1]The word "internalism" is frequently used in other disciplines such as epistemology and the philosophy of language. Neither knowledge internalism nor semantic internalism have a significant relation to the class of views now considered, which belong to the field of "moral psychology" or action theory.

[2]See Darwall (1997).

first view functions as a constraint on reasons: it restricts the range of reasons we can assign to an agent. Existence externalists deny that such a condition exists: whether or not an agent has a reason does not, or at least not always, depend on his motivational state. On the other hand, as we have seen earlier, judgment internalism is the view that an agent's judgments about reasons are related to his or her motivation. This is not a view about what reasons an agent has but about how he typically reacts to judgments concerning those reasons. On this second view, one cannot candidly assert, or believe, that one ought to perform an action without being in fact motivated to do so.[3]

I begin this chapter by showing how High Brow can accommodate judgment internalism. We have already seen that judgment internalism amounts to an intuitive truth.[4] We have not yet seen, however, how High Brow can make sense of this intuition. So the first task is to provide an explanation of the internal connection between normative judgment and motivation (§1). Turning to existence internalism, I explain Bernard Williams's view that there are only internal reasons and briefly review his argument against externalism (§2). As other writers have noticed, this argument is defective since it presupposes much of what it is intended to show. But Williams's discussion also identifies two important features of reason statements. I explain how the High Brow conceptions developed in the previous chapters can accommodate these two features (§3).

An important *desideratum* for a theory of reasons is that it must be able to make sense of the intuitive truth that normative judgments are internally related to motivation.[5] Recall the intuitive truth we noted earlier:

Weak Judgment Internalism If X judges that he has a conclusive reason to φ (or that he ought to φ), then, *ceteris paribus*, he is motivated to do φ.[6]

High Brow needs to explain Weak Judgment Internalism. Moreover, as we have seen, it is not just *per accidens* that ought-judgments motivate.

[3]These two views are independent at least insofar as they do not strictly entail each other. Judgment internalism does not imply existence internalism as it is possible to hold that S cannot believe that she has a reason to φ without being motivated to do so while at the same time holding that a person may have a reason that is "external" in the sense of not being related to any preexisting motivation the agent may have. The converse implication does not hold, either. To say that reasons are, in some way or other, a function of the existing motivation does not entail that an agent who believes he ought to φ is necessarily so motivated.

[4]See §1.5.

[5]See §1.6.

[6]See §1.5.

According to internalism, the relation between the thought that one ought and the corresponding performance is a necessary connection. On the other hand, this does not mean that these thoughts move us with perfect reliability. As the "ceteris paribus" clause indicates, there are exceptions to this rule. The necessity of weak internalism, then, is not of the ordinary type. Although it is necessary that the agent acts on his ought judgments *for the most part*, a variety of conditions may nonetheless, in any given case, prevent the agent from realizing his thought about what he ought to do.[7]

On the theory proposed, the agent's judgment that he ought to φ is best seen as the undertaking of a practical commitment to φ. We may say that undertaking the commitment regularly *causes* the action. This is right as far as it goes. However, if we appeal to psychological regularity, internalism poses a twofold puzzle for the High Brow account:

1. *How can the connection be necessary?* Perhaps we can understand that reasons are related to motivation by reference to the fact that, through behavioral regularities, agents most of the time, though not always, react to their normative judgments with motivationally efficacious decisions. But how could an appeal to disposition vindicate the idea that the relation between reason-judgment and motivation is *necessary* or essential? It seems that behavioral regularities are just empirical generalizations incapable of supporting claims of necessity.

2. *Does High Brow collapse into a dispositional theory?* Above it was argued extensively that a dispositional account of desires fails. Now, however, it seems as if, in order to account for judgment internalism, the notion of a disposition or regularity has returned through the window. Doesn't the strategy pursued here collapse into a purely dispositional account? We may worry that the idea that intentional action requires practical commitments in a strong normative sense precludes us from helping ourselves to the idea of behavioral regularities in explaining the relation between normative judgment and motivation.

To solve the puzzle, let us start with the question of how the mental process behind normative motivation works. The answer will be roughly Sellarsian.[8] The process begins when the agent makes a normative judg-

[7] For a basic account of akratic behavior, see §4.2.

[8] Sellars expounds his view on the logic of intentions in a number of articles. See Sellars (1966), Sellars (1963b), Sellars (1980c), Sellars (1968: ch. 7) and, in particular, Sellars (1973).

ment, for instance the judgment that he ought to call a cab. As we have been insisting, the key to understanding practical thought is to understand the intentional state of practical commitment or intention. Thus we should also understand making an ought-judgment in terms of forming an intention. On this view, a normative judgment is a general kind of intention.[9] Naturally, not all ought-judgments are immediately realized. Even if I judge that I ought to call a cab, there may be other, more important things that claim my attention. What is more, I may not believe that I could in fact call a taxi if I tried, perhaps because it is physically impossible to do so at the moment. But let us suppose that the ought-judgment is both a conclusive and a practical, actionable judgment. The next obstacle is that I need to have an opinion about how to go about calling a cab. In other words, I need to engage in practical reasoning to determine a possible means. If successful, a practical argument yields the intention to, say, raise my arm.

On our Sellarsian view, the intention to raise my arm produces its practical effect, the bodily behavior, by turning into a volition, which is simply a type of intention with immediate relevance to action. Specifically, a volition is an intention to do something *here and now*. We can express intentions as thoughts involving the word or operator "shall". If the agent has the intention

> I shall raise my hand in 30 seconds.

soon the thought will gradually "mature":

> I shall raise my hand in 10 seconds.
> ...
> I shall raise my hand in 1 second.
> I shall raise my hand *here and now*.

[9] Thus Sellars proposes that *ought* "is a special case of shall*w*" (Sellars 1963b: 204). In other words, we should understand ought-judgments as the expressions of intentions, albeit in a special way. What is the difference between an "ought" judgment and a regular expression of intention? The main difference, according to Sellars, is that when we use a sentence containing the moral "ought", we are expressing a community-intention or *we*-intention. In the quote "shall*w*" refers to a variant of "shall" used to express a community intention. Sellars distinguishes between intentions in the first person singular and in the first person plural. The latter are intentions formed *as the member of a group*. On Sellars's Kantian account, the group extends — at least in the moral case — to the community of all rational beings. Sellars's most worked-out version of his ethical view is Sellars (1968: ch. 7). For a slightly different explication of the moral ought, see Sellars (1980c: §100, 92). Note that in the text, we are not necessarily talking about the moral "ought" but rather about the general "ought" of rationality. Nonetheless, it is plausible that even this "ought" can be understood in terms of intentions (perhaps in terms of *we*-intentions).

The result, the volition, is the final piece in the chain of mental occurrences leading to the action. Agents have a reliable disposition to react to a volition to φ by performing the bodily behavior referred to by φ. If all goes well, then, my normative judgment to call a cab leads me to form a volition which causes my raising my arm.

It should be clear that, even in such a simple case, the mental process leading to intentional action is complicated. Consequently, the logic of intention is also complex. Though we cannot explain the logic of "shall" comprehensively here, note that the word "intends" can take as its complement either a *that*-clause or an action-verb and that, conversely, the *shall* operator can operate on both propositions and action-verbs. The latter form is the one more directly connected with action, i.e. when the verb is a bodily action that can be performed directly, such as "I shall raise my arm" or "I shall walk one step to the right". Intentions of the propositional variety do not necessary involve any reference to the speaker, as in the intention expressed by the sentences "It shall be the case that income disparity is reduced." However, if these attitudes are intentions rather than mere wishes, they are linked to action-intentions. As they have consequences for the agent himself, we can understand them as "I shall bring it about that income disparity is reduced", which in turn, depending on further circumstances, may imply a more specific intention such as "I shall vote for Mr. X" and, ultimately, "I shall move my hand such as to write an X on the ballot." Moreover, intentions entail other intentions. As seen above, they do so when they stand to one another in means-end relationships. For example, depending on the circumstances, "I shall open the window" may imply "I shall move my arm in such and such a way". More general intentions, including those expressed by normative judgments, imply more specific ones — the ones leading directly to action. But there are also purely logical relationships. Thus, "I shall quit my job and become a saxophonist" implies "I shall quit my job".

These interrelated features, together with quite a few others, form a whole framework of intention. We have seen in earlier chapters that we can illuminate mental states by understanding them on the model of spontaneous linguistic acts.[10] Thus we can understand doxastic commitments, or beliefs, on the basis of our sound understanding of overt thinkings-outloud. Similarly, we can understand practical commitments, or intentions, on the basis of languagings involving the word "shall", or intendings-outloud.[11] The final element in the chain, the volition, may also be done

[10] See §3.2.

[11] As in the theoretical case, these pieces of linguistic behavior are acts in the sense of actualizations, rather than intentional linguistic actions that can be voluntary or involuntary.

aloud, in which case the performance is caused by a willing-out-loud. As a consequence, in order to count as proficient in the framework of intentions, a person must know the inferential relationships between the various types of sentences containing "shall". Becoming an agent implies learning successfully to navigate "shall" talk.

Returning to our description of normative motivation, we can note that the causal process leading from *X*'s judgment "I ought to φ" to his actually φ'ing can be broken into two majors parts: first the normative judgment causes a volition through a number of intermediate steps; then the volition brings about the bodily performance. Our question is what accounts for the necessary quality of internalism. Focusing on the second half of this two-part process, what ensures that agents who will-out-loud that they φ in fact actually do so? Note that there is a close causal relationship between a volition and actually being moved to performing the action in question. When an agent thinks out loud something to the extent of "I shall φ", we can reasonably expect his behavior to follow suit. In some cases, of course, the agent fails to realize his volition. The agent may be prevented externally from performing the action or he may be the victim of sudden paralysis or forgetfulness. Thus it is not strictly speaking true that whenever an agent intends to do something right now, he will do so. In keeping with weak internalism, we have to allow for exceptions.

The idea is not just that agents typically realize their intentions. An agent's intention to raise his arm *implies*, other things being equal, his raising his arm.[12] What does it mean in this context that the agent's thought *implies* his action? Clearly implication is more than mere uniformity of behavior. To see what the difference consists in, consider the sequence consisting of two events. One moment, the agent says "I shall raise my arm", a moment later he actually raises his arm. The agent responds to a mental occurrence by performing an action that affects the world through his bodily movement. How do we explain the second event in the sequence? The fact that he moves from the first episode, the volition, to the second, i.e. the performance of the action, is due to the fact that making this type of transition is obligatory according to the ought-to-be rules governing "shall" talk. To have one's behavior shaped by the obligatory character of the transition is to be aware, albeit in a practical, non-propositional way, of an entailment between the thought and the performance. That intentions entail, *ceteris paribus*, their realization is part of the logic of intention and is honored by anyone who knows the meaning of the word "shall" and its cognates.

See §4.1.

[12]Cf. Sellars (1980c).

Of course, a speaker sees the entailments between shall-involving sentences only if he has previously learned the framework of intentions. Our question, then, is what it takes for a fledgling speaker to learn the correct use of the word "shall" and its cognates. We can best approach this question by going through a sample learning process, though this, too, no doubt involves a great deal of idealization. Suppose a child, Ben, is for the first time learning the portion of its first language concerned with intention and has not yet mastered the correct use of the expression "shall". Through imitation induced by his parents, he starts making the noises we make when we say "I shall raise my arm". But because Ben doesn't yet understand the meaning of the term, it is not in fact correct to call Ben's performance a *saying* of "I shall raise my arm". Here it does not matter whether we characterize the utterance as reported speech, using the word "shall", or whether we quote it as direct speech using quotes. That Ben says "X" implies that he is using these words with understanding, which of course he still lacks at this point of our story. Thus saying that someone says something relies on certain assumptions about the speaker. Crucially, to characterize his sound as a saying presupposes linguistic competences on his part. Sellars writes:

> when we characterize a person's utterance by using a quotation, we are implying that the utterance is an instance of certain specific ways of functioning. For example, it would be absurd to say:
>
> > Tom said (as contrasted with uttered the noises) 'It is not raining' but has no propensity to avoid saying 'it is raining and it is not raining.'
>
> Thus to characterize a person's utterances by quoting sentences containing logical words is to imply that the corresponding sounds function properly in the verbal behavior in question and to imply that the uniformities characteristic of these ways of functioning are present in his sayings and proximate dispositions to say. (Sellars 1973: 198)

What Sellars says about "not" is true not only about "logical words" but words in general. The reason why it is not appropriate to credit Ben's utterances with meaning is similar to the reason why we do not so credit, say, a bird capable of speech imitation. Ben hasn't yet risen above the level of mere parroting because the way he employs the term does not conform to the rules associated with the term. Recall that, on the view proposed here, to use a word with meaning is to exhibit a certain pattern

in one's linguistic and non-linguistic behavior. To give an example, we may ask what the correct use of the word "red" implies. We can see this if we consider, conversely, what would cast doubt on the assumption that someone employing the noise or string of letters "red" is using it to mean what we mean by that expression. Here are pieces of evidence that would make us wary about attributing mastery of the concept:

- Standing in front of a blue surface, the speaker, whose perceptual capacities are assumed to be in good shape, tends to utter sincerely "This surface is red."
- The speaker says "This surface is uniformly red" but immediately afterwards adds "This surface is uniformly green" without indicating that he has had a change of mind.
- In many different situations in which he is prepared to say "This is red", the speaker is not prepared to say "This is colored."

Any of these incidents would make us question the assumption that the speaker's linguistic behavior is governed by the rules that are part of the concept of redness. Roughly, the rules that correspond to these examples are:

- You ought not to say of a clearly visible non-red surface that it is red.
- If you say that an object is uniformly red, then you ought not to say of the same object, at the same time, that it is uniformly blue.
- If you say that an object is red, you ought to assent to the assertion that the object is colored.

These rules, and many more, are constitutive of the meaning of red. If an agent doesn't show, at least to a certain extent, that he conforms to these rules, we eventually stop counting his linguistic performances as sayings of "This is red", even if he is still using the same sounds. As rules that govern linguistic behavior at its most basic level, these rules are ought-to-be rules that, though they aren't followed explicitly by the speaker, still guide his conceptual activity. As we have seen, we can understand these rules as governing moves in a language game.[13] A speaker starts in a certain position, say, the position which includes saying "This is red", and

[13] See §4.1

transitions to another position, which includes saying or explicit commitment to "This is colored". For our purposes, it is important that the moves prescribed (or prohibited) by the relevant *ought-to-be* rules can be classified according to their relation to extralinguistic reality:

1. *Language-world transitions*: In moves of this type, either the home position or the destination position is an extralinguistic state of affairs in the world. One move of this type is a language-entry transition, which moves from a non-linguistic perceptual situation to the taking up of a linguistic position. This happens with spontaneous observation judgments. Conversely, there are language-exit transitions that lead from a linguistic event to a (normally) non-linguistic piece of bodily behavior. Here the agent reacts to a volition by a piece of bodily behavior.[14]

2. *Language-language transitions*: In moves of this second type, both home position and destination position lie within language. The second and third rule above are examples of rules governing moves of this type. Intralinguistic transitions are inferences from one assertion or position inside the language game to another assertion or game position.

The meaning of words like "red" is collectively fixed by intralinguistic and language-world ought-to-be's. These rules are also what we expect to find active in a speaker's behavior when we characterize an utterance of his as a saying that incorporates the word "red". Like "red" talk, "shall" talk is governed by both inferential rules and language-world rules. Because "shall" figures prominently in the content of volitions, language-world transitions play a major role in fixing its meaning. A meaningful utterance involving "shall" enjoins a corresponding language-exit move. Thus a speaker who says "I shall φ here and now" ought to respond by producing the performance designated by "φ". The fact that this type of response is required is part of the concept "shall". Beyond language-exit transitions, however, the meaning also includes the propriety of intralinguistic moves. As we have pointed out, one can, and often must, infer "Shall[φ]" from "Shall[φ and ψ]". Moreover, there are various inferences that derive from means-end reasoning. In keeping with the remarks about propositional and action-verb intentions above, the logic of intention requires us to move from

[14]There are also linguistic actions or speech acts, where the result of the willing, though extralinguistic in our sense, has to do with language in another sense.

Shall[p]
If I φ, then p.

to

Shall[φ].

On the Sellarsian view, this intralinguistic inference is as much part of the meaning of the word "shall" as the language-exit transitions. Together, these rules make up the functional role of the word.

We can now see what ensures that agents do φ after they will-out-loud to φ. To master the framework of intentions is to have one's tokenings of "shall" governed by, among other rules, a language-exit ought-to-be. This creates a regularity: people who have learned the language of intention have the disposition to go from willing to action. However, appealing to a disposition may cause the worry that this idea throws us back into a dispositionalist conception of intention. Yet the worry is unfounded. It is true that in learning the use of practical vocabulary, associations are formed in an agent's mind. However, the regular connection between volition and action is not a *mere* association. By contrast to mere associations, we can explain the transition by reference to a relation of entailment between the two episodes. The relation is more than a mere reflex of the mind, as the agent proceeds to raise his arm after willing to raise his arm because he *sees* that the intention necessitates or entails acting accordingly. The language-exit move is governed by an ought-to-be rule that says that you ought to do as you intend to. The agent's move from volition to action, then, is not merely a matter of disposition, although a disposition is obviously involved. In making the move, the agent is guided by the relevant entailment and thus by an *ought-to-be* rule.

It is important to emphasize that the rules in question are rules of criticism, or ought-to-be's, rather than rules of actions, or ought-to-do's. As we have seen, this means that it doesn't follow from the fact that a person's linguistic move is guided by a rule that he is explicitly envisaging the rule or that he is aware of it in a propositional form. The rule, as an expression, fulfills its guiding role in the background. Only the language teacher, while he is instilling the relevant propensities in the fledgling speaker, is explicitly following the rule-statement or, to be more exact, the pedagogic rule of action that is derived from it.[15] The speaker's rule-governed behavior is guided by a pattern purposefully inculcated in him

[15] §4.1.

by his trainer. The speaker makes the move *because* of the relations of entailment, since owing to the history of the acquisition of the terms in question, the rule is part of the complete account of the move.

The behavioral uniformities that are acquired alongside linguistic competence are not *brute* dispositions. Instead, the regularities are what we may call *rational dispositions* because the speaker comes to exhibit the pattern-governed behavior in question in the process of being introduced into the social practices that are constitutive of participation in rationality. For this reason, there is no danger that the position described could collapse into a version of causal functionalism. The idea that practical commitments are accompanied by dispositions is perfectly compatible with the claim that what determine their content are their rational or normative links. Another significant respect in which the dispositions that ground concept-use are rational is that they are part of an ongoing collective enterprise and always remain potential objects of rational criticism. A move in a language game can be regarded as inappropriate or challenged. Importantly, such criticism can target not just the linguistic behavior of another speaker but one's own behavior as well. Noticing a pattern in his inferences or extra-linguistic move, the agent may step back from this disposition to ask whether the transition is in fact appropriate. According to Sellars, the ability to be self-critical in this way is the mark of a mature speaker or agent.[16] In McDowell's view, this ability is part of what transforms us into active thinkers.[17] By engaging in critical reflection about one's own conceptual rules, an agent emancipates himself from the kind of brute dispositions we find in trained animals.[18]

With this picture of "shall" talk at hand, we can return to solving the puzzle raised by internalism. As we have seen, Ben's transition from mere behaver to intentional agent is effected by his acquisition of the framework of intention. When he has learned to make the "shall" noises, we can give the following description of his behavior:

> If Ben makes the noise "I shall raise my arm now", he is *ceteris paribus* going to raise his arm.

At this point in the story, the description is a mere causal generalization. His teachers have trained him to react to a parroting in a certain way. But because he doesn't honor the full range of rules yet, he does not yet count

[16]Sellars (2007d: 86).

[17]McDowell (1996: 47).

[18]See also Sellars's distinction between tied and free symbols, between "tied symbol behavior" and free or "rule-regulated symbol activity" (Sellars 1980b: 301).

as *saying* "shall" yet. The connection between the linguistic performance and the doing may have causal or nomological necessity, but it is logically contingent. Fast-forward a few training sessions, Ben has now learned the uniformities enjoined by the inferential and language-world ought-to-be's to an appropriate degree. We can now redescribe his linguistic performance:

> If Ben says "I shall raise my arm now", he is *ceteris paribus* going to raise his arm.

Under the new description, the statement has a new modal status. What used to be a conceptually contingent generalization has become a conceptual truth.[19] The reason is that by ascribing to Ben a saying of shall, rather than a mere parroting, we implicitly presuppose that the agent has the relevant linguistic propensities. It follows that the principle that, absent any preventing factors, an agent's volition to raise his arm is followed by his actually raising his arm, is no longer logically contingent because in ascribing the intentional state and, hence, the accompanying linguistic competence to the agent, we presuppose that, other things being equal, he is going to act on his volition.

Our original puzzlement was about the relation between ought-judgments and motivation. Above we divided this question into two parts: the link between ought-judgments and volitions, and the link between volition and actual motivation. So far we have been concerned with the latter half but, fortunately, the reflections relating to the internalism of "shall" bear directly on the internalism of "ought" as well. As we have noted, judgments about one's own reasons are tied inferentially to intentions. Roughly, judging that one ought to do φ is a general form of committing oneself practically to φ'ing, which in turn entail other commitments. As a result, if the circumstances are right, the judgment that I ought to raise my hand entails the intention of doing so. Again, this inferential relation forms part of the logic of "shall" talk and "ought" talk: it is part of what we learn when we learn to use these words. It follows that the propensity to make the appropriate moves is something we can take for granted when we attribute to an agent that he makes a normative judgment, i.e. uses the word "ought". Willing and willed performance are related by a more-than-accidental connection. The same goes for the connection between ought-judgment and volition. Because they are inferentially related, we presuppose, when describing the agent's utterance as a saying of "shall", that the speaker is prepared to honor these entailments.

[19] Cf. Sellars (1973: 204).

