The Rationality and Normativity of Intention

Jay Jian (Shr-Jie Jian)

Balliol College

Faculty of Philosophy

A thesis submitted for the degree of DPhil in Philosophy

Trinity Term, 2019

The Rationality and Normativity of Intention

Jay Jian

Balliol College; Faculty of Philosophy

A thesis submitted for the degree of DPhil in Philosophy: Trinity Term, 2019

Abstract

This thesis develops a Neo-Humean, functionalist account of practical rationality and normativity. On this view, principles of practical rationality are grounded in the proper function of our agency, while the normative force of a normative reason in calling for a certain response on our part consists in its capacity to engage the function of that response. This functionalist picture, I contend, undermines three long-held and widely-accepted views about practical rationality and normativity. *Contra* orthodox Humeanism about practical rationality, the functionalist account shows that the rational principles governing our intention are not limited to instrumental and formal coherence principles. *Contra* rationalists about the nature of rationality, the functionalist account supports an environmentally-embedded conception of practical rationality and demonstrates that the requirements of practical rationality are not necessary, *a priori* requirements. Finally, *contra* externalists about normative reasons for action, the functionalist account reveals that the normative force involved in normative reasons for action is best captured not by externalist considerations, but instead by our intentions or volitional states.

Table of Contents

Introduction 1

1. The Problem: Practical Rationality and Normativity 1

2. The Starting Point: Intention and Agency 3

Chapter 1 In Defense of the Executive Conception of Instrumental Rationality 9

1. Introduction 9

2. Two Conceptions of Instrumental Rationality 12

3. The Executive Dimension of Instrumental Rationality 18

4. Debunking the Means-end Coherence Conception of IR 23

5. The Supervenience and Boot-Strapping Worry 36

6. Conclusion 46

Chapter 2 The Problem of Practical Disunity and the Integration Principle 48

1. Introduction 48

2. The Constitutive Status of the Principle of Continence 52

3. Intention and Integration 59

4. Integration and Continence 66

5. Defective Intentions 72

6. Conclusion 80

Chapter 3 Misinformation, Subjectivism, and the Feasibility Principle 83

1. Introduction 83

2. The Rationales for Rejecting Misinformed Desire 87

3. The Feasibility Account 95

4. The Feasibility Principle 103

5. Conclusion 111

Appendix 1: Motivation-based versus Liking-based Subjectivism 113

Chapter 4 In What Sense is Practical Rationality Internal? 116

1. Introduction 116

2. Rationality and the Internal Perspective 119

3. Rationality and Coherence 125

4. The Functionalist Conception of Practical Rationality 141

5. The Significance of the Functionalist Conception 156

6. Conclusion 162

Chapter 5 Function and Normativity 164

1. Introduction 164

2. The Structure of Normative Reason 167

3. The Functionalist Account of Normative Force 172

4. The Normativity of Practical Rationality 182

5. The Normative Force of Normative Reasons for Action 186

6. Conclusion 199

Conclusion 202

1. The Rational Principles for Intention 202

2. The Functionalist Picture of Practical Rationality and Normativity 206

3. Final Words 210

References 212

Introduction

1. The Problem: Practical Rationality and Normativity

This thesis is about rationality and normativity. More specifically, it is about the nature of the practical kind of rationality and normativity: What are the rational principles governing our desires? And what are the actions that we have normative reason to do? These are difficult questions. But one trend that we can notice in contemporary philosophy is the general leaning toward *Humeanism about practical rationality and normativity*. On this view, it is our desires that determine what actions we have normative reason to take; yet these desires are only regulated by formal principles of practical rationality, namely, (1) the instrumental principle (which concerns their effective realization) and (2) formal coherence principles (which govern their internal coherence relations). So, on this view, our desires figure as the ultimate sources of practical normativity that are wholly unconstrained by substantive principles of rationality, principles that prescribe us to desire or not to desire certain kinds of end.

This Humean picture represents the received view in decision theory and economics. It is also hugely popular in societies that embrace a moderate form of value pluralism. In fact, even philosophers who disagree with Humeanism broadly accept the instrumental and formal coherence principles proposed by Humeans. It is just that they believe that there are other more substantive principles of practical rationality governing our ends (according to Kantians and neo-Humeans), or that what actions we have normative reason to do are not determined by our desires or mental states in the first place (according to realists). But the Humean picture is still

widely regarded as the most minimal and naturalistically-respectable view about practical rationality and normativity.

Instead of directly arguing that there is more to practical rationality and normativity, my method in this thesis is to first take the minimal Humean picture as our starting-point, and then to unpack the nature and the grounds of Humean rational principles. And my claim is that these rational principles are in fact grounded in the proper function of our agency, so anyone who accepts them has to move beyond the orthodox Humean view and commit herself to a broadly *Neo-Humean, functionalist view about practical rationality and normativity*. This view consists of three central claims, which I will defend in the following five chapters:

(1) *A Neo-Humean view about the rational principles for intention*: Even if we adopt the most austere characterization of agency proposed by Humeans, the rational principles for intention are still not limited to the instrumental principle and formal coherence principles. There are other principles of practical rationality derived from the proper function of our intention and agency (Chapter 1-3).