If this is true, we have, at least in outline, an account of intentions that explains how the internalist principle is true, and non-accidentally so: the systematic relationship between reason-judgment and motivation, which can be explained in terms of the agent's history of language acquisition, is ultimately a matter of conceptual necessity. We can also see how this necessity does not in any way require that the systematic relation is strict. Mastering the concept behind the word "shall" requires having a disposition to respond to volitions by actual motivation to produce the required performance. However, it does not require that the link is perfect. This is expressed by the *ceteris paribus* clause: it is only *other things being equal* that the agent performs the action after forming the volition.[20] Rational links of this type can, and do, break down, a situation which occurs in cases of weakness of the will. In such cases, although the agent should, according to the rules that govern his concepts, show an extralinguistic response, he doesn't actually do it. But far from being surprising or incompatible with the view that these dispositions are a matter of the meaning of the word "shall", the possibility of failures is to be expected. The dispositions are rational dispositions that impose requirements on the agents, and they would not be requirements in this emphatic sense if it was not possible for the agent to flout them by failing to make the move enjoined by the ought-to-be. True, a speaker who consistently violated the constitutive conceptual norms would eventually stop counting as using the word "shall" in the standard way. A global lack of conformity with the conceptual norms is in fact hard to imagine, but this does not touch the possibility of a local breach of the norm. A volition that doesn't motivate is a mark not of having forgotten the meaning of the word "shall" but of a local instance of irrationality.

5.2 Williams on internal and external reasons

Let us now leave judgment internalism behind and turn to existence internalism. The view is formulated and defended forcefully by Bernard Williams in a number of papers.[21] Williams is concerned with the practice of attributing reasons. He holds that there are significant constraints on the normative reasons we can attribute to an agent. Take a normative reason statement:

(NR) X has a reason to do φ.

[20] See §1.5.

[21] Williams (1981a), Williams (1995), Williams (2001).

What does *NR* mean? According to Williams, there are two ways we can read this statement. On the first interpretation, *NR* implies the existence of an element in what Williams calls the agent's motivational set *S*. The agent's motivational set comprises various conative psychological states, notably desires.[22] On the internal interpretation, *NR* entails that *X*'s motivational set includes such a "motive" which would be promoted by the agent's φ'ing. Conversely, on this interpretation *NR* is "falsified by the absence of an appropriate motive" (Williams 1981a: 101). By contrast, on the external interpretation, *NR* does not carry the implication that there is an element in *S* that is appropriately related to φ'ing. On this reading, NR may be true even if there is no "sound deliberative route" from *X*'s *S* to the decision to do φ (Williams 1981b: 120).

The two interpretations are mutually exclusive, which leads to two opposing views. According to externalism, the second interpretation is true. There are "external reasons" as well as "internal reasons": a reason-statement may be true of an agent despite the absence of an appropriate element of *S*. Williams's preferred view, internalism, denies this. If the internal interpretation of *NR* is true, all external reason statements are false. If internalism is true, all reasons for action are internal; there are no external reasons.

If true, internalism imposes significant constraints on normative reason statements as it renders some reason statements false. As an illustration, Williams provides an example taken from Britten's opera "Owen Wingrave". Against the wishes of his father, Owen refuses to join the army. He doesn't seek honor, nor is he especially keen to defend his country. In order to sway him, his father (we may imagine) tells him that he has a reason to join the military, viz. the fact that the men in his family have always done so. By contrast, Owen doesn't find this argument compelling at all. He could not care less about his family tradition: there is no desire to honor the family tradition in his subjective motivational set *S*. It follows that the father's statement is an external reason claim. But according to internalism, such statements are never true. The father doesn't necessarily care whether or not the son has an existing motivation — he

[22] According to Williams, the subjective motivational set *S* can contain desires but also "such things as dispositions of evaluations, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent" (Williams 1981a: 105). Clearly, however, desires are the paradigmatic members of the set, as can be seen from the sentence that follows: "Above all, there is of course no supposition that the desires or projects of an agent have to be egoistic; he will, one hopes, have non-egoistic projects of various kinds, and these equally can provide internal reasons for action" (Williams 1981a: 105). Williams later uses the letter "D" to refer to elements of the set.

doesn't take his claim to imply anything to this effect. According to the father, no matter what the son's present motivational state consists of, he has a reason to enlist. If we can make sense of the statement as intended by the father, we have a vindication of external reasons. But Williams takes this to be impossible. We can't make sense of the statement as an external reasons claim.

Williams's internalism ties normative reasons tightly to the agent's motives in the narrow sense which does not include beliefs or doxastic commitments. On this picture, there are no reasons unless they are anchored in the agent's motivational set and desires and other members of *S* must be the source of our reasons. Though we will review Williams's argument against externalism below, the primary focus of this chapter will not be on a criticism of Williams's internalism.[23] Rather we will try to bring into focus how Williams's nuanced positive position diverges from a simplistic Humean model. Williams pays particular attention to a number of cases and special features of reason-attributions that may be used in arguments against internalism. In responding to these worries by adjusting his position, Williams supplies cues for a conception of normative reasons that is quite different from the expectations his initial characterization of the view may create.

Let us see how Williams applies spit and polish to his view. He writes that "any model for the internal interpretation must display a relativity of the reason statement to the agent's *subjective motivational set*" (Williams 1981a: 102). However, although internal reasons are relative to *S*, that does not mean that reasons are identical to elements of *S*. According to Williams, we cannot simply read off the agent's reasons from his motives. We can see this if we remember that, while we are usually well aware of the contents of our motivational set, we do not necessarily know our internal reasons with perfect accuracy. Williams concedes that thinking that one has a reason to φ is neither a necessary nor a sufficient condition of actually having a reason to φ. To begin with the latter, Williams's famous gin-and-tonic example is intended to illustrate the fact that an agent may believe that he has a reason to φ while in fact having no reason at all to do so. The agent in the example would like a glass of gin and tonic. This leads him to mix tonic water with the liquid in the glass in front of him, which he takes to be gin. As it turns out, he is mistaken; the liquid is actually petrol. Mixing the drink, he takes himself to have a reason to do this. But does he really have a reason? Williams denies this. The natural

[23] See the discussion of some positive reasons for Humeanism in ch. 2. A number of commentators have already correctly pointed out difficulties with Williams's argument. See Cohon (1986), McDowell (1998c) and Millgram (1996).

thing to say is that the agent did not in fact have any reason to mix the liquids or to drink the result, although he thought he did.

But Williams contemplates seriously opting for the contrary answer that the gin lover *does* have a reason to combine the liquids. Here he is pulled in two opposite directions. On the one hand, it makes sense to say that the agent has a reason, given that he in fact mixes the awful drink. The action requires explanation, as the agent certainly does not act for no reason at all. The ascription of a reason helps account for the agent's behavior. This aspect of reasons, which Williams calls its explanatory dimension, counts in favor of a positive answer to the question that the gin lover has a reason to mix the liquids.[24] As Williams emphasizes, our practice of assigning reasons to agents relies on the assumption that sometimes people act for these reason. But there is also a tendency that pulls in the opposite direction. Williams appeals to our linguistic intuition that it is "very odd" to describe the gin-and-tonic case as one in which the agent has a reason (Williams 1981a: 102). On the strength of the intuition behind this verdict, Williams ends by choosing not to assign the agent a reason. According to him, to consider this aspect of reason claims is to consider the agent as a rational being. As Williams puts it, "the internal reason conception is concerned with the agent's rationality" (Williams 1981a: 103). He says it is a mistake made by opponents of his view to think that the conception is concerned exclusively with explanation.

We can already notice two features of Williams's account. The fact that Williams is willing to say that the gin lover has no reason to mix the drink shows that, in fixing the agent's reasons, he is prepared to abstract away from mistaken beliefs on the part of the agent. On his view, an answer to the question what an agent has reason to do depends to some degree on seeing the agent as he should be rather than as he in fact is. Williams's theory of internal reasons, then, involves *idealization*. Second, from the fact that Williams contemplates both a positive and a negative answer, we can see that in edge cases, the determination of an agent's reasons may be a matter of dispute which has no clear-cut answer. His conception makes room for *variability* of reason-ascriptions.

Turning now to the second way in which there may be a gap between the agent's view of his reasons and the reality, thinking that one has a reason is not necessary for having a reason, either. Williams concedes that it is possible to have a reason without knowing it. As an example, suppose a motorist who prefers to stay safe on the road uses summer tires on a mid-winter trip through the Alps because he doesn't know that the weather

[24]Williams (1981a: 102).

conditions require snow tires. Here we say — and Williams presumably will agree — that despite his factual ignorance the driver has a reason to drive on winter tires. Again, our decision whether or not the agent has a reason depends to a certain extent on our seeing the agent as a rational subject. Determining the agent's reasons involves not only removing faulty empirical views but also adding true beliefs. Again, what we are seeing in the example is the effect of the rationality aspect of reasons. Williams's policy for finding true reason statements involves approximating the agent to a certain ideal of rationality.

In the Alpine tourist example, the driver can appropriately be said to have a reason because he is, as Williams puts it, "ignorant of some fact such that if he did know it he would, in virtue of some element in S, be disposed to φ" (Williams 1981a: 103). However, although Williams countenances idealization in this way, he is not willing to permit it across the board. Thus he writes:

> For it to be the case that he actually has such a reason [without knowing it], it seems that the relevance of the unknown fact to his action has to be fairly close and immediate; otherwise on merely says that [the agent] A would have a reason to φ if he knew the fact.

We may say that the driver's driving on summer tires is explained by his ignorance of the danger this brings. Clearly, however, Williams thinks that sometimes one can lack a reason to φ even though there is some piece of factual information p such that one would be moved to φ if one knew that p. It is less clear what kind of example would count as one where the unknown fact is not, in his sense, closely and immediately relevant to the action.[25] Perhaps the following example will do. Suppose a land owner is interested in material wealth. Suppose further that, unbeknown to him, there is an oil source in his back yard. We could express Williams's point by saying that the land owner only has a hypothetical reason. If the man knew about the oil, surely he *would have* a reason to make preparations for drilling. But given that he doesn't have knowledge of that fact, we say that he in fact doesn't have a reason.

This second way in which a gap between having a reason and knowing it can appear shows the element of *idealization* in reason ascriptions. But

[25]Williams writes cryptically that the conditions "must be closely connected with the question of when the ignorance forms part of the explanation of what A actually does" (Williams 1981a: 103). This does not seem to help us understand what he means.

the oil-drilling example also again highlights the *variability* of such ascriptions. The question of which reasons an agent has is not as straightforward as we may have expected. On Williams's view, whether an ascription is appropriate depends on a judgment as to how closely the unknown fact p is to the φ'ing. Williams does not seem to think it is a problem that the set of reasons an agent has is less determinate than we may ordinarily think.[26] A similar variability is at display in his conception of what constitutes the "sound deliberative route" from the subjective motivational set to the proposed action. Simple means-end reasoning counts as correct practical reasoning, of course, but according to Williams, we need a more inclusive conception of correct deliberation. Thus he argues that "finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment", though not a species of instrumental reasoning, is still a sound deliberative route (Williams 1981a: 104). What is more, the imagination may play an important role in the process. He concludes:

> We should not [...] think of S as statically given. The processes of deliberation can have all sorts of effect on S, and this is a fact which a theory of internal reasons should be very happy to accommodate. (Williams 1981a: 105)

There are, then, no clear limits on what processes of deliberation count as rational transitions or on the contents of the agent's motivational set. This suggests that verdicts about reasons are variable in this sense, too. In borderline cases, assessors may disagree as to whether an internal reason statement is appropriate.

To summarize, from Williams's point of view, in order to arrive at correct reason judgments, we need to idealize the agent's attitudes by making different sorts of corrections. What is more, the assessments may vary considerably and depend on factors other than the agent's own attitudes, such as pragmatic considerations. Although this point is not explicitly stated, we can see Williams as proposing as tasks for adequate theories of reasons to account for both the idealization and the variability aspect of normative reasons. In the next section, we will explore the ramifications of these points. For now, however, we have to take a look at Williams's argument against externalism. The argument has two parts, which we can see as corresponding to the two dimensions of reasons mentioned above. The (a) explanatory dimension is this: "If there are reasons for

[26] Cf. Cohon (1986: 555).

action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action" (Williams 1981a: 102). The (b) rational dimension is this: If something is a reason for action, then it must show that the agent is behaving rationally. The agent must be able to acknowledge that he has the reason as a result of rational deliberation.[27]

The first part of Williams's argument exploits the explanatory dimension. Picking up the example of Owen Wingrave, he writes:

> Now no external reason statement could *by itself* offer an explanation of anyone's action. Even if it were true (whatever that might turn out to mean) that there was a reason for Owen to join the army, that fact by itself would never explain anything that Owen did, not even his joining the army. For if it was true at all, it was true when Owen was not motivated to join the army. The whole point of external reason statements is that they can be true independently of the agent's motivations. But nothing can explain an agent's (intentional) action except something that motivates him so to act. (Williams 1981a: 107)

Williams points out that if Owen has an external reason, then this should be able to explain his action, and on his view, something can only explain an intentional action if it motivates the action. For Williams, Owen's external reason R cannot possibly be "something that motivates him to act" because externalism implies that external reasons are independent of Owen's motivational set. His reasoning is this: R is not a member of S so it cannot be something that motivates the agent to act. Hence, we must interpret Williams's principle that "nothing can explain [his] action except something that motivates him so to act" as "nothing can explain [his] action except an element in his motivational set". However, this is a controversial principle, to say the least.[28] In fact, assuming that only members of the agent's motivational set can explain the action seems to assume the very point Williams is trying to argue for: the idea that there are only internal reasons.

Williams seems to assume from the outset that only internal reasons can make sense of the explanatory dimension of reasons. If so, Williams is

[27] The distinction of the two general features of reasons is taken with minor modifications from Cohon (1986: 545–6).

[28] As Cohon (1986: 547–9) makes clear. I follow Cohon's presentation of the issue closely.

begging the question against the externalist. However, Williams's argument also has a second part, which is based on the rational dimension of reasons. This part has the form of a *reductio*: Suppose that Owen actually has an external reason to join the army. Although he has a reason, then, he lacks the relevant element in his *S*. It follows that there must be some "psychological link" that explains how he comes to be moved to do so. This link must be that, prompted by his father, Owen comes to believe that his family tradition is a reason for him to enlist. Williams grants that, once the agent has developed this belief, he is moved to act accordingly. However, he asks how it comes about that the agent comes to have the belief that he has a reason to act. He writes:

> The basic point lies in recognizing that the external reasons theorist must conceive *in a special way* the connexion between acquiring a motivation and coming to believe the reason statement. For of course there are various means by which the agent could come to have the motivation and also to believe the reason statement, but which are the wrong kind of means to interest the external reasons theorist. Owen might be so persuaded by his father's moving rhetoric that he acquired both the motivation and the belief. But this excludes an element which the external reason theorist essentially wants, that the agent should acquire the motivation *because* he comes to believe the reason statement, and that he should do the latter, moreover, because, in some way, he is considering the matter aright. (Williams 1981a: 108–9)

What Williams is working with here is a contrast between two ways in which an agent may come to believe the external reason statement advanced by his father (that p). On the one hand, the agent may acquire the belief that p by brute rhetorical force, or perhaps through a blow to his head. If he does, he does not acquire the belief through rational means. But as the rationality condition (b) above highlights, it is one function of reasons for action to reveal the rationality of agents.[29] If the agent comes to believe p in a brute fashion, this important condition is violated. This way of becoming motivated does not validate that, before Owen developed the motivation, he had an external *reason* to join the army. The externalist cannot base his claim that p is a reason on this non-rational way of becoming motivated.

On the other hand, of course, there are rational ways of coming to believe reason statements and thus of becoming motivated. Williams ex-

[29] Cf. Cohon (1986: 547).

plains that the only way for the acquisition to be appropriate is for it to occur under the conditions of correct deliberation.[30] For Williams that means that the original external reason statement must entail that "if the agent deliberated correctly, then, whatever motivation he originally had, he would come to be motivated to φ" (Williams 1981a: 109). However, Williams argues that the agent could not possibly acquire the motivation simply through correct rational deliberation. For we have assumed that the agent does not have a relevant element in his motivational set to deliberate *from*. For Williams, Owen could not acquire the motivation to join the army because he cannot reach that motivation by "seeing things aright", which he regards as necessarily involving practical reasoning which takes his existing motivation as its point of departure. For Williams, the externalist faces the impossible task of showing how the new motivation was arrived at rationally, yet

> at the same time it must not bear to the earlier motivation the kind of rational relation which we considered in the earlier discussion of deliberation — for in that case an internal reason statement would have been true in the first place. (Williams 1981a: 109)

Williams concludes that these conditions cannot be met simultaneously. Now he is certainly correct that the defender of the claim that Owen has an external reason to φ must show that the agent can reach the decision to φ in some way that is not blatantly irrational, i.e. in a way other than using a blow to the head. That is, he must show that there is a ground for attributing the reason to the agent.[31] Owen's father cannot simply invent reasons for the agent — as we might say, they must be in some way "anchored" in the agent. On the other hand, Williams takes it for granted that the agent could rationally be moved to φ'ing only if he engages in the kind of instrumental deliberation he recognizes as valid practical reasoning.[32] The idea is that no rational process could generate a new motivation, given that no prior motivation exists from which it could be derived.

[30] Cf. Cohon (1986: 547).

[31] Cf. Cohon (1986: 554).

[32] As noted above, Williams accepts other reasonings than simple means-end reasoning as valid practical reasoning, including finding what is constitutive of a given aim. But like means-end reasoning, constitutive reasoning is also dependent on a preexisting end or motivation. In this respect, Williams's slightly extended conception of deliberation is no different from the classical instrumentalist view.

5.3 Ideal and variable reasons-ascriptions

We are now in a position to see that Williams's argument relies on a controversial understanding of what counts as a rational transition to a new motivation. Williams holds that the only way reason can produce a new element in S is by deriving it from another preexisting member of S. Jay Wallace has usefully called this interpretation of practical reasoning the desire-out-desire-in principle: no new element can come about unless it build on an earlier one.[33] Williams demands, correctly, that a theory of reasons must take into account the rationality aspect of reasons. It must be possible to react to an external reason by acquiring the corresponding motivation rationally. But we need not accept Williams's restrictive conception of what constitutes an exercise of rationality.

We have already extensively criticized the Humean conception of rationality.[34] The alternative, High Brow conception we have developed does not restrict rational deliberation in the way envisaged by Williams. According to High Brow, to have a normative reason is to have a practical commitment to φ. Practical arguments are inferential moves from one practical commitment to another. What transitions are rational is determined by the judgments of the assessors of our inferential practice. We have seen that practical commitments are not immune to criticism. In a similar vein, we should not rule out the possibility that the agent may acquire a practical commitment without deriving it from a prior practical commitment.[35] There is much to learn about the ways such inferential practices work. However, the point now is simply that Williams leaves his assumption about the appropriate shape of rational deliberation unsupported by argument. The desire-out-desire-in principle seems to be little more than a reflection of his initial conviction that internal reasons do not exist. Williams's situation, then, is similar to the one he finds himself in with respect to the first part of his argument. To assume that (broadly) instrumental reasoning is the only way to come to have motivations is to assume a large part of what he is trying to prove.

Let us return to the positive payoff we extracted from Williams's treatment of our practice of attributing practical reasons. We noted above

[33] Wallace (2006a: 30).

[34] See ch. 3–4.

[35] In §4.4, we have already noted an example of a valid inference from a theoretical starting point to a practical conclusion: "It is 5 p.m. Thus, I shall have the 5 o'clock tea now.". The existence of examples of this type seems to contradict Williams's idea that new rational motivation needs to be derived from some element in the agent's subjective motivational set.

that, despite his Humean roots, Williams countenances a certain degree of *idealization* in reason-ascriptions and that he appears receptive to the idea that such ascriptions are not entirely determinate, i.e. that they display *variability*. Now, as we said, High Brow sees having a normative reason to φ as being practically committed to φ'ing. This may give rise to the worry that High Brow leaves idealization and variability out of the picture: either you are committed to φ or not. To put it differently, given that practical commitments were equated with intention, either you intend to φ or you don't. Does this picture leave room for the two features we found in Williams's article?

When we attribute an external reason or when we attribute an internal reason that the agent is not aware of, we are essentially criticizing the agent's present assessment as to what he ought to do. We have already noted that practical commitments are amenable to rational criticism, but we can now describe in more detail how such criticism works.[36] To begin with the problem of how the attribution of intentions can be subject to the sort of idealization Williams admits, we should note that crediting a speaker with an intentional state is not as straightforward as it may seem first sight. Thus according to Daniel Dennett, attributing belief is what we do when we start seeing the behavior of a creature as understandable through rational explanation involving intentional states.[37] For Dennett, in order to be properly regarded as an intentional system, an organism needs to fulfill a number of criteria. On Dennett's view intentional systems often have the ability to communicate verbally, but he does not regard speaking a language as a requirement for counting as an intentional system.

Obviously these creatures cannot tell us what they are thinking, yet we can still explain what they do by positing beliefs and intentions. We cannot read their minds, so for these creatures at least the sole basis we have for attributing intentional states is their observable behavior.[38] Thus when see a creature standing under the tree rather than in the rain, we can attribute to it the intention not to get wet. However, it is immediately clear that this explanatory strategy is hardly foolproof. For our evidence that the creature has the intention we attribute to it is its staying under the tree. But the creature's intention to stay dry is relevant to the account only on the further assumption that it also believes that the tree gives it shelter. What would be evidence that the creature believe that the tree

[36]See §3.5.

[37]Cf. Dennett (1981: 16–22).

[38]In fact, according to Dennett, even direct neurophysiological access to their brains would be of no use. Cf. Dennett (1981: 21).

shelters it from the rain? We can certainly attribute this belief to it if it stays under the tree when it is raining. However, the belief explains its conduct only on the assumption that it also intends to stay dry.

In other words, attribution of intentional states on a behavioral basis occurs under delicate conditions: they are caught up in a circle of implications between intentional concepts.[39] We cannot single out one intentional state and explain it individually in non-intentional terms. Instead, intentional states hang together in such a way that they resist a safe and entirely reliable procedure of identification: we can only attribute the belief that p conditionally, on the assumption that the creature also has the intention to φ; and we can attribute the intention to φ only conditionally, on the assumption that the creature also has the belief that q; and so on. Now of course, the creatures we are interested in — and the ones Dennett is most interested in — have a greater variety of giving expression to their intentional states. Persons not only exhibit their beliefs and intentions behaviorally but also produce linguistic performances that reveal what they think. Thus it may be thought that the intentional circle does not apply to speaking creatures: when in doubt about their beliefs and intentions, we simply ask them. However, this turns out to be an illusion. It is true that the speaker's utterance "The tree gives me shelter" is good evidence that he in fact believes that it does. However, the evidence is conditional on the assumption that the agent's utterance is candid, that the agent does not want to deceive us, and so on. Furthermore, we need to assume that we know what the speaker means by his utterance, which again presupposes a number of further assumptions. If anything, interpreting the mind of a speaker is more challenging than explaining that of a silent creature.

We need to make the best of our situation as interpreters of the minds of others. On the other hand, it is true that we could hardly ask for a better source of evidence for attributions of beliefs and desires than the person's words. For Dennett, avowals — i.e., in the Sellarsian terms introduced above, intendings-out-loud and thinkings-out-loud — are also what allow us to break into the intentional circle. There are, on his view, two preconditions of a speaker's being a rational system: (1) "normally, more often than not, if x believes that p, p is true" and (2) "normally, more often than not, if x avows that p, then he believes that p" (Dennett 1981: 18). Unless these prerequisites are met, the creature we are dealing with cannot count as a rational believer (or intender). The agent's beliefs about his environment are, by and large, true; and what he says is, by and large, what he thinks. We can make use of this assumption to establish

[39] Dennett (1981: 19).

a basic interpretation of the agent's mind. We can in general say that the agent's assertions reveal his beliefs *unconditionally*. Most of the time, the agent's actions indicate what he intends. As Dennett writes, "we get around the "privacy" of beliefs and desires by recognizing that in general anyone's beliefs and desires must be those that he "ought to have", given the circumstances" (Dennett 1981: 19). It should be clear, however, that while this is true as a general rule or globally, it need not be true locally. An adequate theory of interpretation must allow that, in any particular case, an agent's belief may deviate from the truth, his desires from what is obvious.

In other words, even if we can assume, as Dennett urges, that the agent's beliefs are largely correct, we still need to carry out the task of identifying attributing beliefs manually in any particular case. As we said, we can draw on two sources of evidence: behavioral and verbal. Regarding the first, when we attribute a belief to an agent because of something he does, we very often have to attribute to him a true belief. In doing so, we are guided by what Dennett calls the rationality norm. According to this norm, we ought to favor attributing to the agent true beliefs and also rational beliefs, i.e. beliefs that are not incoherent. Roughly: When in doubt, do not assume the agent is mistaken or contradicts himself! On the other hand, when we base our attributions on the agent's utterances, we are guided by what he calls the accuracy-of-avowal norm. Roughly: When in doubt, avoid an interpretation according to which the agent says something he doesn't sincerely believe!

As Dennett writes, these norms often work in a complementary way, but this need not always be the case. On occasion an agent produces an utterance that gives rise to a conflict between the two norms.[40] Suppose what a speaker says is clearly contradicted by the empirical evidence, perhaps evidence that the agent himself is aware of. We have two ways of dealing with such a situation. On the one hand, we may, as Dennett puts it, "lean on the myth that a man is perfectly rational" (Dennett 1981: 20). In other words, we may assume that the agent does not in fact mean what he is saying. That is, we may choose not to attribute the belief the utterance is evidence for on the grounds that it would be incompatible with other beliefs he has already expressed. However, to do so, we must concede that the norm noted above — that most of the time assertoric utterances are genuine avowals — fails in this particular instance. On the other hand, we may "lean on the [the agent's] right as a speaker to have his word accepted" (Dennett 1981: 20). In doing so, we cling to the norm

[40]Dennett (1981: 19).

that utterances are genuine avowals, but we have to flout the norm not to attribute to the speaker incoherent combinations of intentional attitudes.

For Dennett, we often cannot achieve complete conformance to both norms — we either have to opt not to take the speaker by his words or assume that his beliefs are at least partially irrational. As Dennett writes, neither position "provides a stable resting place" (Dennett 1981: 20). In particular, an ascription of irrationality can be problematic since, as we have seen, the coherence of his attitudes is a condition of the possibility to make sense of the agent. If the agent crosses a certain threshold of irrationality, we would be forced to stop treating the creature as an intentional system altogether. But we can tolerate a certain amount of incoherence without shedding the intentional stance.

Thus when we say of a person that he has a belief, we potentially confront conflicts of this type. In this connection, Robert Brandom has noted that the word "belief" is ambiguous.[41] In the first sense of the word, attributing a belief implies that the target of the ascription avows the belief or is prepared to do so when prompted and the situation is appropriate. A belief in this sense is obvious to the subject that has the belief. But in another sense, it may well come as a surprise to the subject that he has the belief attributed to him. In this sense, even though the subject believes that p, he may not know that he does. Beliefs in this sense are not subconscious beliefs but rather inferential extensions of avowed beliefs held by the speaker. Whether the subject is consciously aware of the belief or not, it can be ascribed to him on the basis of other commitments he acknowledges.

Suppose a person concedes both that $\neg q$ and that if p, then q. This person, in a way, has also admitted that $\neg p$. Think of the remark, often heard in debates, that "you have just admitted that ..." In this case, $\neg p$ is the logical consequence of beliefs the subject has expressed. Still he may not have noticed that the result follows logically from propositions he holds true. Or he may not be prepared to admit the consequence due to a psychological hang-up. In the avowed sense he can be said *not* to admit or believe the proposition in question. But this should not be confused with the idea that in *another* sense he does believe or admit the proposition. In other words, it is possible for the subject to believe $\neg p$ in the latter, rational or ideal sense, and nonetheless fail to believe $\neg p$ in the first, conscious sense which implies preparedness to avow.

According to Brandom, it is an advantage of his normative scorekeeping account that the ambiguous notion of belief can be replaced by the univo-

[41] Brandom (1994: 193–6).

cal concept of a doxastic commitment.[42] On his view, undertaking a doxastic commitment (forming a belief) is to be understood as a genus comprising two different positions in the game of giving and asking for reasons. It may be understood as the agent taking up the normative attitude of attributing the commitment that p to himself. Undertaking the commitment directly in this way implies a preparedness to avow the claim. But for Brandom, undertaking a commitment is not a species of attributing the commitment to oneself; in fact, the converse is true. Thus one way to undertake the doxastic commitment that p is indirectly, by undertaking another commitment whose inferential consequences include the claim that p. An agent may be correctly taken as *really* committed to p even though he lacks the disposition to acknowledge it in the appropriate circumstances.

Dennett speaks primarily about attributions of beliefs, but the tension he has identified exists equally for intentions. Similarly the ambiguity noted by Brandom with respect to doxastic commitments is echoed in the related concept of a practical commitment.[43] A person may undertake a practical commitment in either of two ways. He can directly undertake the commitment by asserting "I shall ..." or (more commonly) by just doing what he intends, which is also a way of acknowledging the commitment. But he may also undertake the commitment consequentially, by virtue of its being entailed by another commitment he is prepared to acknowledge. A restaurant diner's practical commitment to have the nutria entails the practical commitment to have rodent meat for dinner, whether one is aware of the entailment or not. What one is committed to practically is not exhausted by one's acknowledged plans but also includes plans one is undertaking commitment to indirectly, claims commitment to which is attributed to one by other speakers.

We can now return to the subject of our High Brow conception of the practices of ascribing reasons. The view we have taken from Dennett and Brandom helps us understand reasons for action as well. Let us see how High Brow accounts for the two features we found in Williams's view above, idealization and variability.

a) idealization: As noted, making the normative reasons statement "X has a reason to φ" is attributing to X a practical commitment to φ'ing. When we attribute a consequential practical commitment to an agent, we are choosing to see him as a rational being. We are taking an idealizing

[42] Brandom (1994: 196).

[43] Cf. Brandom (1994: 259–71).

view on his commitments. Nobody actively attributes to himself all the consequences that follow from his avowed commitments — that would be an impossible task. Taking the idealizing view, we see the agent's commitments as they ought to be, rather than as the agent happens to be prepared to acknowledge them. It follows that attributing a practical commitment to an agent has a significant idealizing component.

According to Williams, reasons have an explanatory dimension. By seeing normative reasons as commitments, we can accommodate this dimension. When we provide an intentional explanation of an agent's behavior, we do so by seeing it in relation to possible practical commitments that would make sense of the action. Doing φ intentionally is the expression of the practical commitment to φ. On the view we have developed, acknowledging the commitment to φ here and now — willing to do so — is the causal antecedent of the performance. Rational agents are trained to react to acknowledgments of commitment of this type by exhibiting the behavior in question. Through the volition, an intention or commitment is related causally and conceptually to the action intended. As practical commitments, reasons for action have an important explanatory role.

As we noted, however, Williams also says that his conception is "concerned with the agent's rationality" (Williams 1981a: 103). In addition to their explanatory role, reasons for action have a rationalizing aspect. Indeed, intentional explanations can be successful only to the extent that the agent is seen as rational. High Brow can accommodate this point as well. In saying that *X* has a normative reason to φ, we say that he is committed to φ'ing, but this may also mean that he has undertaken the commitment unwittingly, not by acknowledging the commitment directly either verbally or behaviorally, but by undertaking another commitment whose consequences include the commitment to φ'ing. Normative reasons can be consequential commitment as well as acknowledged commitments. In other words, as assessors of another's reasons, by ascribing a *consequential* commitment we gain the ability to take an idealizing stance on what he has reason to do.

b) variability: We saw above that, besides idealization, Williams also allows for a certain amount of variability in reason-ascriptions. We can see the truth in this if we follow Brandom in paying attention to the social dimension of discursive commitments.[44] As Brandom urges, in allowing consequential commitments as well as acknowledged ones, we bring into the picture the essential perspective of the attributor. For Brandom, there are two socially distinct perspectives: the one of the subject ascribing

[44]Brandom (1994: 197).

commitments to himself by acknowledging them, and the one of other speakers ascribing commitments to the subject. From the perspective of the subject himself, there is no difference between the commitments he acknowledges and the ones he is *really* committed to. But the attributor assessing a subject can distinguish between the claims the subject acknowledges and the one he has *really* undertaken. What emerges is, in Brandom's words, the distinction between the actual deontic status and the deontic attitudes of the subject.

Suppose a person who knows the melting point of different metals acknowledges commitment to the claim that this pipe is made of copper. He still may not acknowledge the claim C that this pipe melts at 1084℃. An assessor, on the other hand, may nonetheless take the subject to be committed to C, albeit only consequentially. By attributing to the subject the commitment to C, the assessor is assigning the subject a deontic status. Or again, suppose a person who knows (perhaps in theory) that nutria is a rodent announces her intention to have nutria for dinner. Still she may not acknowledge commitment to the plan P to have a rodent for dinner. However, an assessor may nonetheless take the subject to be committed to P, albeit only consequentially. The assessor is not attributing to her any deontic attitude but the deontic status of *really* being committed to P.

It follows that consequential commitments, doxastic or practical, display a fundamental relativity to the social perspective of an attributor. What is more, what a person is committed to consequentially depends crucially on further collateral hypotheses that are assumed in drawing out the consequences of an acknowledged commitment. That the pipe melts at 1084℃ follows from the avowed claim that it is made of copper only on condition of the auxiliary claim that copper melts at that temperature; the volition "I shall eat a rodent" follows from the intention to have nutria for dinner only on condition of the auxiliary claim that nutria is a rodent. As a result, there are two ways in which attributions of consequential commitments are relative. First, as attributors, we may choose either to base our attributions on the collateral hypotheses of the subject himself or to draw the collateral hypotheses from our own set of substantive commitments. If we choose the latter course, we use our own beliefs and intentions to significantly expand the person's commitments. Even if the person does not know the melting point of copper, we can still attribute to him commitment to the claim that the pipe melts at 1084℃ because there really is a relation of implication between "X is made out of copper" and "X melts at 1084℃", whether the subject is aware of it or not. Second, insofar as different scorekeepers may bring different sets of

collateral hypotheses to bear on the subject, what claims a speaker undertakes consequentially varies depending on which scorekeeper is performing the assessment.

We find the same relativity of consequential commitments in normative reasons sentences. Take Williams's gin-and-tonic case. Williams poses the question whether the gin lover has a reason to mix the liquid in front of him with tonic water. We can now see why Williams hesitates as to whether a negative answer is appropriate. On the one hand, we may say that, given his own doxastic commitments, the agent is practically committed to mixing the liquids. If we base our attribution on the subject's own view — which includes the belief that the liquid in the bottle is gin — we can infer that the agent is consequentially committed to mixing the drink. On the other hand, an attributor who knows the real contents of the bottle may choose to draw on his own beliefs to determine the gin-lover's reasons. If he does, the assumption that the bottle contains gin is not available as a collateral hypothesis that would license the inference from "I shall have a glass of gin-and-tonic" to "I shall mix this liquid with tonic water". From this perspective, the agent is *not* consequentially committed — or even entitled — to the plan of mixing the liquids. Even if he acknowledges the commitment by actually mixing the liquids, as attributors we may still evaluate this performance as incorrect.

The attitude of taking someone to be committed to a claim or plan encodes these different ways of correcting for false beliefs on the part of the subject. Conversely, the Alpine tourist we mentioned above lacks the belief that a winter trip to the mountains requires snow tires. When Williams says that someone in such a position has an internal reason, he suggests that we should correct for the lack of a relevant belief as well. In Brandom's scorekeeping terms, the tourist does not acknowledge commitment to the plan of using snow tires, but as assessors we may nonetheless take him to be committed to doing so consequentially. When we say that the tourist has a reason to use snow tires, we attribute a commitment based on collateral hypotheses drawn from our own set of doxastic commitments. Depending on the scorekeeper performing the attribution, what deontic status — what reason — the agent is taken to have may vary, insofar as collateral hypotheses are also liable to variations. This explains why Williams sometimes does and some does not attribute a reason despite the agent's own view to the contrary. Attributions of reasons are not a clear-cut practice because they depend on what material inferences a scorekeeper regards as good.

In cases of conflict, we can see the effects of the tension Dennett has identified as inherent in the attribution of intentional states. Suppose an

Alpine tourist is dimly aware, in theory, of the danger of accidents in the mountains due to inadequate tires. Our subject of attributing normative reasons is governed by both the accuracy-of-avowal norm and the rationality norm. On the one hand, we may ascribe to the Alpine tourist the intention of crossing the Alps on summer tires. To do so is to lean on the accuracy-of-avowal norm: he reveals his intention practically by taking the trip despite the danger, and we may judge his mind by his actions. This requires us to ascribe to him a blatant inconsistency in the set of intentions we ascribe to him. On the other hand, we may opt to say that the agent *does* have a reason to put winter tires on his car. This allows us to follow the norm of interpreting the agent as rational. But this means that we cannot take him "by his word" or judge him by his actions. We have to conclude that he does not understand his trip as a dangerous crossing of the Alps, perhaps because he thinks the trip is short or that it may not be snowing. In either way, we have to make a concession with respect to one of the norms.

Thus we can capture the reason statements Williams wants to mark as true on a High Brow conception of normative reasons as practical commitments. In fact, Brandom's distinctions between acknowledged and consequential commitments, and between different perspectives on consequential commitments, allow us to disambiguate the talk of reasons to some extent. As we have seen, Williams condones idealization of normative reasons, and High Brow can explain the different ways of idealization from the perspective of a scorekeeper. However, Williams is only willing to idealize in one dimension, as he only countenances correcting for mistaken empirical beliefs. In our terminology, an agent who fails to acknowledge the practical commitment to φ because he lacks, as a collateral hypothesis, the doxastic commitment p may still be taken by a scorekeeper to be committed consequentially to φ'ing if the scorekeeper can supply the claim that p from his own commitments. But in a Williamsian view, the same is not true for practical commitments. For Williams, if the agent fails to acknowledge being committed to φ'ing because he lacks, as a collateral hypothesis the *practical* commitment of ψ'ing, a scorekeeper is not allowed to supply that practical commitment from his own set of commitments. For Williams, unless the agent has a relevant "motivation", he cannot have a normative reason.

This is a pronounced asymmetry in Williams's treatment of doxastic commitments, on the one hand, and practical commitments, on the other. For Williams, practical commitment cannot be corrected for or idealized by a scorekeeper because they ultimately have to come from the agent's personal, subjective motivational profile. In this judgment, his Humean

conception of reasons and reasoning becomes evident. The last chapters, however, have revealed a number of ways in which the Humean conception falls short. In particular, practical commitments are states that aim at the good and are subject to rational norms that help determine their content. We should not assume that practical commitments are immune to rational criticism. We cannot decide the issue here. Perhaps some practical commitments are inherently personal and subjective so that it is "up to" the agent to decide whether to undertake them or not. But it should not be assumed that all practical commitments are subjective in this way. If there are intersubjective practical commitments as well as personal ones, we cannot rule out the possibility that a scorekeeper may assess another agent as having a practical commitment based on his own view of what he should be practically committed to. *Pace* Williams, such a practice would be intelligible. To ascribe a reason in this way, independent of a preexisting motivation, would be to idealize and correct for, not mistaken empirical beliefs, but wrong-headed, foolish or conspicuously absent practical commitments.

Chapter 6

Rational Requirements

6.1 Rational requirements

Our reasons for belief and our reasons for action are plentiful. That the street is wet is a reason for me to believe that it has been raining. An open window will let in fresh air — that's a good reason to open the window. The bird's red color is a reason for me to believe that the animal is a robin. That the man in the train needs assistance is a reason for me to help him. Reasons such as these are substantive: they are grounded in something that actually counts in favor of so believing or of so acting. However, we sometimes attribute reasons that are not in this way substantive. If you believe that Argos is a dog, then you ought to believe that Argos is a mammal. The former belief, we say, gives you a reason to have the latter belief, but although it is a reason, it is not necessarily a substantive reason. After all, if it is not true that Argos is a dog, it may be that from an objective standpoint you have no reason at all to think that Argos is a mammal. Again, if you intend to become a doctor, you ought to go to medical school. The former intention gives you a reason to intend the latter, but again the consideration need not be a substantive reason. Suppose that you are squeamish when you see blood and dislike interacting with patients. Since you're unsuited for the medical profession, you ought not to become a doctor in the first place. But then surely neither do you have any good reason to intend to go to medical school.

Even if Argos is in fact not a dog but a lizard, it would nonetheless be irrational for you not to believe that Argos is a mammal given that you

think he is a dog. Rationality requires you to have that belief. Once again, it may well be true that it is a bad idea for you to become a doctor. Yet, the foolishness of the plan notwithstanding, given your prior intention it would be irrational for you not to intend to go to medical school. Rationality requires you to have that intention. We have to do here with a particular type of reason. When we give a substantive reason, we point to, as Niko Kolodny puts it, "some feature of the *actual situation*" (Kolodny 2005: 509). But in special cases like the two we described, there may not be any relevant feature of the situation to ground the reason. Instead the reason arises from the agent's combination of existing commitments. There are, then, reasons of rationality as well as substantive reasons.

The existence of reasons of rationality gives rise to two puzzles that have been discussed extensively in the literature. As we will see, reasons of rationality are connected with a set of principles which are often called "rational requirements". A discussion of reasons of rationality is a discussion of rational requirements. The first puzzle concerns the logical form of these requirements. The literature is divided as to whether the principles involved take wide scope or narrow scope. The second puzzle asks what the relation between reasons of rationality and substantive reasons might be. This chapter aims to make progress towards solving the puzzles. I start by introducing the notion of a rational requirement (§1). Next I move to the discussion of the first puzzle and consider a well-known argument against the wide-scope view (§2). In the following sections I successively develop a positive conception of rational requirements as inferential principles (§§3–4). The positive conception makes use of the technical vocabulary and conceptual tools developed in earlier chapters. Any contribution the High Brow framework makes to the solution of the puzzles discussed lends additional support to the framework itself. Returning to the first puzzle, I apply the positive conception of rational requirements to the dispute between wide-scopers and narrow-scopers (§5). Finally, I close the chapter by addressing the second puzzle (§6).

We can state the relations that generate our reasons of rationality in the form of general principles. Consider again the two examples given above. In the first example, we assume that the subject believes some proposition, p, and believes that p implies another proposition q. The conclusion — that the subject ought to believe that q — relies on a principle I will call MP, for *modus ponens*. MP is naturally expressed using a conditional:

> (MP) If you believe that p and believe that $p \longrightarrow q$, then you ought to believe that q.

In the second example, we assume that the person intends an action, φ, and believes that φ'ing requires performing another action, ψ. The conclusion — that the agent ought to intend to ψ — relies on the principle IP, short for *instrumental principle*:

> (IP) If you intend to φ and believe that doing φ requires doing ψ, then you ought to intend to ψ.

To decide whether an agent violates MP or IP, we need to turn our attention to his intentional states considered as such. After abstracting away from features of the actual situation, we are left with a view of the agent's attitudes and how they are related to one another. For this reason, MP has also been called the principle of belief coherence while IP has been called the principle of means-end coherence. There are some differences among philosophers as to the correct formulation of the principles. As to IP, John Broome contends that we must add to the conditional's antecedent a clause to the effect that you believe you will not ψ unless you intend to ψ. According to Broome, an agent is only making himself vulnerable to accusations of irrationality if he doesn't end up doing φ anyway. And he also thinks that MP should have a clause to the effect that it matters to you whether q because "rationality does not require you to clutter your mind with pointless beliefs" (Broome 2005: 322–3). Whether these clauses are required is a matter of some dispute. In any case, as the modifications don't change the situation in a material way, we can safely ignore them for the present purposes and assume that an adequate formulation of the principles can be found.[1]

Two questions have dominated recent discussion of the two rational requirements:

1. Do rational requirements have wide scope or narrow scope?

2. Do reasons of rationality give us substantive reasons? Or, equivalently, are rational requirements normative?[2]

[1]Opinions have varied, not just about the definition of the individual requirements, but about the exact content of the catalog of requirements as well. Thus Broome proposes, as a requirement, the principle not to believe p and to believe $\neg p$ at the same time (Broome 2005: 322). Kolodny proposes as a principle that you ought to believe p if you believe you have conclusive evidence that it is true that p (Kolodny 2005: 521). Kolodny's principle will be discussed below.

[2]Note, however, that the claim that reasons are normative allows different interpretations. See §6 below.

Each question has generated a puzzle. Before explaining the questions further, it will be useful to set out the puzzles associated with them. The *first puzzle* starts from the observation that using only rational requirements and straightforward premises about the psychology of an agent, we can derive apparently objectionable oughts. Given the premise that a person believes that p and that $p \rightarrow q$, we can construct the following argument:

> X believes that p.
> X believes that $p \rightarrow q$.
> If X believes that p and X believes that $p \rightarrow q$, then X ought to believe that q.
> Thus, X ought to believe that q.

From seemingly innocent premises, we can derive the conclusion that X ought to believe that q. But suppose it is not the case that q; then it is not the case that he ought to believe that q, for there can hardly be a requirement to believe false propositions. A similar problem arises from the instrumental principle. Given as a premise that a person intends to φ and that he believes that φ only if ψ, we can argue:

> X intends to φ.
> X believes that doing φ requires doing ψ.
> If X intends to φ and X believes that doing φ requires doing ψ, then X ought to intend to ψ.
> Thus, X ought to intend to ψ.

Assuming the premises allows us to conclude that X ought to intend to ψ. But this is problematic: it may in fact be the case that X ought not to ψ at all because it is immoral or crazy or imprudent. And it can hardly be both the case that X ought to ψ and that he ought not to ψ. Here it may be held that the conflict is only apparent because the oughts in question are only *prima facie* oughts. In an election, perhaps, you ought to vote for candidate C because he promises to improve the schools, but at the same time you ought not to vote for C because he will cut funding for the arts. These two statements do not contradict each other because the oughts in question are not all-things-considered but *prima facie* oughts. But that isn't necessarily so in the present case. The ought in the argument's conclusion does not seem defeasible in any way. What is more, if there are moral reasons, it may be true categorically and indefeasibly that X ought not to do ψ. Finally, *prima facie* oughts can be weighed against each other, but we do not compare, or assign a score to, doing what is

rational, on the one hand, and doing what is substantively correct, on the other.

The problem is this: we can conclude that X unconditionally ought to ψ or, as it is sometimes put, we can detach the conclusion from the conditional principle IP (or MP). Accordingly the problem has been called the "Detaching Problem".[3] In his important paper "Normative Requirements", John Broome explicitly proposes a way out of the difficulty by introducing a distinction concerning the logical structure of principles such as IP and MP.[4] The principle's surface grammar suggests that we should understand the "ought" in the formula IP as taking narrow scope. We can use parentheses to mark the distinction:

> (IP-NS) If you intend to φ and believe that doing φ requires doing ψ then you ought to (intend to ψ).

But as Broome points out, another reading of the principle is possible:

> (IP-WS) You ought to (intend to ψ if you intend to φ and believe that doing φ requires doing ψ).

In this second interpretation, the conditional is entirely within the scope of the deontic operator — hence the name "wide scope". The difference between wide and narrow scope readings has important consequences for the logical behavior of the statement. In particular, while it is possible to detach the conclusion of the conditional from the narrow-scope principle, this is not possible with the wide-scope principle. From IP-WS and the psychological premises noted above, nothing interesting follows. In particular, it does not follow that the agent ought to do or intend to ψ.[5] Similarly, according to Broome, MP should not be interpreted as

[3] Cf. Way (2010).

[4] Broome (1999).

[5] Relatedly, philosophers have struggled with what has come to be called the Bootstrapping Problem. According to the principle I+, you ought to intend to φ if you believe that you have conclusive reason to φ. Now suppose that I think that I have conclusive reason to φ. Given this assumption and the principle, we can infer that I ought to intend to φ. But surely if it is the case that I ought to intend to φ, I have a reason to φ. Thus if we assume a narrow-scope enkratic principle, just from the fact that I believe that I have a reason to do something it follows that I really do have a reason to φ. Yet even when it comes to reasons, just thinking that something is the case cannot *ipso facto* make it the case. We cannot bootstrap reasons into existence like that. Cf. Broome (1999: 404). For a detailed treatment of the problem with a different moral, see Kolodny (2005: section 1). Broome's reaction to this problem is to reject the narrow-scope version of I+ in favor of a wide-scope version.

The Bootstrapping Problem is, however, just a special case of the Detaching Problem, albeit one that makes the problem particularly vivid. The problem lies in the fact that we can

> (MP-NS) If you believe that p and believe that $p \rightarrow q$ then you ought to (believe that q).

but rather as

> (MP-WS) You ought to (believe that q if you believe that p and believe that $p \rightarrow q$).

Not all writers have accepted this conclusion. Philosophers have different views concerning the question whether rational requirements such as IP and MP actually have wide or narrow scope. For Broome, the solution to the Detaching Problem is to interpret all rational requirements as wide-scope principles whose logic disallows detachment.[6] Against this argument, some philosophers have insisted that at their core rational requirements must have narrow scope.[7]

The *second puzzle* is about the distinction we started with, between substantive reasons and reasons of rationality, and about the question whether there is a connection between the two. In his initial treatment of the topic, Broome assumes without explicit argument that there is a strong connection, viz. that rational requirements provide us with reasons:

Reasons Claim If one is rationally required to φ then one has conclusive reason to φ.[8]

In later writings, Broome is more skeptical of, though still sympathetic to, the Reasons Claim and regards it as an open question whether, as he puts it, rationality is normative.[9] Why ought we to be rational? Other writers have explicitly argued that no substantive answer to this question is possible.[10] In particular, the Reasons Claim faces a challenge: if rational requirements give us substantive reasons, what type of reasons do they give us? For Broome, the most likely candidates are instrumental reasons,

---

detach oughts which seem objectionable. Because it relies on beliefs about conclusive reasons, it only occurs with the rational requirement I+ and related principles. This restriction makes the difficulty less general than the Detaching Problem, which is why in §5 I discuss only the latter. The discussion in that section will be applicable also to the Bootstrapping Problem.

[6] Other defenses of the wide-scope approach include Broome (2005), Broome (2007), Brunero (2010), Way (2011).

[7] Schroeder (2004), Kolodny (2005).

[8] This definition is taken from Kolodny (2005: 539). Note that Kolodny's main aim in his paper is to argue against the *Reasons Claim*.

[9] Broome (2007), Broome (2005).

[10] Kolodny (2005).

but some writers have pointed out problems with this approach.[11] On the other hand, it has been said that, the problems notwithstanding, the idea that rational requirements are normative has intuitive plausibility.[12]

6.2 Arguments against the wide-scope view

For now let us put the purported normativity of rationality to one side and focus on the first puzzle. As a first step toward deciding the dispute between wide-scope and narrow-scope interpretations of rational requirements, we should examine what is distinctive of wide-scope principles. Consider the wide-scope reading of the instrumental principle:

> (IP-WS) You ought to (if you intend to do φ and you believe that doing φ implies doing ψ, then you intend to do ψ).

The two antecedent conditions and the consequent of this conditional refer to attitudes of the agent. As the parentheses indicate, the ought covers the whole conditional. The "if… then" connective in IP should be interpreted as a simple material conditional, the logical connective which is true just in case its antecedent is false or the consequent is true.[13] As a material conditional is defined purely in terms of its truth conditions, the principle can be read equivalently as a tripartite disjunction:

> (IP-WS′) You ought to (not intend to do φ *or* not believe that doing φ implies doing ψ *or* intend to do ψ).

As an example, suppose that Sarah intends to clean up the kitchen and believes that cleaning up the kitchen requires taking out the trash. Suppose further that she fails to intend to take out the trash. What IP says about her is that something is amiss with her rationally: she is in a combination of states that she ought not to be in. Call this a conflict-state.

The fact that Sarah is in a conflict-state is indicated by the fact that all three of the statement's disjuncts are false. Conversely, for the requirement to be satisfied it is sufficient that one of the three disjuncts is made

[11] Cf. Kolodny (2005).

[12] Cf. Broome (2005: 321).

[13] This is assumed in most of the contributions to the debate, including Broome's later papers. In his early paper "Normative Requirements", however, Broome says that the relation is, not an ought *simpliciter*, but an ought "with determination added, from left to right" (Broome 1999: 402). Broome does not appear to pursue this interesting idea further and does not explain what the "determination" consists in. Also cf. Hussain (n.d.: 42–3).

true. This has consequences for the revision of her attitudes: she ought, on the one hand, to avoid getting into such a conflict and, on the other, to remove the tension if she finds herself in a conflicted state. With regard to the latter, three changes could occur which would resolve the conflict-state by making her conform to the wide-scope principle:

a) *Sarah comes to intend to take out the trash.* This is the most natural way to resolve the conflict. Perhaps Sarah just realized that the kitchen needs to be cleaned up. Knowing that this involves taking out the trash, she forms, by a logical train of thought, the intention of doing so.

b) *Sarah ceases to intend to clean up the kitchen.* Doing so allows her to retain her means-end belief and still not to intend to take out the trash, while at the same time escaping the conflict. On the plausible assumption that taking out the trash is an onerous task, this reaction is not hard to understand. To be sure, our thoughts frequently run along similar lines. Nonetheless, the rationality of such a reaction seems dubious.

c) *Sarah ceases to believe that to clean up the kitchen she needs to take out the trash.* This way of escaping the conflict seems even more questionable than the preceding one. Still as far as the wide-scope principle is concerned, this is a valid way of getting out of a state that she ought not to be in.

The wide-scope instrumental principle tells the agent that there is something wrong about her combination of commitments, but it doesn't single out a specific commitment-state to revise, i.e. either undertaking a commitment not held before or dropping a commitment presently held. As far as IP-WS is concerned, all ways out of the conflict are on a par. Now according to both Mark Schroeder and Nico Kolodny, this leads to problems for the wide-scope view.[14] To begin with the first criticism, Schroeder argues that the three options (a)...(c) are not equally appropriate. In fact, he regards only the first option as the right thing to do. It is rational to respond to a conflict between intending to φ and failing to intend to ψ by coming to intend to ψ. Dropping the intention to φ, as we might say, is a cop-out. Although it is a human and perhaps excusable reaction to give up a plan when confronted with its unpleasant implications, it is not what the instrumental principle recommends. The instrumental principle cannot condone taking option (c), which is not just a surprising move but

[14]Schroeder (2004), Kolodny (2005).

has the shape of a deeply irrational reaction: surely dropping an instrumental belief in response to the conflict is incorrect from the standpoint of reason.

Schroeder points out that the wide-scope principle implies that the result of choosing option (b) or (c) is that the agent is no longer irrational. After all, as one disjunct out of three, (a) is in no way privileged. As such, the principle appears to have a laissez-faire aspect about it: it condones whatever way to resolve the conflict, instead of insisting on the proper way. The formulation of the principle does nothing to mark what we would describe as the regular way to exit the conflict. For Schroeder, any wide-scope principle "posits a symmetry between different ways in which it might be fulfilled" (Schroeder 2004: 346). Given that the principle has been labeled a *rational* requirement, it ought to tell us how to respond rationally to the conflict; and from the standpoint of rationality, the options available hardly seem all on a par. For this reason, Schroeder maintains that "the symmetry predicted by the Wide-Scope account [...] is not sustained" (Schroeder 2004: 339). As a consequence, he rejects IP-WS and MP-WS in favor of a narrow-scope reading of rational requirements.

The second argument, given by Kolodny, starts by proposing a way of determining whether a given principle has wide scope. To Kolodny, it is a necessary condition of any wide-scope principle that it passes what he calls the Rational-Response Test:

> Suppose it is claimed that the process-requirement governing the conflict between *A* and *B* is wide scope: i.e., one is rationally required (either not to have *A*, or not to have *B*). For this claim to be true, it must be the case that (i) one can rationally resolve the conflict of having *A* and *B* by dropping *B* and (ii) one can rationally resolve it by dropping *A*. (Kolodny 2005: 519–20)[15]

Building on this foundation, Kolodny goes on to ask what constitutes the rational resolution of a conflict. His answer is that a way to exit the conflict is rational only if it can be explained by the subject's "awareness of what is amiss in the state (*A, B*)" (Kolodny 2005: 520). Such an explanation is possible when the agent reasons from the content of one attitude to revising the other. This is reflected by what he describes as a restatement of the first test, the Reasoning Test:

[15]Note what Kolodny is chiefly interested in here is conflicts between pairs of attitudes represented by "A" and "B", whereas IP and MP involve two premises and one conclusion. The reason is that Kolodny prefers a different set of rational requirements which have this format. See the discussion of enkratic principles below.

> The process-requirement governing the conflict between *A* and *B* is wide scope — that is, one is rationally required (either not to have *A*, or not to have *B*) — only if, from a state in which one has conflicting attitudes *A* and *B*, (i) one can reason from the content of *A* to dropping *B* and (ii) one can reason from the content of *B* to dropping *A*. (Kolodny 2005: 520–1)

According to Kolodny, we can use the Reasoning Test to determine whether a given principle really has wide-scope. For him, rational requirements do not pass the test and thus cannot have wide scope.[16] To see if this is so, we can apply the Reasoning Test to the instrumental principle. Note that *A* and *B* can represent multiple attitudes, and that they can represent the lack of an attitude as well as the existence of an attitude. On this way of seeing things, a conflict of attitudes is being in attitude-state *A* and attitude-state *B* at once.[17] In Sarah's case, *A* consists of the intention to φ (i.e. to clean the kitchen) and the belief that φ requires ψ (i.e. that doing so requires taking out the trash); *B* consists of the *lack* of the intention to ψ (to take out the trash). Sarah's conflict lies in the fact that she is in both attitude-state *A* and attitude-state *B*, so she has the choice between giving up *A*, by ceasing either to intend to φ or to believe that φ only if ψ, on the one hand, and, on the other, giving up *B*, by forming the intention to ψ. As to the first condition of the Reasoning Test, it is plainly possible to reason from the content of *A* to the dropping of *B*. The content of the compound state *A* allows the reasoning:

> Shall[φ]
> φ only if ψ
> Thus, Shall[ψ]

As the conclusion of this reasoning is to form the intention to ψ and its result is the removal of state *B*, the first condition is met. Is it equally possible to start from the content of *B* and to proceed to dropping *A*? As Kolodny rightly says, this is impossible. To be sure, Sarah might notice that she doesn't intend to ψ and move from there to the conclusion that either she ought not to intend to ψ or that φ'ing does not imply ψ'ing. Such transitions are not uncommon, in particular if ψ'ing is an unpleasant task, but they do not count as *reasoning* in a proper sense of the word. Sarah cannot rationally move from *B* to *A* by a rational argument starting

[16] Or at least many such requirements do not pass the test.

[17] As I use it, the word "attitude-state" means having or lacking one or several attitudes. For example, an attitude-state may be "does not believe *p* or *q*" or "intends φ".

from the lack of the intention to take out the trash and leading to giving up the intention to clean the kitchen. Such a transition would be, not reasoning, but a form of self-deception.

Moving from *A* to *B* follows the current of proper reasoning, but going in the reverse direction — "going upstream" — does not. For Kolodny, there is no such thing as upstream *reasoning*; there are only upstream irrational transitions. As Kolodny writes, "[r]ationality requires one, in forming, retaining, or revising one's attitudes, to follow the downstream current" (Kolodny 2005: 529). Rational requirements are a stream with a built-in direction. According to the Reasoning Test, a principle has wide scope only if it allows both downstream and upstream transitions. The instrumental principle does not satisfy the second condition of the Reasoning Test, so according to Kolodny it cannot be a wide-scope principle.[18]

Although I agree with Kolodny that rational requirements do not satisfy the second condition, I do not think that this warrants the conclusion that rational requirements cannot have wide scope. In other words the Reasoning Test is not a good indicator of whether or not a principle has wide scope. Kolodny portrays the Reasoning Test as a natural extension of the Rational Response test. But though the Rational Response test, in my view, correctly captures the nature of a wide-scope principles, the Reasoning Test does not. Thus it is true that, with any wide-scope principle, there isn't one unique rational way to exit a given conflict-state, because depending on the circumstances, it may sometimes be correct to drop *A*, sometimes to drop *B*. But this doesn't entail that, in any given instance, it is possible to reason one's way, following a direct path, from *B* to dropping *A* as well as from *A* to dropping *B*. In Sarah's case the Rational Response test says that it is sometimes correct to develop the intention to take out the trash, but that at other times it is correct for her to drop the intention to clean the kitchen. But this doesn't entail that there is a correct reasoning path leading from her reluctance to take out the trash to dropping the intention to clean the kitchen. The rational response doesn't have to consist in a piece of reasoning from *A* to *B* or a piece of reasoning from *B* to *A*.

[18] I am simplifying Kolodny's argument a little here. Kolodny actually proposes two arguments. The first argument holds that in order for reasoning from *B* to *A* to be possible, *B* must have a content associated with it. If *B* consists in the lack of an attitude, there is no content to reason from, so no reasoning is possible. The second arguments holds that reasoning from *B* to *A* must always be downstream. I take the second argument to be more general than the first because it also applies to cases that the first does not cover. Focusing on the upstream reasoning part does not, I think, weaken Kolodny's argument.

That there can be multiple rational exits from a rational requirement is shown by two examples. Suppose that Karl believes that p and believes that $p \rightarrow q$ (A) but fails to draw the conclusion that q (B). He is in violation of *modus ponens*. As in Sarah's example above, when we convert the conditional of MP into a disjunction, he has a total of three options, namely

a) forming the belief that q,

b) dropping the belief that p and

c) dropping the belief that $p \rightarrow q$.

In many cases, no doubt, forming the belief that q is appropriate. Perhaps having recently learned that it is true that p and having good evidence for the conditional, it is most natural for Karl to go on to conclude that q. But it isn't obvious that this choice is invariably the best. For one thing, suppose that Karl has good independent evidence that makes it likely that, in fact, it is not the case that q; and suppose further that he has little evidence for the claim that p. In this scenario, it seems that it would be preferable for Karl to respond to the conflict by giving up his belief that p. To decide whether option (a) or (b) is more rational, then, we need to consult the wider context of the beliefs. The choice between (a) and (b) is familiar from philosophical arguments: As it is sometimes said, one man's *modus ponens* is another man's *modus tollens*. Next, suppose that Karl has good evidence that p and also that $\neg q$, but that his belief that $p \rightarrow q$ is merely a wild conjecture. In these circumstances, it would be entirely appropriate to give up the conditional belief while holding on to the belief that p. Thus without taking the wider context into account, we cannot decide whether the rational result would be dropping state A, by picking option (b) or (c), or dropping state B, by picking option (a).

A similar point applies to practical reasoning. Going back to Sarah's case, her situation is that she has to choose between taking the means (ψ), giving up the end (φ) and stopping to believe that the end requires the means (φ only if ψ). Now it is certainly often appropriate to react to the conflict by taking the means, but in some cases a different reaction is warranted. For instance, as before it may be that the evidence for the empirical judgment that φ *only if* ψ is weak, in which case the perceived conflict would prompt the agent to reconsider the unfounded empirical assumption. Furthermore, it may be that Sarah is deeply committed to another project which in effect precludes ψ'ing. In this case, it is appropriate for her to backtrack and to reevaluate her unconditional commitment to achieving

the end, ψ. The result of this reexamination may well involve giving up the intention to ψ. In both of these cases, the rational requirement would be satisfied by dropping state *A* rather than by dropping state *B*.

The wide-scope readings of IP and MP, then, leave open alternative ways of removing the conflict. Sometimes it is better to prefer one of the alternative paths to the default option, and whether this is so depends on the further context, which includes the evidence for the belief-premises and the strength of commitment to the intention-premises. The wide-scope principles cannot give specific advice about which way out of the conflict is best because they do not, and cannot, take these further factors into account. Notice that, in the examples given, the alternative ways of resolving the conflict really are rational resolutions. If so, wide-scope principles pass the Rational Resolution test: one can rationally resolve the conflict by having *A* and dropping *B* (the default path) *and* one can rationally resolve the conflict by having *B* and dropping *A* (the alternative path).

Still it would be wrong to describe the revisions involved, with the Reasoning Test, as "reasoning from the content of *A* to dropping *B*" and "reasoning from the content of *B* to dropping *A*". In the theoretical example, what Karl realizes, as a consequence of noticing the conflict, is that his evidence for the belief that *p* is weak. He engages in reasoning which leads him to dropping the belief that *p*, but it is not a case of reasoning which starts from the fact that he doesn't believe that *q*. Instead it is an argument that starts from independent premises related to the evidence for the claim that *p*. The resolution of the conflict is rational in virtue of the fact that it is the result of reasoning, but not of upstream reasoning. The relevant reasoning is a different, independent line of downstream reasoning leading to the revision of the premises. In the practical example, Sarah becomes aware that the end to which she has committed herself is questionable. But she does not perform an upside-down argument, starting from the fact that she isn't committed to the means and leading to dropping the commitment to the end; that would be "upstream reasoning", which is to say not reasoning at all. Instead she engages in a separate piece of reasoning from independent premises, which causes her to drop the commitment to the end. The resolution of the conflict counts as rational because it is the upshot of a line of downstream reasoning leading to the revision of the premise.

As the examples show, the principles pass the Rational Response Test but do not pass the Reasoning Test. This gap shows that the latter is not a more elaborate version of the former. Instead of taking the observation that IP and MP do not allow *both* reasoning from *A* to dropping *B* and

from *B* to dropping *A* to show that the principles are not wide scope, we should reject the Reasoning Test as insufficient.

Before concluding this section, we need to consider a complication. The argument developed by Kolodny requires us to pay attention to the difference between his catalog of requirements and ours. In his paper, Kolodny is not primarily concerned with IP and MP but with a different set of rational requirements.[19] The two corresponding principles are:

> (B+) Rationality requires one to believe that *p*, if one believes that there is conclusive evidence that *p*.
>
> (I+) Rationality requires one to intend to *X*, if one believes that there is conclusive reason to *X*. (Kolodny 2005: 521)[20]

To distinguish these principles from IP and MP, I will call B+ and I+ enkratic principles.[21] Supposing that these principles also count as rational requirements, there should be wide-scope versions of these. Do they resemble MP and IP in allowing multiple ways to exit a conflict? If there is a unique rational revision of attitudes when conflicts of this type occur, B+ and I+ do not conform to the Rational Response Test. This would undermine the idea that a correct wide-scope reading of all principles is available.

Suppose that the Inspector believes that he has conclusive evidence that Smith killed Jones (A), but he does not go on to form the belief that Smith killed Jones (B). The belief about the evidence is a meta-belief. According to the Rational Response Test, he can react to the conflict in either of two ways: by coming to believe that the victim was killed by Smith — dropping A — or by giving up the belief that he has conclusive evidence that

[19]Kolodny briefly considers *modus ponens* in Kolodny (2005: 541–2). He conjectures, however, that all rational requirements, including IP and MP, can be reduced to enkratic principles.

[20]Kolodny also admits two negative principles, B- and I-. According to B-, rationality requires you not to believe that *p* if you believe that there is not sufficient evidence that *p*. According to I-, rationality requires you not to intend to *X* if you believe that you lack sufficient reason to *X*. I will ignore these further principles mainly to simplify things. But it is worth pointing out that the principles are more controversial than their positive counterparts. After all, I believe many facts without remembering what evidence I drew on when I learned them in the first place, and I intend many things without having the explicit belief that there is, in fact, conclusive reason to intend them. That does not make these beliefs and intentions irrational. For a defense of the negative principles, see Kolodny (2005: 527).

[21]The term "enkrasia" or "krasia" (or "enkrateia") is used for principles similar to B+ by a number of writers, including Broome (2007) and Way (2010). In this usage, an agent is "enkratic" if he is not akratic.

Smith killed Jones — dropping B. More than in the case of MP, it seems intuitive that the only rational option is to do the former; giving up the belief about evidence would be irrational or "upstream reasoning". But it is not hard to imagine a situation in which revising the belief about evidence would be entirely rational. For believing that you have conclusive evidence for the belief that *p* entails having one or more particular beliefs that make it probable that *p*. Thus suppose the Inspector believes:

> E1: Smith was seen hurriedly leaving the crime scene.
> E2: Smith is in debt and in dire need of money, which he could rob from the victim.
> E3: Smith owns a gun.

Together these pieces of evidence make it probable that Smith committed the murder, so the Inspector is correct to believe that he has conclusive evidence for the claim. If we have no further information about the Inspector's commitments, the most rational procedure would be to drop A. But suppose, further, that the Inspector is aware that the witness who reports Smith's whereabouts is unreliable; that Smith is about to inherit money to pay off his debts; and that the caliber of Smith's gun doesn't match that of the murder weapon. If so, it would be rational to react to the conflict described by B+ by giving up the assumption that E1…E3 support the hypothesis that Smith is the killer. By giving up his reliance on the particular pieces of evidence, the Inspector also stops believing that he has conclusive evidence that Smith committed the murder. In doing so, he resolves the conflict between *A* and *B* by dropping *A*. The revision is rational because it consists in reasoning, but he doesn't reason from his lack of belief that Smith was the killer (A) to giving up his belief about the evidence (B): he gives up the belief after reasoning from facts about E1…E3 being insufficient evidence for his hypothesis.

Despite the initial appearances, then, it is not true that B+ can be resolved in only one way. Which of his commitment-states needs revision depends on the context of the Inspector's further commitments. In this, B+ is no different from *modus ponens* and the instrumental principle: sometimes it is rational to drop *A*, sometimes it is rational to drop *B*. The belief about evidence, which plays a crucial role in B+, is only a placeholder for the substantive considerations that form the evidence. The real evidential support for the conclusion that Smith killed Jones lies in the Inspector's substantive beliefs about the homicide, and there is nothing irrational about giving up these beliefs in the case of a conflict if they are themselves badly supported and if there is an independent line of thought that supports the claim that Smith is innocent. A similar point applies to I+.

As with B+, the real work in supporting the conclusion is done by the individual reasons — considerations which may themselves be built on shaky grounds. If so, there is nothing irrational about reconsidering the original particular reasons in the light of an incompatibility. Then the situation is the same as with the instrumental principle and we have little reason to think that the enkratic principles are special in this regard. Kolodny's argument fails to show that there are no wide-scope versions of these principles.[22]

To summarize, Kolodny's argument against wide-scope interpretation does not succeed. Although the Reasoning Test is not accurate, he is right that all wide-scope principles must pass the Rational Response Test. Against Kolodny, however, we have concluded that the rational requirements we have considered pass the latter test, and this is true for his favored enkratic principles as well. This move allows us to respond to Schroeder's criticism as well. Schroeder says that wide-scope principles posit a symmetry between different ways in which they might be fulfilled, an asymmetry between revising the requirement's antecedent conditions, on the one hand, and, on the other, revising the consequent attitude. Now as we have seen, it is indeed true that, depending on the wider context, it may be correct either to revise the antecedent or the consequent.[23] But this does not amount to positing a symmetry or to saying that a wide-scope principle authorizes all possible revisions. Rather than representing a laissez-faire policy, the principle simply remains silent on the question of how to respond to the conflict. The principle only tells you *that* you are in a conflict and *that* you ought to revise your attitudes so as to remove the conflict, but it doesn't specify *how* you ought to revise your attitudes.

Schroeder's grounds for rejecting the wide-scope interpretation of requirements is that they predict a symmetry between different outcomes that, on closer scrutiny, is not sustained. But defenders of the wide-scope view are not committed to the claim that requirements predict a symmetry. The job of the principles is not to prescribe specific attitudes but to

[22]There is a reason, however, to focus on the principles we have been working with, IP and MP, rather than the enkratic principles. As the latter principles start from premises that concern what reasons (or evidence) we have for a given proposition, they are more reflective, higher-order principles than the former: they involve judgments about reasons and their relative weights. While enkratic principles operate on a meta-level, IP and MP are first-order principles. If so, they require less cognitive sophistication and are thus more basic: we do not, in everyday reasoning, reflect on our own reasons but rather directly act upon them.

[23]As Way puts it, "Wide-Scope requirements do not discriminate between the different ways in which you might avoid irrationality" (Way 2011: 229).

inform the agent of the presence of a conflict. Consequently, Schroeder's argument does not give us reasons to reject wide-scope rational requirements as incorrect. Interestingly, Schroeder at one point explains that "[w]ide-scope principles are good at predicting what is wrong with an agent *at a time*. But they are not good at predicting the rational ways for an agent to *change* her situation" (Schroeder 2004: 346). According to him, because the wide-scope readings do not capture this important aspect of the instrumental principle, and of other requirements, they ought to be rejected. But the diachronic character of the rational requirements should not lead us to reject the wide-scope principles. Instead we should accept that they exist and are valid, while remembering that they only have the limited role of informing us of a conflict.[24]

Does this mean that rational requirements must have wide scope? To draw this conclusion would be rash. Although Kolodny's and Schroeder's arguments have failed to show that rational requirements cannot have wide scope, Schroeder's point that there must be diachronic principles which account for revisions of attitudes is important.[25] As wide-scope principles cannot satisfy this condition, they are not by themselves sufficient. Our attitudes cannot be governed by wide-scope principles alone, so we need to supplement them with additional rational principles. The nature of these additional principles will be explored in the following sections, and this exploration will lead us back to the question whether these additional principles have wide or narrow scope. Although the present section has been chiefly critical of Kolodny's and Schroeder's arguments, both authors, as we will see, are right to insist that the most important rational principles have narrow scope. What all this doesn't show is that there is no place at all for wide-scope principles. It is really true that ra-

[24]In a similar vein, Kolodny distinguishes between state-requirements, which "simply ban states in which one has conflicting attitudes", and process-requirements, which "say how, going forward, one is to form, retain, or revise one's attitudes so as to avoid or escape such conflict-states" (Kolodny 2005: 517). His arguments against wide-scope views of rational requirements assume that at least some requirements are process-requirements. He goes on to discount the importance of state-requirements, if they exist at all. In the text it was claimed that his argument that there are no wide-scope rational requirements is not convincing. However if what Kolodny wants to say is just that rational requirements, if they have wide scope, cannot also be process-requirements, then I agree with him. There is a place for narrow-scope process-requirements, and in fact in my view they play an important role, as will be apparent in what follows. I only take issue here with Kolodny's argument that rational conflicts cannot be resolved in more than one way and that we should, for this reason, reject wide-scope requirements.

[25]Officially, Kolodny's argument purports to show only that *some* rational requirements have narrow scope, while other principles may have wide scope (Kolodny 2005: 515). But he clearly thinks that ultimately all rational requirements can be reduced to narrow scope principles. By contrast, it is claimed here that although there are wide-scope principles, there must be narrow-scope versions of the same principles as well.

tionality requires us to (either not believe that p or not believe that $p \longrightarrow q$ or believe that q) and Broome is right to defend the existence of IP-WS and MP-WS.

6.3 The function of rational requirements

The preceding argument suggests that rational requirements consist of more than the wide-scope versions of the principles MP and IP. Let us take a closer look at the requirements to pinpoint their principal form. We will also try to substantiate the suspicion that this form will have narrow scope. Starting with some preliminary work, what is the function of rational requirements? What do principles like *modus ponens* and the instrumental principle accomplish? A first answer might be this: The principles describe the way our minds work on a basic level, and their topic is a particular type of regularity of our psychological states in relation to one another. On this view, *modus ponens* says that people who think that p and that $p \longrightarrow q$ tend to think in appropriate circumstances also that q; the instrumental principle says that people who intend to φ and believe that φ'ing requires ψ'ing tend in the appropriate circumstances also to intend to ψ. Our mental machinery is designed in such a way that attitudes of one type tend to occur together with attitudes of another type.[26]

This is true as far as it goes. It is true that rational requirements correctly describe how our minds work. Although we violate the requirements with some frequency, we do follow them in the vast majority of cases. But as our criticism of dispositionalism leads us to expect, the view does not fully capture the principles.[27] According to *modus ponens* we ought to believe that q if we believe that p and that $p \longrightarrow q$. The dispositional view reads the "ought" in this principle as the ought of prediction, as in the remark that the train ought to arrive shortly. But this is clearly not what the word should mean in this context. It means a genuine requirement: the principles express norms properly speaking, not in the attenuated sense of a feature that it is statistically normal for human reasoners to have. When an agent fails to conform to *modus ponens*, she has broken a rule, not just gone against a regularity.

This observation suggests a second view. When an agent makes a mistake of this kind, we can take her to task for this violation of the requirement.

[26] This is the view proposed in Pettit (2002).

[27] See §§2.4–5.

It may be a consideration like this that has inspired the view that the requirements are normative, albeit in a less-than-complete sense. The idea is that principles like *modus ponens* are standards whose use is limited to evaluation, coupled with the thought that standards of evaluation need not necessarily be norms. Thus we can use the rational principles as a yardstick to assess the performance of an agent as rational or irrational and, as the case may be, criticize her. Yet according to this view, rational requirements fall short of real normativity: they are not in force for us in the way that prudential or, perhaps, moral demands are.[28]

It is true that the principles, in the form in which they have been presented, are well suited for the assessment of the performance of other agents. The reason is that, in giving direction about what is to be done, the formulations embody a third person (or second-person) stance. Their point is to reveal relations between attitudes attributed to an agent, and this is just what an onlooker, as opposed to a deliberating subject, does. And it is true that we often use the requirements in this way to assess or criticize another subject; this interpersonal aspect of the requirement is crucial. Again this is correct as far as it goes. But the proposal goes beyond merely stating that this is an important aspect of the principles. It states that their use is purely evaluative. This does not seem true since it hides the fact that rational requirements have an important first-person aspect in addition to a third-person use. The principles have a part to play in rational deliberation as well as in evaluation.

As we may put it, the requirements guide our rational conduct. Kolodny also take note of this first-personal character.[29] Kolodny rightly observes that the pure evaluative view fails to notice another normative dimension. But he adds that this dimension is the "experience of being bound by a rational requirement to believe or intend" a proposition (Kolodny 2005: 554). Or again we "feel that we ought to respond as [the requirements] require" (Kolodny 2005: 555). This suggests a phenomenological or experiential aspect of normative force that, as we have pointed out, is questionable.[30] The important thing is not that there is something it is like to be aware of a rational requirement. Rather it is that we are conscious of these requirements and that these requirements are operative in our choices.

In the light of these circumstances, we can conclude that the function of rational requirements is to be operative in an agent from the first person standpoint rather than to play the role of a mere regularity or of an

[28]For this idea, see Scanlon (2007) and Kolodny (2005: 551–6).

[29]See also Scanlon (2007: 87).

[30]Cf. §2.3.

evaluative standard. The function is that of a guide of our theoretical or practical reasoning. To be sure, the role of MP and IP in reasoning is important, and it will be discussed in detail in the following sections. However, while establishing this point we need to avoid two pitfalls. The first is the assumption that to play their role in reasoning, the principles must be premises; the second is the assumption that the pieces of reasoning we are concerned with are directed at attitudes. Let us take these two sources of difficulties in turn.

Starting with the first issue, we can ask what role principles such as *modus ponens* play in our actual reasoning processes. One may assume that they appear in the shape of premises to the arguments which constitute the reasoning.[31] Take as an example Ron, who believes that Argos is a dog and that dogs are mammals. If rational, Ron also believes that Argos is a mammal. Let us assume that, having discovered Argos's canine nature, Ron concludes that Argos is a mammal. Clearly, in this reasoning he is being guided by *modus ponens*, but in which way exactly does the principle provide guidance the reasoning? On the assumption being considered, the rational principle is a further, unstated premise to the argument. The argument, the line of thought might run, has roughly this form:

(1) I believe that Argos is a dog.

(2) I believe that if Argos is a dog, then Argos is a mammal.

(3) Thus, I believe that Argos is a mammal.

As a hidden premise, we would have to assume the rational requirement:

(4) If I believe that p and that $p \rightarrow q$, then I ought to believe as well that q.

Adding this additional proposition arguably adds to the perspicuity of the inference. On a possible view, rational requirements exert their guiding function by being present as a further, essential premise, which can be omitted for the sake of abbreviation. But this view cannot be correct. First, we should note that when we actually rehearse an argument mentally, rational requirements do not usually figure as premises. Nor does it seem that they merely slip from our minds. More importantly, the sequence (1),(2),(3),(4) is not a typical or simple theoretical argument. When we go through the steps of a theoretical argument, we do not prefix each step with "I believe that" or a similar clause. Instead, our reasoning is more likely to have a simpler structure:

[31] Cf. §4.4.

(5) Argos is a dog.

(6) If Argos is a dog, Argos is a mammal.

(7) Thus, Argos is a mammal.

Clearly this is the elementary form of reasoning. But now consider the result of adding the requirement, in the guise of premise (4) above, to the argument (5), (6), (7). Adding this further premise in no way improves the perspicuity of the argument simply because it does not help the argument go through. The premise fails to make contact with the premises (5) or (6) because the antecedent of (4) has as its subject matter Ron's beliefs, whereas the former premises aren't concerned with anyone's beliefs. In a similar way, the premise fails to make contact with the conclusion, (7), because (4)'s consequent is concerned with beliefs, whereas there is no mention of beliefs in (7). The statement of *modus ponens* cannot be simply incorporated into the argument in the form of a premise. The same point applies to the instrumental principle. Suppose that Sarah intends to clean up the kitchen and believes that cleaning up the kitchen requires taking out the trash. If so, then according to the principle, she ought to intend to take out the trash. But what moves her to the intention would be a practical syllogism:

> Shall[I clean up the kitchen]
> Cleaning up the kitchen requires taking out the trash.
> Thus, Shall[I take out the trash]

As in the theoretical case, the individual steps do not refer to psychological states. But then adding the premise

> If I intend to φ and I believe that φ'ing requires ψ'ing, then I ought to intend to ψ.

does not fit with the premises or the conclusion of the syllogism, whose subject matter is not psychological at all. The requirement cannot assume the form of a regular premise. If the instrumental principle cannot be an explicit premise in an argument, it cannot be a hidden or suppressed premise, either.

Consider now the second pitfall mentioned above. The point, which was already implicit in the first pitfall, is that our reasoning proceeds on the level of content, rather than attitudes. Just above we observed that regular theoretical syllogisms do not aim at the conclusion that *I believe that* ...

and that regular practical syllogisms do not aim at the conclusion that *I intend to* ... Now consider a distinction offered by Scanlon. He asks how we should interpret the antecedent of a rational requirement, the part marking its circumstances of application. One way to conceive it is as "a judgment about the adequacy of reasons for holding the attitude in question".[32] As an example, Scanlon explains that "it might be that an agent who judges there to be conclusive reason to believe that p must, insofar as he or she is not irrational, believe that p." But as Scanlon rightly points out, these types of judgments, in which we are explicitly concerned with reasons for an attitude, are artificial. By contrast, our judgments are usually content-directed. In Scanlon's words,

> In deciding whether to believe that p, we "direct our attention to the world" and ask whether p is true, and a judgment leading to an intention to do A at t is likely to be a judgment about the merits of doing A. (Scanlon 2007: 91)

If most if not all reasoning is content-directed rather than attitude-directed, then this must be reflected by the rational requirements. For this reason, the principles should be interpreted as content-directed. Thus in the example of a theoretical syllogism above, we can imagine Ron reasoning out loud:

> Argos is a dog. A dog is a mammal. So Argos is a mammal.

The propositions do not contain reference to psychological states. The reasoning process comprises a series of expressed contents. The content of the premises consists in what Ron believes, not in his believing it, and the content of the conclusion is what he ought to believe, not his act of doing so.[33]

We can make the same point using the tools introduced in earlier chapters. The elements of a piece of reasoning are propositions, or contents, which represent commitments, pragmatic or doxastic. But there is a difference between attributing a commitment to a person and expressing the commitment. The two differ in subject matter. Attributing a commitment, e.g. by saying "John believes that snow is white", is a way of talking about an agent's psychology. By contrast, expressing a commitment, e.g. by saying "Snow is white", does not normally concern the psychology of an agent. It is true that by attributing a commitment one also

[32] Scanlon (2007: 90).

[33] Cf. §1.4.

undertakes and expresses a commitment, namely one about someone's psychology. But the difference ought to be sufficiently clear. It is easy to attribute to someone the commitment that p without oneself undertaking the commitment that p — without thereby endorsing the truth of p oneself.[34]

Thus the actual reasoning process consists of a series of commitments expressed. Accordingly, the requirements of rationality should be content-directed in this sense: they concern the contents of the commitments expressed, not the attitudes attributed. In following the requirements, we "direct our attention to the world". The function of the rational principles, then, is to guide our reasoning in a content-directed way. We do not comply with the requirements because they figure as premises in our reasoning. Instead, the requirements appear transparent. This is related to a point made by Scanlon. While discussing the way in which the rational requirements constitute norms, he writes:

> The behavior of a rational agent will exhibit (at least to a significant degree) the regularities described by requirements of rationality. But this is not because the agent sees this way of behaving as required by principles that she must be guided by. A rational agent who believes that p does not accept arguments relying on p as a premise *because* she sees this as required by some principle of rationality to which she must conform. Nor does she generally do it "in order not to be irrational". Rather, she will be willing to rely on p as a premise simply because she believes that p. (Scanlon 2007: 85–6)[35]

[34]One source of confusion is the fact that it is possible for me to attribute a commitment to myself autobiographically. In doing this publicly, I talk about my own psychology. Thus just as I can attribute to John the belief that snow is white, I can say "I believe that snow is white", and just as I can attribute to John the intention to close the window, I can say "I intend to close the window." Still self-attributing a commitment is not the same as expressing it. To see that the two are different it suffices to see that it follows logically from "Snow is white", but not from "I believe that snow is white", that snow is not black; and that it follows logically from "I intend to close the window", but not from "I shall close the window", that I presently have an intention. Now it is true that in everyday speech, we do not always observe this distinction clearly. Often we use the expression "I intend to φ" to mark the *expression* of the intention without purporting to comment on our own psychological state. Similarly the expression "I believe that p" need not have the function of performing a self-attribution of a doxastic commitment. Sometimes it is used just to endorse the claim that p, while at other times the phrase has a different use altogether, such as marking one's uncertainty. Still, philosophically speaking, we should be careful not to blur the line between attributing and undertaking a commitment.

[35]The requirement on which this passage is based is the requirement to be prepared to take a proposition as a premise in further theoretical arguments. We may say that this is similar to the enkratic principles noted above.

Scanlon rejects the idea that what makes agents reason in accordance with rational principles is their explicit or propositional awareness that the behavior in question is required by the principle. The agent relies on the premise that *p* and argues in accordance with *modus ponens*, not because of a premise-form awareness of the principle but simply because he is committed to *p* being the case. The principle itself remains transparent. Similarly, one might mistakenly think that the agent draws the conclusion she draws because she would rather not be irrational. Scanlon rightly rejects this suggestions as well. Although it would be true that, if the agent failed to conform to the principle, she could be accused of irrationality, surely this is not the agent's operative concern. The concern is with the content *p* and with all that this commitment implies.

Scanlon points out that what is true for modus ponens also holds for the instrumental principle. If an agent takes the required means to an endorsed end, this is not in the light of the rational requirement, but rather because of the endorsed end itself. Nor is she likely to take action in order to evade the charge of irrationality. Summarizing his criticism, Scanlon writes:

> Ideas of rationality and irrationality belong to a higher-order form of reflective thought that we need not engage in when, for example, we see that we have reason to do what will advance one of our ends. (Scanlon 2007: 86)

To suppose that means-end reasoning or *modus ponens* always involves the conscious reflection involving the notion of rationality or irrationality is to overintellectualize the simple process of the theoretical or practical syllogism. Regular reasoning is object-level reasoning, not reasoning about reasoning. The requirements are content-directed and remain transparent to the agent.

6.4 Inferential requirements

To summarize, rational requirements play an important role in reasoning. Principles are not just dispositions or evaluative standards, so a conception will be adequate only to the extent that it shows how the requirements are norms in the full sense. Further, rational requirements should play a role in first-person deliberation. Even if they do not constitute premises of reasoning, they play a role in guiding our reasoning. Finally, we have observed that the requirements must be content-directed: they

must be interpreted as operating on the level of what is believ*ed* or intend*ed*, rather than, as the wide-scope principles we have examined suggest, on the level of attitudes.

Moving on to a positive conception of the principles, we can draw upon results we have obtained earlier in our reflections on the nature of practical reasoning.[36] Recall the lessons we drew from Caroll's story of Achilles and the Tortoise. On an adequate picture of reasoning, it is hopeless to try, like Achilles in the story, to make explicit everything that is responsible for making a piece of reasoning go through as valid. Inevitably, some aspects of the reasoning are bound to remain implicit in a shared practice of common ways of treating inferential transitions as good. Moreover, as we have seen, we should countenance short arguments moving from a single premise to a conclusion. We can see short arguments such as

> It's raining. Thus I shall open the umbrella.

as complete, materially valid arguments rather than as enthymemes. Similar, consider the single-premise inference:

(8) Argos is a dog.

(9) Thus Argos is a mammal.

This inference, as we emphasized, is a good inference as it stands. Nonetheless, it is still possible to challenge the argument. Through Socratic prompting, we may be led to wonder whether this inference really is one that ought to be accepted. As we have seen, a response may introduce a conditional, in the simplest case:

(10) If Argos is a dog, Argos is a mammal.

This conditional makes explicit that it is correct to move from (8) to (9). By adding (10) to the short argument, we explicitate a propriety that was already implicit in the original argument. Now (10) is a rather specific proposition, which involves the proper name "Argos". Clearly another, more general premise would do the same job:

(11) For all x, if x is a dog, then x is a mammal.

or

[36] §4.4.

(12) Dogs are mammals.

Like the more fine-grained conditional about Argos, this principle licenses the inference leading to "Argos is a mammal". But it also licenses many other inferences like "Kerberos is a dog, thus Kerberos is a mammal" and "Fido is a dog, thus Fido is a mammal." Its function is similar to that of (10), but it has more general application.

As we have seen, this challenge can be repeated. Thus in response to doubts about an inference, we can point to a proposition that makes the propriety of this first-level inference explicit. In the same spirit, we can ask whether the argument comprising (8), (9) and (10) is a valid inference. To answer this challenge, we can follow Achilles in producing another proposition:

(13) If (i) Argos is a dog and (ii) if Argos is a dog then Argos is a mammal, then (iii) Argos is a mammal.

Again this premise makes explicit what is implicit in regarding the previous argument as valid. Notice that this is a conditional proposition which itself includes a conditional premise. This proposition is of a higher order than (10). As to responses to the first challenge, (11) and (12) were more general alternatives to (10). Now in the second challenge, there is again an alternative way of making the inference explicit. To see this, notice that (13) has a pragmatic force similar to:

(14) "Argos is a dog" and "If Argos is a dog then Argos is a mammal" imply "Argos is a mammal".

We can generalize this statement to:

(MPI) "p" and "If p then q" imply "q".

In going from (10) to (11), we replaced the individual constant Argos with a variable x ranging over a set of individuals. The step from (14) to (MPI) is similar: we replace specific propositions with schematic letters p and q, whereas p and q range over whole propositions. Thus we move from a formulation that licenses one particular inference to a more general formulation that governs not just this inference but many inferences besides.

Of course, what we arrive at is just a formulation of the rational requirement MP. It follows that we can conceive of a rational requirement as the

result of a general way of making explicit the inference underlying a regular argument. As we saw, MP is a natural extension of what we do when we make explicit what is implicit in a single-premise inference. Compared to the proposition (10), extracting a rational requirement means going one step further or climbing another rung on the ladder of explicitness. Whereas the earlier step was captured in a conditional, MPI makes explicit the transition from premises which include a conditional.

What MPI in effect says is that a certain sort of inference is good. Thus we can see it as equivalent to the inference-schema:

> *p*.
> If *p* then *q*.
> Thus, *q*.

How does this compare to the way we have talked about rational requirements in previous sections? Recall the original formulation of *modus ponens*:

> (MP) Rationality requires of you that, if you believe *p* and you believe (if *p* then *q*), then you believe *q*.

We ought to regard the inference schema above as an alternative way of expressing the rational requirement MP. It is worth noting that the inference-schema is clearly a content-directed principle rather than an attitude-directed principle. The present strategy can be applied to the other rational requirement we have been concerned with:

> (IP) Rationality requires of you that, if you intend to φ and you believe that ψ'ing is a necessary means to φ'ing, then you intend to φ.

Following the train of thought above, this principle can be seen as a way of making explicit what we regard as a good inference, namely an inference of the form:

> Shall[φ]
> φ requires ψ
> Thus, Shall[ψ].

This inference-schema is general in contrast with specific relations of inference between, say, intending to clean the kitchen and intending to

take out the trash. Compared to the latter inferences, it applies to a large number of intended contents.

According to the present view, then, rational requirements can be conceived as explicitations of general inferential relations. Introducing a rational requirement does not add a new element but brings to the fore something that is already active in the practice, although one needs to as it were read between the lines to see it. It is for this reason that rational requirements seem both familiar and exotic at the same time. They are familiar because we rely upon them in all reasoning, even if we aren't necessarily explicitly aware of them. Yet they are also exotic because we do not commonly ascend to the level of abstraction required to give them expression in the form of explicit claims or schemas.

A few clarifications are in order. To begin with, rational requirements are a tool of abstraction, but conditionals can function as tools of abstraction in a similar way. Thus it will be useful to briefly compare requirements to this use of the conditional. As we have seen, conditionals can be thought of as explicit inference-licenses: if you have the antecedent in your commitment-box, you are entitled to having the consequent in your commitment-box.[37] Similarly, rational principles are explicit inference-licenses which entitle you to put the consequent in your commitment-box if you have the antecedent conditions in your commitment-box. One may think that it follows that rational principles must be conditionals, considering that Broome's original formulation has the "if… then" structure. But whereas Broome's formulation IP is in terms of the ascription of attitudes, the inference-schema operates on the level of content. A form aimed at attitudes allows the use of the conditional, but the content-directed form does not.

In fact, the new formulation of the principles cannot be put in terms of the conditional because of two evident problems. First, suppose that we formulate the content-directed principles using the conditional:

> If p and if p then q then q.

This statement is a mere logical tautology which doesn't capture the inference licensed by the rational requirement. Second, we cannot use the conditional to express specifically practical requirements such as the instrumental principle. The difficulty is that the natural suggestion,

> If Shall[φ] and φ only if ψ, then Shall[ψ].

[37] See §4.4.

is not a viable option because it is doubtful that the sentence is logically well-formed.[38] The logic of "shall" doesn't allow us to embed the operator in the antecedent or, for that matter, in the consequent of a conditional. It is true that, the "will" or "shall" of the future tense can be embedded. But this is not true for the expression of intention. That there are major problems doing this can easily be seen by considering the fact that "Shall[φ]" doesn't have a truth-value. Although an intention can be satisfied or disappointed, it is not *true* when satisfied or *false* when disappointed. If it is true that one can falsely attribute an intention to an agent, the agent cannot express a false intention. Embedding, on the other hand, requires truth-apt expressions.

Thus although we can understand what rational requirements are on the model of the conditional, a rational requirement ought not to be understood as a kind of conditional but instead as represented by the inference-schema. The conditional is an elegant instrument to make explicit inferential relations between propositions, but its use, however versatile, is not universal: it cannot be used in general to express rational requirements on the content level. The reason might be that rational requirements are very general, compared to the relations commonly expressed using conditionals. As to why the conditional isn't capable of doing this job, we can only speculate. Our natural language hasn't developed an elegant tool such as the conditional to make explicit the very general kinds of inferential relations involved. We have to make do with inferential relations that cannot be put into conditional form.

Further, inferential relations are governed by ought-to-be rules.[39] Accepting the premises of *modus ponens* in a sense compels you to accept the consequence as well. The compelling force is not, of course, the force of psychological or physical necessity. It is rationality that forces you to accept the claim. This doesn't mean that the agent cannot possibly accept the premises without admitting the conclusion. But it does mean that there is something wrong rationally speaking if he doesn't make this inferential step. At the fundamental level, the principles that ground MP and IP are inference licenses (and sets of inferential prohibitions).

In keeping with the distinction introduced earlier between consequential and acknowledged commitments, we can see how we can criticize an agent for failing to honor a rational requirement.[40] By making one claim X, a speaker undertakes a further commitment Y without necessarily being aware that he does. In other words, the speaker need not

[38] Cf. Sellars (1968: ch. 7).

[39] See §4.1.

[40] For the distinction, see §5.3.

acknowledge a consequential commitment. Applied to rational requirements, we can say that the agent is already committed to the conclusion, even if he is not prepared to acknowledge it or doesn't actually follow out the inference.

Consider an agent who holds that Argos is a dog and that dogs are mammals. Rationality requires that he also believe that Argos is a mammal. This means that he is consequentially committed to the claim that Argos is a mammal. He may not realize this. He may even believe that Argos is not a mammal. In that case his irrationality would be evident, as he would be committed to a claim and to its contradiction at the same time. Again suppose that the agent intends to clean the kitchen and that he believes that this entails taking out the trash. This means that he is consequentially committed to taking out the trash. He may not have noticed this, but it is a consequence of his other commitments, one that he cannot evade. Whether he acknowledges the practical commitment or not, it is required by principles of rationality.

This can be exploited by a critical observer when an agent doesn't accept the conclusion of the inference-schema corresponding to MP or IP. He can remind the agent of his original commitment — for example to p or to doing φ — and of the consequences he has to accept on account of it. In fact the agent has already undertaken the commitment to p or to do φ by endorsing other commitments. The critic can point out that by failing to acknowledge the consequence, the agent fails to do something that he is rationally supposed to do. The relevant inferences aren't optional, and it is not entirely up to the agent whether or not to acknowledge the commitment, given that he has bound himself to (what presents itself as) the premises of the inference-schema. By appealing to rational requirements we can urge a subject to acknowledge his own rational obligations.

Interpreting rational requirements in the way just sketched meets the three conditions we started with. To begin with, (i) rational requirements are content-directed because as inferential practices they operate on the object level rather than on the attitude level. What is more, (ii) they provide guidance to the subjects. It is true that we do not necessarily grasp the principles, but they are responsible for they way we reason in the form of ought-to-be rules. As one writer usefully puts it, although we do not reason *from* the principles, we reason *with* the principles.[41] When an agent reasons from a proposition, she needs to have it before her mind,

[41]Hussain writes that "an agent S reasons from one set of attitudes to another by reasoning with principles of rationality" (Hussain n.d.: 2). He explains that "to reason with a principle of rationality does not require *believing* that a norm, N, is a norm S ought to

but that isn't true for the principles she reasons with. We have seen that the rational principles are implicit in our practice when we treat the relevant transitions as good. We are committed to these principles: we have a rational disposition to conform to the principles and we are committed to the correctness of behavior conforming to them. It is true that we are not committed to the principles in exactly the same way in which we are explicitly committed to a proposition. Instead when we reason with the principles, we are taking part in a practice that involves regarding the inferences permitted by the principles as good. To use a Brandomian phrase, we have an *inferential commitment* to these principles.[42]

Finally, (iii) according to the proposal rational requirements are genuine norms. Being committed to *modus ponens* and other principles is part of what it is to be rational. The reason for this is that the principles are constitutive of the conceptual contents of our intentional attitudes. Part of what it means to be committed to the claim that p is to base one's further reasoning on the premise that p. In particular, it entails being committed to detaching the consequent from a conditional whose antecedent is p. In other words, part of what it means to believe that p is to have an obligation or responsibility to follow *modus ponens* — to be committed to *modus ponens*. Similarly, the instrumental principle is constitutive of having an intention — of being practically committed to doing something. One cannot have a practical commitment without being inferentially committed to the instrumental principle.

6.5 Narrow scope and the Detachment Problem

Equipped with an alternative conception of rational requirements, we can now turn to the two questions raised at the beginning of this chapter: the dispute between narrow-scopers and wide-scopers, and the alleged normativity of rational requirements. Starting with the first, we have already made progress towards an answer. Wide-scope principles are concerned with combinations of attitudes. Now we have seen that some arguments against the existence of wide-scope principles fail. It is not true that the wide-scope formulations of the principles are false. But we

follow or that N is the norm rationality requires S to follow. Indeed, such beliefs by themselves could never be sufficient for reasoning since S would always need to be reasoning *with* some other principle in order to reason from — in order to use in her reasoning — any such belief".

[42]Cf. Brandom (1994: 247ff) for the introduction of the term "inferential commitment" in the context of practical inferences.

have also seen that wide-scope principles have crucial deficiencies. Our theoretical and practical reasoning is evidently guided by principles, but the wide-scope versions cannot fulfill this function.

What has not been shown yet is that in the version in which they are represented by inference-schemas, the principles have narrow scope. To see that they do, notice that the requirements are operative in our actual thought processes. They are *inferential* commitments: they are commitments to make certain transitions. But an inference rule must give us specific instructions as to which contents to accept. In other words, it must be possible to detach a particular conclusion from the rule.

Korsgaard provides a helpful comparison when she asks us to consider what the wide-scope view would mean for cookbook writing.[43] A normal, narrow-scope, recipe for a pasta sauce would be comprised of instructions such as "After you sauté the tomatoes and the mushrooms, you should add a little salt to the mixture." Now imagine the recipe written by a wide-scoping chef. This would not instruct us to use particular ingredients in given circumstances. Indeed the wide-scoping chef would insist that these instructions are faulty. After all, you might have added salty olives or bacon instead of mushrooms, in which case adding salt would be a mistake. The wide-scoper would argue that the instruction should read: "Either you should add a little salt *or* you should previously have added olives by mistake." Evidently, instructions like this would make it hard to actually finish a meal. They do not prescribe which ingredient to add; they only state that certain combinations of ingredients do not go together well. But this is not sufficient for preparing a dinner. Far from guiding our cooking, they would leave us disoriented.[44]

As Korsgaard's analogy suggests, inference rules are not merely concerned with constellations of attitudes because that would mean that the lessons we could draw from the rule would be vague. The wide-scope recipe only tells us that if you add both salt and olives, something is wrong — perhaps you ought not to have added olives, perhaps you ought not to add salt. Similarly from the wide-scope principle, we only learn that something is amiss with respect to the commitments involved, but we do not learn how to go about resolving the conflict. If they allowed no detachment, inferential principles would leave us disoriented and would

[43] Korsgaard (2009).

[44] As Korsgaard points out, "If you hope ever to get your dinner made, you want to avoid recipes written by the wide-scoping chef. If the job of rational requirements is to govern the activities of thought and deliberation, and the point of those activities is to direct us to belief and action, then rational requirements cannot be wide scope, since wide scope requirements cannot do the job" (Korsgaard 2009: 29).

thus be useless. But inference rules are evidently not useless: they form the backbone of our reasoning.

It is true that, as we formulated them, the inference-schemas for MP and IP don't include the word "ought", but that doesn't mean that the question of wide or narrow scope doesn't arise. An inference-schema represents a rule of inference, and all rules involve an ought. We could bring the "ought" out in this way:

> 'p' and 'p $\longrightarrow$ q' ought to be followed by 'q'.

Here the ought cannot take wide scope, because the rule is an instruction of the form "When in situation *X*, you ought to do *Y*." It gives us particular instructions, so it must have narrow scope. Where does this leave us with the original question? Rational requirements are primarily inference rules which are usually implicit in practice. We identified these with inference-schemas operating on the level of content. According to the present view, they are narrow-scope principles. On the other hand, as we have seen in §1, attitude-directed wide-scope principles exist, although they are of secondary importance. These two sets of principles can coexist without causing a conflict. Their function is complementary. Wide-scope principles inform us of a conflict in a given combination of attitudes. But as they only give us vague advice, narrow-scope inference rules are required which govern our reasoning. These latter principles recommend a particular content to accept, which makes them suitable as principles of reasoning.

In their principal form, rational requirements have narrow scope. As we have seen, however, an important motivation for the wide-scope view, and the reason why many philosophers have rejected narrow scope requirements, is that from the narrow-scope formulations, along with simple psychological facts, we can infer purportedly objectionable conclusions. Suppose that David, who intends to get the promotion, believes that the only way to get the promotion is to kill his colleague, Don. The instrumental principle says that, if David has these attitudes, he ought to intend to kill his colleague. Given that David has these attitudes, we can infer that David ought to kill his colleague. But doing so would be cold-blooded murder – surely David ought not to kill his colleague! It cannot be true that David both ought and ought not to kill Don. Using the narrow-scope principle, we have detached a seemingly unacceptable normative statement. A similar problem arises in the theoretical case. Suppose that Sarah believes she has $500 in her wallet and that if she

does, she can afford a new bicycle. *Modus ponens* says that, if Sarah has these attitudes, she ought to believe that she can afford a new bicycle. But how could she? Really her wallet is empty and she can't even afford a bus ticket. The narrow-scope principle has enabled us to detach a seemingly unacceptable normative statement.

To avoid these unwelcome consequences, we do not need to give up the narrow-scope view in favor of a wide-scope view. The natural reaction to the Detaching Problem – and, I think, the right reaction – is to appeal to the distinction between two kinds of oughts.[45] When we say to David: "You really ought not to kill your colleague", we mean that doing so would be wrong no matter what his own psychological state is. What we mean is that David has a substantive or objective reason not to act, and he has this reason whether he knows it or not. Furthermore, David's objective reasons for not killing his colleague are conclusive: they cannot be overridden by other considerations. On the other hand, without giving up our view that murder is unacceptable, we could still say to David something along the lines of: „From the point of view of your own attitudes, you ought to kill your colleague." In other words, although it would be objectively wrong for him to commit murder, David nonetheless ought subjectively to kill Don. The subjective ought flows from his own commitments and does not directly track his moral and other substantive reasons.

It is not hard to see how to apply the distinction of two sorts of ought to the detachment problem. It is true that the narrow-scope instrumental principle allows us to detach the normative conclusion that David ought to kill Don, but the "ought" detached in the conclusion should not be confused with the substantive ought of objective reasons. Rather, it only follows from the principle that David ought subjectively to kill his colleague. Similarly, it is true that narrow-scope *modus ponens* permits detaching the conclusion that Sarah ought to believe that she can afford a bicycle. But she doesn't have a substantive reason to believe this proposition; she only has a subjective reason, and subjective doxastic reasons do not track truth. "X subjectively ought to φ" means that it would be irrational for X not to φ. For this reason, the subjective ought can also be called the ought of rationality.[46]

However, Broome regards the distinction between subjective and objective ought as dubious. He remarks that he lacks a firm grasp of what it

[45] Cf. Kolodny (2005).

[46] Some philosophers reserve the word "ought" to substantive reasons or oughts. For them, there is no ought of rationality. Instead of saying that *S* ought subjectively to do φ, they say that rationality requires S to do φ. This is just a terminological difference. As I find it natural to speak of subjective oughts, I will use the word and disambiguate if necessary.

means to have a subjective rather than an objective reason, or to be subject to a subjective ought rather than an objective ought. As he writes, he is clear about the meaning of "X ought to do (or believe) Y" but he claims not to be clear about the change of meaning when we say instead that "X ought only subjectively to do (or believe) Y".[47] I disagree with Broome's contention that the notion of a subjective ought is incomprehensible. Hence it will be useful to give a more detailed account of the distinction. We can take his remarks as a challenge to clarify the distinction.

Before proceeding to a defense of this difference, however, we need to consider an objection raised by Jonathan Way.[48] According to Way, we cannot escape the Detachment Problem by introducing the ought of rationality since in some cases, even with the distinction in mind, it can still be problematic to say that the agent ought subjectively to intend the objectionable action or believe the objectionable proposition. In our example, Sarah believes wrongly that she has $500 in her pocket. But now suppose that here idea is not just wrong but outright irrational — in Way's words, "obviously crazy". Then the original belief could be the result of a non-rational belief-forming process such as self-deception or wishful thinking. In that case, Way contends, we can no longer say that rationality requires Sarah to follow the consequences of the belief because that belief is itself irrational. It strikes Way as unintuitive that Sarah is subject to an ought of rationality derived from an irrational premise.

We should be clear about what from Way's perspective is unintuitive about the result. Owing to its provenance in wishful thinking, Sarah's resultant attitude, the belief that she can afford the bike, is itself irrational. And according to Way, rationality cannot require you to hold an irrational attitude. The idea is that the fact that the premise was due to wishful thinking is a taint that, by way of the argument, is transferred from the premise to the conclusion. However, this idea should give us pause. For one thing, it raises the question whether the intuition that the requirement cannot mandate a belief with irrational ancestors depends

[47]Broome writes:

> "Ought" is our most basic normative term. I understand it well. But "subjectively ought" and "objectively ought" are philosophers' terms, and their meaning needs to be specified. What is the meaning of "You subjectively ought to *F*?" It is evidently supposed to assign some normative property to your *F*ing: the word "ought" indicates that much. "You subjectively ought to *F*" is supposed to say something a bit like "you ought to *F*". But what like it, exactly? (Broome 2005: 326)

[48]Way (2011: 228f).

on how far back in the inferential ancestry of a premise the taint lies. Is it still unintuitive that Sarah rationally ought to have the belief if the belief has, as it were, an irrational great-grandfather? If not, where should we draw the line? If a belief early on in my inferential history was acquired irrationally, surely it should not for that reason be impossible for *modus ponens* to mandate my believing things that derive from it. More generally, it is not clear that irrationality, as a feature of belief, is a property that can be transferred from premise to conclusion in argument. The idea that irrationality is such a property seems to construe a belief's being rational or irrational as an inheritable property. But whether or not a belief is rational can only be decided by checking whether it is justified, which in turn often means that it must be the result of good reasoning. A belief's status as irrational often depends on the inferential processes it was based on, but Way's argument requires that we can identify a belief as irrational globally, without attention to its context.

The rational requirements we are concerned with are local rather than global requirements. I agree with Kolodny, who points out that "[r]ational requirements [...] ought to be local. In each instance in which one is under a rational requirement, what it ought to require of one is to avoid or resolve some specific conflict among one's attitudes — as opposed to, say, to satisfy some global constraint on all of one's attitudes" (Kolodny 2005: 516). Although there might be global constraints that apply to the whole set of our attitudes, it is clear that *modus ponens* and the instrumental principle are about specific conflicts in a certain constellation of attitudes. As a consequence, it is possible that several rational requirements are in play simultaneously. The case Way imagines seems to be of this type. On the one hand, there is a requirement according to which forming the original belief is illegitimate because, as wishful thinking, it lacks any sound evidential basis. On the other hand, *modus ponens* requires us to acknowledge the commitment implied by earlier commitments. If we properly take into account that rational requirements are local, we should not be surprised by the fact that such a constellation is possible. That it follows from *modus ponens* that, as far as this requirement is concerned, Sarah ought to form the belief that she can afford the bicycle is not incompatible with the fact that this may not be right thing to believe all things considered. To assume, as Way seems to do, that individual requirements necessarily yield the substantively right result is to mistake them for global requirements.

If this is right, then Way's objection does not succeed in raising additional worries about the narrow-scope view. Nonetheless it might still seem inappropriate to say that it follows from a rational requirement that an

agent ought subjectively to do things that he really ought objectively not to do. To dispel this appearance, we need to clarify what it means to say that an agent ought subjectively to have a certain belief or intention. The key to the answer is that, as we have construed rational requirements, the inference-schemas operate on the level of content. Rational requirements are inferential commitments, and the inferential relations they represent cover, not attitudes — not the belief that p and the belief that $p \rightarrow q$ and the belief that q — but contents: "p", "p→q" and "q". We should not interpret the rules as telling you that you ought to form the belief that q but, quite simply, as rules that mandate that q.

To return to the example, *modus ponens* entails that Sarah ought subjectively to believe that she can afford the bike. The inference-rule that governs her commitments causes her to put the commitment that q into her commitment-box. The result of her doxastic deliberation is that q. The narrow-scope principle functions on the level of the expression of commitments. So she does not attribute to herself the doxastic commitment that q but rather expresses it in thought. This contrasts with wide-scope requirements, which can only be interpreted as attributions of commitments. With wide-scope requirements, we take the perspective of an observer who attributes commitments to the agent, and the only way an observer can make claims about the reasoning of another agent is by attributing to her commitments and entitlements: they concern the agent's attitudes. The observer can only evaluate or criticize the agent from the outside. By contrast, as narrow-scope requirements, *modus ponens* yields contents that ought to be adopted. This is something only the agent herself is capable of doing, as she is the only one who can reason from within her own system of commitments on the level of content.

To say that Sarah ought subjectively to believe that she can afford the bike is to attribute to her a commitment. It is to say that, whether she knows it or not, she has undertaken a commitment to the propositional content. As we have said, she may not be prepared to acknowledge the commitment, but her earlier commitments, about the contents of her pocket and the conditional assumption, are binding. The observer deploys the subjective ought to attribute a consequential commitment, one that is appreciable from the subject's first-person perspective. By contrast, to say that Sarah ought objectively to have the belief in question is not just to ascribe the commitment but also to endorse it. In this linguistic act, the ascriber not only takes another person to be committed to a belief but also undertakes the commitment in question himself. To do this, for a belief, is to hold the belief in question true.

Notice that the two oughts express two different linguistic acts. Whereas

the objective ought involves a substantive commitment on the part of the ascriber, the subjective ought allows the ascriber to abstain from undertaking any commitments of his own about what should in fact be done or believed: he only expresses a view about what the subject, from his own standpoint, is committed to. Accordingly, the narrow-scope principles do not prescribe what attitudes the subject ought to have, as that would inevitably involve endorsement of the particular contents in question. Narrow-scope principles, which are inferential principles, do not, and could not, prescribe what is the substantively correct thing to believe or do. Their job is to ascribe consequential commitments, which they do because they only say which content follows from the existing commitments, from the perspective of doxastic or practical deliberation.[49]

We can conclude, then, that the Detachment Problem is not a real difficulty. Broome's complaint that rational requirements yield objectionable oughts only seems cogent on the assumption that the oughts in question involve the endorsement of attitudes. On this assumption, the requirements appear to force us observers to agree with the attitudes that rationality requires Sarah to have. But if we recognize that the requirements only yield contents and subjective oughts rather than attitudes and objective oughts, we see that they do not imply that the observer has to agree with what it would be best for the agent to think or do. Rational requirements imply simply that the observer takes the subject to be consequentially committed to a propositional content. What we can detach

[49]Here it is perhaps appropriate to register a reservation about the use of the words "subjective" and "objective" in this context. These expressions have connotations that may mislead. When we say that an agent (really) ought to have a belief, this is no more objective (in the sense of "truly true") than what we say when we say that an agent ought, rationally speaking, to have a belief. In both cases, it is possible that what we are claiming is false, and in both cases what is believed is purported to be correct. Rather than in a possible degree of reality, the difference between the two lies in the *person* who is committed to the correctness or truth of the belief. With a "subjective" ought, the responsibility lies with the agent: it is she who ought to know better. Even by her own lights, she is committed to the claim. An "objective" ought, on the other hand, places the responsibility on the observer: it is possible that an agent ought to believe that p even though she is not aware of it, through no fault of her own. But it is not necessary that the observer's perspective is more objective, or more true, than that of the agent. Likewise, the ought of rationality is not subjective in the sense having an inherently personal quality. It is true that this ought is relative to the point of view of the agent concerned. But objective oughts are also, in a sense, relative to a point of view, albeit not only to that of the subject but also to that of the ascriber. The crucial difference is that when we say that a person ought objectively to believe something, we thereby ourselves endorse the claim, whereas with a subjective ought, although we assign to the person a commitment to the claim, we do not ourselves endorse the claim. The contrast pertains to who is committed to the content. Perhaps the unhappy terminology is to blame, at least in part, for Broome's insistence that he doesn't know the difference between the two types of ought.

from a narrow-scope requirement is not that the person in question ought objectively to have the attitude in question; that would imply that we, as observers, ought to endorse the attitude. As we have seen, it may well be that the agent is committed to a claim that is false or an intention that is foolish, as seen from the observer's perspective. The principle entails not that from the observer's perspective a certain attitude is appropriate, but that from the subject's perspective a certain content follows from other contents held. We can attribute consequential beliefs and intentions without thereby endorsing them.

6.6 The normativity of rational requirements

We have yet to discuss the second puzzle raised at the beginning of this chapter: do rational requirements give us reasons? Broome considers the question whether rational requirements are normative in a series of papers.[50] An answer to this question requires becoming clearer about what Broome means by the word "normative" in this context. We can distinguish three possible interpretations of what an affirmative answer to the question might mean:

a) Rational requirements are *rule-like*. To be subject to a rational requirement is to be subject to a norm.

b) Rational requirements are valid principles that apply to us. Rational requirements have *normative force for us*.

c) Rational requirements provide us with *substantive or objective reasons*. Whenever the antecedent condition is true of an agent, he has some good reason to adopt the attitude the principle requires of him.

Beginning with the question of rulishness, we have already seen that in this rather basic sense rational requirements are evidently normative. As inference principles, they form the backbone of our reasoning processes. The inference-schemas they correspond to have the form of an inference-rule: they mandate which content is appropriate, which content ought to be inferred from which content, and so forth. However, it is not this interpretation that Broome has in mind when asking this question.[51] He accepts that, as a system of requirements, the principles of rationality

[50]Including Broome (1999), Broome (2005), Broome (2007) and Broome (2010).

[51]Broome (2007: 162).

have to do with correctness according to rules. But he notes that some systems of requirements such as the rules of freemasonry, which require you to roll up one trouser-leg in certain circumstances, should not be seen as telling you that you *ought* to do so.[52] On the other hand, there are other sources of requirements that do tell you what you ought to do; these requirements are normative in a stronger, more contentious sense.[53]

Broome's question is whether, unlike freemasonry, rationality falls into the category of *bona fide* normative sources. One way to tackle this question would be to attempt to vindicate rational requirements — to show that the requirements have normative force for us. This interpretation, item (b) in the list above, is not the approach Broome pursues directly, so we will put it aside for the moment. Instead Broome bases his discussion on interpretation (c) and asks whether rational requirements give us substantive reasons. He writes:

> Most of us take it for granted that we ought to be rational — to have the bundle of dispositions and abilities that constitute the faculty of rationality. Most of us also take it for granted that we ought to satisfy various individual requirements of rationality: we ought not to believe it is Monday and also believe it is not Monday; we ought to intend to catch the 12.50, if we intend to get to a meeting and believe catching the 12.50 is the only way to get there; and so on.
>
> There is a genuine question whether these things are so: ought we to be rational, and ought we to satisfy the individual requirements of rationality? (Broome 2005: 321)

Three immediate comments are in order. According to Broome, the question is whether having a rational requirement apply to oneself entails an ought with respect to one's attitudes. To him, that X ought to do φ in turn implies that X has a reason to do φ. In effect the ought amounts to having a reason that trumps all other reasons one might have. The chief point at issue is the Reasons Claim:

Reasons Claim If one is rationally required to φ, then one has conclusive reason to φ.[54]

[52] Broome (2005: 324).

[53] Broome suggests that morality may be one such source of requirements that is normative in the more forceful sense, though he still seems to be undecided whether or not this ultimately will turn out to be the case (Broome 2007: 165). Way (2010) suggests that prudence and epistemic evidence belong to this category as well.

[54] See §1.

The contrast here is between having a *pro tanto* reason, on the one hand, and a conclusive reason, on the other. As the distinction will not be important in what follows, I will assume, with Kolodny and Broome, that the contentious point is whether it is the case that whenever rationality requires an attitude of us, we have a conclusive reason to adopt the attitude.

Second, in the passage above, Broome mentions two distinct claims. The idea that we have a reason to be rational may mean that *in each particular case* in which a rational requirement applies, we *ipso facto* have a reason to conform to the requirement. This claim, which is captured by *Reasons Claim*, contrasts with the claim that there is *in general* a reason to do the things that rationality requires of us. This latter idea is that we have a generic reason to have a certain disposition, namely the disposition to conform to a system of rational rules. Broome takes this second claim to be less contentious and also less interesting. The idea will be relevant below, but for the moment it will be useful to follow Broome in disregarding the idea to focus on the first — the claim that there are particular reasons to follow the individual requirements of rationality, which Broome insists is a genuine question.[55] Broome finds this question intriguing because, in his view, the Reasons Claim may turn out to be false despite its intuitive plausibility, despite the fact that we "take it for granted".

Third, we distinguished above between subjective oughts and objective oughts, and we can correspondingly distinguish between subjective reasons and objective reasons. The Reasons Claim is intended to mean that we have objective or substantive reasons to follow the requirements. This is clear because claiming that rational requirements give us subjective reasons would not be controversial because what you subjectively ought is a content that follows from your other commitments, which is just what rational requirements prescribe. What is more, Broome countenances only one type of ought, so the ought he talks about must be of the substantive or objective kind.[56]

This last point leads us to the chief difficulty for the Reasons Claim: it is hard to see how it can be true in general that rational requirements

[55] This is worth emphasizing because initially Broome did not doubt the reason-giving character of the requirements, as in his early paper, "Normative Requirements" (Broome 1999). Later he came to question this assumption. The change of mind is reflected in the terminology used in later papers. Whereas in the early paper he is happy to call rational principles like *modus ponens* "normative requirements", in later papers (Broome (2005), Broome (2007)) he chooses to speak of "rational requirements" instead, thereby leaving open the question whether the principles really are normative in his sense.

[56] That Broome has more in mind than just the ought of rationality is also clear from his examples (Broome 2005: 332–333).

can give us objective reasons. Sometimes we are rationally required to take the means to an end although the end itself is foolish. In this case, it seems, there is no substantive reason to pursue the end, and there may be substantive reasons not to pursue it. But then neither is there a substantive reason to take the means, whose only claim to being something we ought to pursue derives from its relationship to the end. It would seem that it is not true that we always ought substantively to do what we are rationally required to do because sometimes we have no reason at all to follow the dictates of rationality.

The same point can be made using scorekeeping vocabulary. If I say that you ought objectively to undertake a commitment, I thereby endorse the commitment. But surely it is possible to attribute to you the consequential commitment without thereby endorsing it. This is what we do when we say that a rational requirement applies to another subject. But if so, then the Reasons Claim seems to lead us back to the implausible thesis rejected above that in saying that a requirement applies, we necessarily endorse the commitment mandated by the requirement.

It would, however, be too rash to give up the Reasons Claim immediately. In response to the obvious difficulty, Broome writes:

> if rationality is indeed normative, that seems likely to be because of what we can achieve by being rational. It seems likely to be for instrumental reasons, as I shall put it. (Broome 2007: 171)

The idea is that we have a reason to be rational because being rational is useful. The usefulness of rationality lies in the fact that it enables us to achieve goals that we actually have reason to achieve, goals that we ought substantively to achieve. This is what Broome has in mind when he writes:

> In general, there are some *G*s such that you ought to *G*. Satisfying the requirements of rationality seems plausibly a good way of coming to *G* in many instances when you ought to *G*. (Broome 2007: 171)

Notice that, except for some very simple goals, we have to engage in instrumental reasoning in order to achieve our goals. If my goal is to eat and I have nothing but a coconut, then I need to crack open the coconut in order to satisfy my hunger. Suppose that it is true that I ought to eat. However, imagine also, *per impossibile* no doubt, that I lack instrumental

rationality entirely. In that case, I would not make the inference required, from "Let me eat" to "I shall crack open the coconut". Perhaps it would not even occur to me that cracking the nut is what I need to do. As a consequence, the fact that I do not engage in correct instrumental reasoning would lead to my failing to do something I ought to do, viz. eating. So the fact that, contrary to this imaginary scenario, I do honor the instrumental principle is useful for my doing what I ought to do and my getting what I ought to have. Being rational opens up these possibilities to me.[57]

Still by all accounts this is only one half of the story.[58] The description of the case included the assumption that it is in fact the case that I ought to eat. But suppose that, as a matter of fact, this is not so: eating really would be foolish — perhaps I have already eaten enough or someone else is more in need of the available food than I. In that case, doing what the instrumental principle instructs me to do is not something that I ought to do. Doing the action might even prevent me from performing a different action that I objectively ought to do. If so, conforming to the requirement is not, to put it simply, a good thing: it does not help me achieve goals that I ought to pursue and might even actively hinder me from achieving them. Surely in such a case, all things considered it would be better if I didn't follow the requirement.[59] It may be that doing what rationality tells you to do leads you to do things you ought to do in many cases, but it clearly isn't always or necessarily so. It is not in general true that, in any given instance, you have a substantive instrumental reason to follow the dictates of rationality.

We should concede that in a majority of cases we have a reason to do what rationality prescribes. According to many philosophers, this means that we have a general reason to adopt the attitudes required by rationality — a reason to have a tendency to be rational. That is what Broome is thinking of when, in the passage cited above, he says that we plausibly have a reason "to have the bundle of dispositions and abilities that constitute the faculty of rationality". We are tempted to say that we are better off with the capacity to be instrumentally and epistemically rational than without it, because if we didn't have that capacity we would never reach any goals. Then it would seem to follow that we have an instrumental

[57]Note that when Broome assumes, in the passage cited, that "there are some *Gs* such that you ought to *G*", he does not take "you ought to *G*" to be tantamount to "you desire to *G*" or some other form of extrarational preference. There may be *Gs* that you ought to do although you have no desire for them.

[58]As Broome readily admits (Broome 2007: 172). See also Kolodny (2005: 543).

[59]Here again we need not construe "better" as "helping me better to satisfy my preferences". Broome assumes that a realist conception of what I ought to do is available that also accounts for the relation of one action being objectively better than another.

reason to be follow the rational requirements.

However, there are two problems with this way of conceiving of the normativity of rational requirements. The first problem is specific to the idea that the substantive reasons we have for being rational are reasons *to be disposed* to be rational. We may concede, for the sake of the argument, that it is useful to have the disposition to be rational and that this usefulness gives us a substantive reason to have this disposition. Still it does not follow that we have an instrumental reason to be rational in particular cases. The general reason to be rational does not transfer to particular instances. Suppose again that I shouldn't really eat the coconut and that, as I know, my eating it requires cracking it open. It may be that I have a general reason to act rationally. Nonetheless in this particular case it is not instrumentally helpful to do as rationality requires. Compare an analogical case. I have good instrumental reason to have the disposition, in general, to drive on the right side of the road in continental Europe. But perhaps today, although the road is almost empty, there is a drunken driver speeding toward me, driving on the wrong side of the road. Then surely today, at least, because of the danger of a crash, I have no reason to follow the right-driving rule. That I have a reason to have a general disposition to follow the traffic rules does not entail that I have a reason, just now, to drive on the right side. Similarly I may have an instrumental reason to be disposed to follow the instrumental principle, but perhaps today my plans are foolish and, if executed, will lead me to trouble. In these circumstances, surely I have no substantial reason to follow the rules of rationality. If a reason is generated by the disposition, it can only be a subjective reason.

Broome is aware of this difficulty but he doesn't consider it a fundamental problem.[60] Although the difficulties move him to become agnostic about the question whether rationality really is normative, he seems to think of the issue as a mere technical difficulty.[61] But even if we waive the difficulties associated with the transfer of reasons to have a disposition to individual reasons, there is a second and more fundamental problem: there are principled grounds for doubt that the normativity of rationality can be constituted by instrumental reasons.

To see what is problematic, suppose that Tom has a substantive reason to get something to eat and that he knows that the only way to get food is to crack open the coconut. According to the instrumental principle, Tom ought (subjectively) to intend to crack open the coconut. Broome

[60]Broome acknowledges the problem in Broome (2007) and Broome (2005).
[61]Broome (2007: 177).

claims that Tom really does have a reason to follow the dictate of the requirement. What justifies his claim that Tom has a substantive reason to open the coconut? His answer is that there are instrumental reasons for following the dictates of rationality. Specifically, following the instrumental principle is part of the best way of achieving something that he ought objectively to achieve, i.e. bringing it about that he gets food. Doing the rational thing has instrumental value, which Broome seems to regard as independently valuable, as something we automatically have a substantive reason to pursue.

However, our topic is the question what reason there is to do what the *instrumental principle* prescribes. If doing the rational thing has only instrumental value, it is something we ought to pursue only to the extent that the instrumental principle is capable of providing us with reasons. But this is exactly the point we were unsure about in the first place. In other words, it seems that Broome's argument presupposes the very principle he is trying to defend. Because Broome's story about how rational requirements give us reasons relies on instrumental values and, thereby, on the instrumental principle, he cannot appeal to the same idea in a defense of the instrumental principle itself. Accordingly, we can once again ask whether Tom really has a substantive reason to do what *this* instrumental connection dictates. Does the fact that it is useful to be rational give him a substantive reason to this? Here Broome seems forced to repeat the account. And when he says that doing what the instrumental principle tells him to do is part of the best way of achieving something that he ought to achieve anyway, he seems already embarked on a vicious regress of justification.

The trouble is not just with the instrumental principle but with the idea of invoking instrumental reasons to support the normativity of rational requirements in general. As we have seen, the status of rational requirements is elementary. They represent inferential relations that we are committed to simply by virtue of having particular doxastic or practical commitments. Rational requirements govern the inferential transitions that make up the very core of our reasoning lives. For this reason, claims about the usefulness of the requirement should be taken with a grain of salt. Our reasoning is governed by the inference-rules whether they are useful or not. As principles with this fundamental status, rational requirements are not susceptible to instrumental justification. On the contrary, all instrumental justification presupposes that a network of inferential relations, including those set up by principles such as *modus ponens* and the instrumental principle, are already in place.

What is more, as we have seen, undertaking a practical commitment by, say, adopting the goal of getting food already involves undertaking further commitments, some of which are determined by rational requirements. In particular, one cannot commit to doing φ without thereby committing to doing ψ should it turn out that ψ'ing is required for φ'ing. In other words, having an intention involves having a responsibility to respect its inferential consequences. The very fact of being practically committed, then, already suffices to conceding authority to the instrumental principle. Given this status, a further justification of the requirement in terms of substantive reasons is redundant. On our conception of rational principles as an implicit practice that can be made explicit using inference-schemas, we need not and cannot find instrumental reasons to support them.

In the light of these problems we ought to give up the project of looking for reasons that support rational requirements. However, we may still wonder why philosophers find this project attractive or why Broome holds that the notion that rational requirements are normative is intuitively true. We can find two motivations in Broome's papers. Imagining an agent who fails to follow a rational requirement, he writes:

> When we are accusing someone of irrationality, we are surely criticizing her. How could we be entitled to do so if there is no reason for her to satisfy the requirements of rationality in the first place? (Broome 2007: 177)

We can see this passage as containing an argument in favor of the claim that rationality must be normative. The idea behind Broome's remark is that we can criticize the agent for an irrational act by citing rational requirements only on the conditions that she has a substantive reason to satisfy the requirement. If it is true that criticism or evaluation inevitably require the existence of a substantive reason to latch on to, our desire to criticize agents for rational failings explains why we should be looking for reasons generated by rational requirements. However, it is simply not true that criticism presupposes a substantive reason. It is true that, in order to criticize an agent for an action or attitude, there needs to be an ought that applies to her and that forbids the action. But this need not be an objective ought that involves the endorsement of the act in question. External criticism draws on all considerations that count against the act. Internal criticism, on the other hand, draws on other commitments of the agent. To criticize an agent internally is to appeal to the subjective

oughts associated with narrow-scope requirements, in which case there is no need to invoke any further substantive reasons.

We can criticize an agent for not making the inference she ought to make according to the inference-rule that underlies the rational requirement. To do this, we can point out to her that, for instance, from "p" and "p→q" the proposition "q" follows, or that from "Shall[φ]" and "φ only if ψ" the proposition "Shall[ψ]" follows.[62] Internal criticism as it were reminds the agent of her inferential obligations. For instance, we may point out to her that in the past she has followed this schema for other inferences. Such a reminder requires no substantive reason on the part of the agent. We need not prove to the agent that it is in her best interest to make the inference; it is enough to display the inference as valid.

There is a second, related motivation of the search for a substantive reason behind rational requirements. If we are looking for a substantive reason to be rational, we might as well ask whether doing as rationality mandates is warranted. In other words, we may doubt the validity of principles such as MP and IP. Thus an agent who consciously fails to conform to *modus ponens* may be interpreted as unwilling to acknowledge the force of the principle. He may ask, "How can we be so sure that the principles we know really are correct?" This skeptical question may elicit an attempt to prove that the requirements of rationality are valid. Here we have reached the final sense of the word "normative" that was mentioned at the beginning of this section. Interpretation of the question "Are rational requirements normative?" is: "Do rational requirements have normative force for us?" This is a demand for the vindication of the principles. It is true that we generally assume without hesitation that our conforming to rational requirements is warranted. This explains why Broome finds it "intuitively plausible" that rationality is normative.[63] We do not usually question the validity of principles such as MP or IP. But as philosophers, we can nonetheless ask if our practice of relying on the principles can be vindicated.

In my view, the vindication of rational requirements is an interesting and worthwhile challenge. In practice we arguably have no choice except to assume that the requirements are valid, but it doesn't follow that the demand for a justification of the inferential rules is illegitimate. However, pursuing this question is beyond the scope of the present project.[64] Before

[62]Cf. Kolodny's transparency account of rational requirements (Kolodny 2005).

[63]Broome (2007: 177).

[64]One way to approach the task is this. We may be prompted to justify an inferential practice when an agent, after being criticized for irrational behavior, refuses to accept the assessment. Refusing to conform to the rational principle presents a skeptical challenge to

concluding, I will therefore only make a final point that connects with what was said above. Thus Broome's exploration of instrumental substantive reason for being rational may be seen as one way of approaching the project of vindicating rationality. On Broome's view, the basic normative unit is the reason for action or *ought*.[65] A general method of vindicating a principle is to identify underlying substantive reasons. In some area, this is in fact possible. If our goal is to vindicate our continental European practice of driving on the right side of the road, a good way to do this would be to find substantive reasons that all the participants in the practice share. To the question "Why drive on the right?" we would reply that such a rule leads to fewer accidents and that we have a substantive reason to avoid accidents. Now on the assumption that everything there is to normativity must have to do with substantive reasons, it is natural to conclude that a vindication even of so basic a principle as *modus ponens* ultimately needs to invoke such reasons.[66] Following this line of reasoning, we are led to ask, as in the title of Kolodny's paper: Why be rational?[67] Broome's assumption seems to be that an answer to this question concerning rationality would not be unlike an answer to the request for a vindication of the traffic rule.

However, from the above it is clear that we cannot agree with Broome about the way to go about justifying the principles of rationality. Vindicating a principle like *modus ponens* by supplying substantive reasons is impossible if, as we have argued, rational requirements do not supply

the validity of the principle in question. The skeptic may question whether it is really true that I ought to accept the claim just because doing so is required by *modus ponens*. We can think of vindicating the principles as a response to a skeptical challenge of this kind. Such a response could take different forms. Hussain proposes to respond to the normative question by quasi-metaethical considerations (Hussain n.d.: 46–56). His approach is similar to the way some writers in metaethics respond to challenges to the institutions and principles of morality. Hussain's suggestion is to start with the observation that the rational principles outlined are the ones we actually use in practice and then to use the Rawlsian method of reflective equilibrium to show that the norms of rationality hang together in a coherent way. If successful, this method would lend support to the whole edifice of principles. A different approach is taken by Nicholas Southwood, who argues that rational requirements belong to the class of what he calls "first-personal, standpoint-relative demands" (Southwood 2008: 28). On his view, the normativity of rational requirements is a matter of our honoring our first-personal authority.

[65] As we saw above, Broome says that he has a good grip on what "X ought to φ" means, whereas he does not know what the role of a subjective *ought* is supposed to be.

[66] Cf. Raz's view: "The normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons [...] So ultimately the explanation of normativity is the explanation of what it is to be a reason, and of related puzzles about reasons" (Raz 1999: 67).

[67] Kolodny (2005). Although he poses the "why"-question in the title of his article, Kolodny does not share Broome's optimism that we have a good answer to the skeptical question that involves substantive reasons.

us with substantive reasons. But this does not mean that we should despair of finding a way of justifying our inferential practice. Normativity comprises more than just substantive reasons. Our conclusion should be, then, that the vindication of rational requirements as inferential principles must proceed in a way that does not rely on the existence of substantive reasons to be rational.

Bibliography

Alvarez, Maria (2008). "Reasons and the Ambiguity of 'belief'". In: *Philosophical Explorations* 11, pp. 52–65.

Ammereller, Erich (2005). "Die Gründe des Handelnden". In: *Rationale Motivation*. Ed. by Erich Ammereller and Wilhelm Vossenkuhl. Mentis.

Anscombe, G. E. M. (2000). *Intention*. Harvard University Press.

Austin, J. L. (1953). "How to Talk – Some Simple Ways". In: *Proceedings of the Aristotelian Society* 53, pp. 227–246.

Ayer, A. J. (1990). *Language, Truth and Logic*. Penguin.

Bilgrami, Akeel (2004). "Intentionality and Norms". In: *Naturalism in Question*. Harvard University Press.

– (2006). *Self-knowledge and Resentment*. Harvard University Press.

Bittner, Rüdiger (2001). *Doing Things for Reasons*. Oxford University Press.

Boghossian, Paul (2002). "The Rule-following Considerations". In: *Rule-following and Meaning*. Ed. by Alexander Miller and Crispin Wright. Acumen.

– (2008). "Is Meaning Normative?" In: *Content and Justification*. Oxford University Press.

Boyd, Richard (1989). "How to Be a Moral Realist". In: *Essays on Moral Realism*. Cornell University Press.

Brandom, Robert (1994). *Making It Explicit*. Harvard University Press.

– (2000). *Articulating Reasons*. Harvard University Press.

– (2001a). "Modality, Normativity, and Intentionality". In: *Philosophy and Phenomenological Research* 63, pp. 587–609.

– (2001b). "What Do Expressions of Preference Express?" In: *Practical Rationality and Preference: Essays for David Gauthier*. Cambridge University Press.

– (2007). "Inferentialism and Some of Its Challenges". In: *Philosophy and Phenomenological Research* 74, pp. 651–676.

Brink, David (1989). *Moral Realism and the Foundations of Ethics*. Cambridge University Press.

Broome, John (1999). "Normative Requirements". In: *Ratio* 12.4, pp. 398–419.
– (2005). "Does Rationality Give Us Reasons?" In: *Philosophical Issues* 15, pp. 321–337.
– (2007). "Is Rationality Normative?" In: *Disputatio* II.23, pp. 161–178.
– (2010). "Rationality". In: *The Blackwell Companion to the Philosophy of Action*. Ed. by Timothy O'Connor and Constantine Sandis. Blackwell.
Brunero, John (2010). "The Scope of Rational Requirements". In: *The Philosophical Quarterly* 60.238, pp. 28–49.
Carroll, Lewis (1895). "What the Tortoise Said to Achilles". In: *Mind* 4.14, pp. 278–280.
Cohon, Rachel (1986). "Are External Reasons Impossible?" In: *Ethics* 96, pp. 545–556.
Dancy, Jonathan (2000). *Practical Reality*. Oxford University Press.
Darwall, Stephen (1985). *Impartial Reason*. Cornell University Press.
– (1997). "Reasons, Motives, and the Demands of Morality. An Introduction". In: *Moral Discourse and Practice*. Ed. by Stephen Darwall, Allan Gibbard, and Peter Railton. Oxford University Press.
– (2002). "Ethical Intuitionism and the Motivation Problem". In: *Ethical Intuitionism: Re-evaluations*. Ed. by Philip Stratton-Lake. Oxford University Press.
Darwall, Stephen, Allan Gibbard, and Peter Railton (1992). "Towards Fin De Siècle Ethics. Some Trends". In: *The Philosophical Review* 101.1, pp. 115–189.
Davidson, Donald (2001a). "Actions, Reasons and Causes". In: *Essays on Actions and Events*. 2nd ed. Oxford University Press.
– (2001b). "Intending". In: *Essays on Actions and Events*. 2nd ed. Oxford University Press.
De Caro, Mario and David Macarthur (2010). "Science, Naturalism and the Problem of Normativity". In: *Naturalism and Normativity*. Ed. by Mario De Caro and David Macarthur. Columbia University Press.
Dennett, Daniel (1981). "Intentional Systems". In: *Brainstorms*. MIT Press.
Dreier, James (1990). "Internalism and Speaker Relativism". In: *Ethics* 101, pp. 6–26.
Enoch, David (2007). "An Outline of an Argument for Robust Metanormative Realism". In: *Oxford Studies in Metaethics* 2, pp. 21–50.
Falk, W. D. (1947). "'Ought' and Motivation". In: *Proceedings of the Aristotelian Society* 48, pp. 492–510.
Firth, Roderick (1952). "Ethical Absolutism and the Ideal Observer". In: *Philosophy and Phenomenological Research* 12, pp. 317–345.
Foot, Philippa (1972). "Morality as a System of Hypothetical Imperatives". In: *Philosophical Review* 81, pp. 305–316.

Frankfurt, Harry (1982). "Freedom of the Will and the Concept of a Person". In: *Free Will*. Ed. by Gary Watson. Oxford University Press.
Gauthier, David (1986). *Morals by Agreement*. Oxford University Press.
Gibbard, Allan (2003). "Thoughts and Norms". In: *Philosophical Issues* 13, pp. 83–98.
Hampton, Jean (1995). "Does Hume Have an Instrumental Conception of Practical Reason?" In: *Hume Studies* 21, pp. 57–74.
Harman, Gilbert (2007). "Moral Relativism Defended". In: *Ethical Theory: An Anthology*. Ed. by Russ Shafer-Landau. Wiley-Blackwell.
Heuer, Ulrike (2004). "Reasons for Actions and Desires". In: *Philosophical Studies* 121, pp. 43–63.
Hornsby, Jennifer (2008). "A Disjunctive Conception of Acting for Reasons". In: *Disjunctivism*. Ed. by Adrian Haddock and Fiona Macpherson. Oxford University Press.
Horwich, Paul (1998). *Truth*. 2nd ed. Oxford University Press.
Hubin, Donald (1999). "What's Special About Humeanism". In: *Noûs* 33, pp. 30–45.
Huemer, Michael (2007). *Ethical Intuitionism*. Palgrave.
Humberstone, I. L. (1971). "Two Sorts of Oughts". In: *Analysis* 32, pp. 8–11.
Humberstone, Lloyd (1992). "Direction of Fit". In: *Mind* 101, pp. 59–83.
Hume, David (1975). *An Enquiry Concerning the Principles of Morals*. Ed. by P. H. Nidditch and L. A. Selby-Bigge. 3rd ed. Oxford University Press.
– (1978). *A Treatise of Human Nature*. Ed. by P. H. Nidditch and L. A. Selby-Bigge. Clarendon Press.
Hursthouse, Rosalind (1991). "Arational Actions". In: *The Journal of Philosophy* 88.2, pp. 57–68.
Hussain, Nadeem. "The Requirements of Rationality". Unpublished manuscript, Aug 2007. URL: http://www.stanford.edu/~hussainn/StanfordPersonal/Online_Papers_files/HussainRequirementsv24.pdf.
Hyman, John (1999). "How Knowledge Works". In: *The Philosophical Quarterly* 49, pp. 433–451.
James, William (1907). "Pragmatism's Conception of Truth". In: *Pragmatism*. Longmans.
Kant, Immanuel (2003). *Kritik der praktischen Vernunft*. Ed. by Horst D. Brandt and Heiner F. Klemme. Meiner.
Kenny, Anthony (1994). *Action, Emotion and Will*. Thoemmes.
Kolodny, Niko (2005). "Why Be Rational?" In: *Mind* 114.455, pp. 509–563.
Korsgaard, Christine (1996). "Skepticism About Practical Reason". In: *Creating the Kingdom of Ends*. Cambridge University Press.

Korsgaard, Christine (2008). "The Normativity of Instrumental Reason". In: *The Constitution of Agency*. Oxford University Press.
– (2009). "The Activity of Reason". In: *Proceedings and Addresses of the American Philosophical Association* 83.2, pp. 23–43.
Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Basil Blackwell.
Lewis, David (2002). "Psychophysical and Theoretical Identifications". In: *Philosophy of Mind*. Ed. by David Chalmers. Oxford University Press.
Mackie, J. L. (1976). *Ethics. Inventing Right and Wrong*. Penguin Books.
McDowell, John (1996). *Mind and World*. 2nd ed. Harvard University Press.
– (1998a). "Are Moral Requirements Hypothetical Imperatives?" In: *Mind, Value, and Reality*. Harvard University Press.
– (1998b). "Functionalism and Anomalous Monism". In: *Mind, Value, and Reality*. Harvard University Press.
– (1998c). "Might There be External Reasons?" In: *Mind, Value, and Reality*. Harvard University Press.
– (2009a). "Conceptual Capacities in Perception". In: *Having the World in View*. Harvard University Press.
– (2009b). "Sellars on Perceptual Experience". In: *Having the World in View*. Harvard University Press.
McGinn, Colin (1984). *Wittgenstein on Meaning*. Blackwell.
Miller, Alexander and Crispin Wright, eds. (2002). *Rule-following and Meaning*. McGill Queen's University Press.
Millgram, Elijah (1995). "Was Hume a Humean?" In: *Hume Studies* 21, pp. 75–93.
– (1996). "Williams' Argument Against External Reasons". In: *Noûs* 30, pp. 197–220.
Milton, John (2005). *Paradise Lost*. Ed. by Philip Pullman. Oxford University Press.
Moore, G. E. (1959). *Principia Ethica*. Cambridge University Press.
Nagel, Thomas (1970). *The Possibility of Altruism*. Oxford.
Nozick, Robert (1993). *The Nature of Rationality*. Princeton University Press.
O'Shea, James (2007). *Wilfrid Sellars. Naturalism With a Normative Turn*. Polity.
Parfit, Derek (1997). "Reasons and Motivation". In: *Aristotelian Society Supplementary Volume* 71.1, pp. 93–130.
Pettit, Philip (2002). "Three Aspects of Rational Explanation". In: *Rules, Reason, and Norms*. Oxford University Press.
Plato (1997). "Meno". In: *Complete Works*. Ed. by John M. Cooper. Trans. by G. M. A. Grube. Hackett Publishing Company.

Quinn, Warren (1995). "Putting Rationality in Its Place". In: *Virtues and Reasons. Philippa Foot and Moral Theory: Essays in Honour of Philippa Foot*. Ed. by Rosalind Hursthouse. Oxford University Press.
Railton, Peter (1997). "On the Hypothetical and Non-Hypothetical in Reasoning About Thought and Action". In: *Ethics and Practical Reason*. Ed. by Garrett Cullity and Berys Gaut. Oxford University Press.
Raz, Joseph (1975). *Practical Reason and Norms*. Hutchinson.
– (1999). "Explaining Normativtiy: Reason and the Will". In: *Engaging Reason*. Oxford University Press.
– (2002). *Engaging Reason*. Oxford University Press.
Robertson, John (2001). "Internalism, Practical Reason, and Motivation". In: *Varieties of Practical Reasoning*. Ed. by Elijah Milgram. MIT Press.
Rosenberg, Jay (1974). *Linguistic Representation*. D. Reidel.
Ross, W. D. (1973). *The Right and the Good*. Oxford University Press.
Russell, Bertrand (1921). *The Analysis of Mind*. George Allen & Unwin.
Ryle, Gilbert (1963). *The Concept of Mind*. Penguin Books.
– (1971). "'If', 'so' and 'because'". In: *Collected Papers, vol. 2*. Hutchinson.
Scanlon, Thomas (1998). *What We Owe to Each Other*. Harvard University Press.
– (2007). "Structural Irrationality". In: *Common Minds. Themes From the Philosophy of Philip Pettit*. Ed. by Geoffrey Brennan et al. Oxford University Press.
– (2010). "Metaphysics and Morals". In: *Naturalism and Normativity*. Ed. by Mario De Caro and David Macarthur. Columbia University Press.
Schmidt, Thomas (2004). "Moral Values and the Fabric of the World". In: *Normativity and Naturalism*. Ontos.
Schroeder, Mark (2004). "The Scope of Instrumental Reason". In: *Philosophical Perspectives* 18.1, pp. 337–364.
– (2007). *Slaves of the Passions*. Oxford University Press.
Searle, John (1983). *Intentionality*. Cambridge University Press.
Sellars, Wilfrid (1957). "Counterfactuals, Dispositions, and the Causal Modalities". In: *Minnesota Studies in the Philosophy of Science*. Ed. by Herbert Feigl, Michael Scriven, and Grover Maxwell. Vol. 2. University of Minnesota Press, pp. 225–308.
– (1963a). "Empiricism and the Philosophy of Mind". In: *Science, Perception and Reality*. Ridgeview Publishing Company.
– (1963b). "Imperatives, Intentions, and the Logic of 'Ought'". In: *Morality and the Language of Conduct*. Ed. by Hector-Neri Castañeda and George Nakhnikian. Wayne State University Press.
– (1966). "Thought and Action". In: *Freedom and Determinism*. Ed. by Keith Lehrer.
– (1968). *Science and Metaphysics*. Routledge & Kegan Paul.

Sellars, Wilfrid (1973). "Actions and Events". In: *Noûs* 7, pp. 179–202.
– (1975). "The Structure of Knowledge". In: *Action, Knowledge and Reality: Studies in Honor of Wilfrid Sellars,* Bobbs-Merrill.
– (1980a). "Behaviorism, Language and Meaning". In: *Pacific Philosophical Quarterly* 81, pp. 3–30.
– (1980b). "Language, Rules, and Behavior". In: *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars.* Ed. by Jeffrey F. Sicha. Ridgeview Publishing Company.
– (1980c). "On Reasoning about Values". In: *American Philosophical Quarterly* 17, pp. 81–101.
– (2007a). *In the Space of Reasons: Selected Essays of Wilfrid Sellars.* Ed. by Kevin Scharp and Robert Brandom. Harvard University Press.
– (2007b). "Inference and Meaning". In: *In the Space of Reasons: Selected Essays of Wilfrid Sellars.* Ed. by Kevin Scharp and Robert Brandom. Harvard University Press.
– (2007c). "Language as Thought and as Communication". In: *In the Space of Reasons: Selected Essays of Wilfrid Sellars.* Ed. by Kevin Scharp and Robert Brandom. Harvard University Press.
– (2007d). "Meaning as Functional Classification". In: *In the Space of Reasons: Selected Essays of Wilfrid Sellars.* Ed. by Kevin Scharp and Robert Brandom. Harvard University Press.
– (2007e). "Some Reflections on Language Games". In: *In the Space of Reasons: Selected Essays of Wilfrid Sellars.* Ed. by Kevin Scharp and Robert Brandom. Harvard University Press.
Setiya, Kieran (2007). *Reasons Without Rationalism.* Princeton University Press.
Smith, Michael (1987). "The Humean Theory of Motivation". In: *Mind* 96, pp. 36–61.
– (1994). *The Moral Problem.* Blackwell.
– (1998). "Galen Strawson and the Weather Watchers". In: *Philosophy and Phenomenological Research* 63, pp. 449–454.
Southwood, Nicholas (2008). "Vindicating the Normativity of Rationality". In: *Ethics* 119.1, pp. 9–30.
Stevenson, C. L. (1952). "The Emotive Meaning of Ethical Terms". In: *Reading in Ethical Theory.* Ed. by Wilfrid Sellars and John Hospers. Appleton-Century-Crofts.
Stocker, Michael (1979). "Desiring the Bad". In: *Journal of Philosophy* 76, pp. 738–753.
Strawson, Galen (1994). *Mental Reality.* MIT Press.
Street, Sharon (2008). "Constructivism About reasos". In: *Oxford Studies in Metaethics.* Vol. 3. Oxford University Press.

Tenenbaum, Sergio (2003). "Accidie, Evaluation, and Motivation". In: *Weakness of the Will and Practical Irrationality*. Ed. by Sarah Stround and Christine Tappolet. Oxford University Press.
Velleman, David (2000a). "The Guise of the Good". In: *The Possibility of Practical Reason*. Oxford University Press.
– (2000b). "The Possibility of Practical Reason". In: *The Possibility of Practical Reason*. Oxford University Press.
Wallace, R. Jay (2005). "Moral Motivation". In: *Contemporary Debates in Moral Theory*. Ed. by James Dreier. Blackwell.
– (2006a). "How to Argue About Practical Reasoning". In: *Normativity and the Will*. Oxford University Press.
– (2006b). "Reason, Desire, and the Will". In: *Normativity and the Will*. Oxford University Press.
Watson, Gary (1982). "Free Agency". In: *Free Will*. Ed. by Gary Watson. Oxford University Press.
Way, Jonathan (2010). "The Normativity of Rationality". In: *Philosophy Compass* 5, pp. 1057–1068.
– (2011). "The Symmetry of Rational Requirements". In: *Philosophical Studies* 155, pp. 227–239.
Williams, Bernard (1973). "Deciding to Believe". In: *Problems of the Self*. Cambridge University Press.
– (1981a). "Internal and External Reasons". In: *Moral Luck*. Cambridge University Press.
– (1981b). "Ought and Moral Obligation". In: *Moral Luck*. Cambridge University Press.
– (1993). *Ethics and the Limits of Philosophy*. Fontana Press.
– (1995). "Internal Reasons and the Obscurity of Blame". In: *Making Sense of Humanity*. Cambridge University Press.
– (2001). "Postscript. Some Further Notes on Internal and External Reasons". In: *Varieties of Practical Reasoning*. MIT Press.
Wittgenstein, Ludwig (1958). *Philosophical Investigations*. Ed. by G. E. M. Anscombe. Macmillian.
– (1975). *Philosophical Remarks*. Ed. by Rush Rhees. Trans. by Raymond Hargreaves and Roger White. Barnes & Noble.
Zangwill, Nick (1998). "Direction of Fit and Normative Functionalism". In: *Philosophical Studies* 91, pp. 173–203.