(2) *The functionalist conception of practical rationality*: Principles of practical rationality are ultimately grounded in the proper function of our agency. An intentional state pattern is proper and prescribed by practical rationality in virtue of how it is constitutively and distinctively involved in the proper function of our agency (Chapter 4).

(3) *The functionalist account of normative force*: The normative force of a reason in calling for a certain response (e.g. belief, intention, or action) consists in its capacity to properly engage the function of that response. Moreover, the normative force of a reason for action resides in its capacity to engage the function of our agency in fulfilling an intention of ours (Chapter 5).

These three ideas, which I will clarify shortly, present the most minimal picture of practical rationality and normativity that we should accept in place of the orthodox Humean view. Given that my main target is the Humean view, I will remain largely agnostic about whether there is anything beyond this minimal picture. That is, in this thesis, I will leave it open whether normative realism is true, or whether there are other more substantive rational principles governing our intention. In fact, the overall functionalist picture proposed here is compatible with both Kantian and realist accounts of normativity. This is because, as we shall see later, Kantians and realists can incorporate, and build on, the functionalist picture by adopting more substantive conceptions of the proper function of agency.

2. The Starting Point: Intention and Agency

In developing my functionalist picture of practical rationality and normativity, I choose to focus on intention. This is because intention plays an integral role in the proper function of our *agency*, and, on my view, it is this particular kind of function that grounds the principles of practical rationality and normative reasons for action. So, what is an intention? Intention, as I take it, is the archetype of a special class of

desire with volitional character.[1] It is primarily defined by its distinctive functional role in our agency: Intention first settles our practical question concerning what we are to do with its all-out, committed volitional character, and it then leads us to realize the prospective state of affairs that we have settled on and committed to. It is the output of our decision-making mechanism, and it in turn provides the input to our executive mechanism for intention-fulfillment. Intention is constitutively involved in the core functions of our agency in making practical decisions and in setting out to fulfill them. That is why the investigation on the nature of intention serves as a good starting point for my functionalist account.

Indeed, as I will first argue, a fuller characterization of the proper function of intention and agency reveals that the orthodox Humean picture regarding the rationality of intention is not entirely correct (Chapter 1-3). Intention aims at its own realization, and it is in virtue of this executive function that intention is governed by the instrumental principle as its core rational constitutive principle. This is what Humeans get right (Chapter 1). However, intention also has a settling function. It also functions to settle, with its all-out, committed volitional character, our practical question concerning what actions we are to do. But an intention resting on practical disunity and inconsistency cannot usually carry enough volitional all-out-ness to settle our practical question and to play this settling function well. That is why there is also a rational *integration principle*, which requires us to form an intention by reaching a kind of practical integration in our first-person perspective (Chapter 2). What is more, intention aims at not just its own realization. An intention's executive function, strictly speaking, consists in how

[1] In this thesis, I will use the term "desire" to refer to any pro-attitude or motivating state that has a world-to-mind direction of fit. In line with this understanding of desire, I take volitional state to be a subclass of motivating state, and intention as a subclass of desire.

it functions to lead us to act and to realize a prospective state of affairs *in virtue of this very intention*.[2] But an intention to bring about a prospective state of affairs that cannot be realized normally cannot play this executive function well. That is why there is a further rational *feasibility principle*, which forbids us from having intentions that are not even minimally feasible (Chapter 3). These two further principles, along with the instrumental principle, are then the core rational constitutive principles for intention, which are respectively derived from its settling and executive functions. Thus, contrary to what Humeans have claimed, the rational principles for intention are not just limited to instrumental and formal coherence principles. In fact, the feasibility principle, as I will explain later, is actually a semi-substantive rational principle precluding intentions with a certain kind of content.

This general picture about the rational constitutive principles for intention in turn provides support for the functionalist conception of practical rationality and a volition-based view about normative reasons for action (Chapter 4-5). As I will argue, a certain intentional state pattern is practically rational neither in virtue of how this pattern is regarded as proper by the agent herself, nor in virtue of how the relations between the intentional states involved in this pattern are proper and coherent in a necessary manner. Rather, an intentional state pattern is proper and prescribed by practical rationality primarily in virtue of its constitutive involvement

[2] I will take the content of an intention to be a prospective state of affairs that we aim to realize through our agency. Although the talk about the intention *to do something* is more natural than the talk about the intention *that we bring about a certain state of affairs*, the former kind of content is actually expandable into the latter. This is because when we intend to perform a certain bodily movement (e.g. intend to wave our hand) or to bring about a certain end result with our action (e.g. intend to make the world a better place), what we intend can be said to be a prospective state of affairs that we aim to bring about (e.g. that our hand moves in a certain way and that the world becomes a better place). I will talk interchangeably between the intention to do something and the intention that we bring about a certain state of affairs. I will explain in Chapter 1 why the content of an intention can often be reformulated in an infinitive phrase.

in the proper function of our agency: We have the requirement of non-contradiction in intention just because it contributes to our intention-formation, which is supposed to resolve our practical ambivalence and settle us on a single course of action; we have the instrumental requirement in virtue of how it is indispensable for our intention-fulfillment; finally, we have the Enkratic requirement just because it realizes a form of normative self-governance over our intentions. This functionalist conception of practical rationality, as I will explain, in turn reveals how the rationalist account of rationality is not entirely true, at least when it comes to practical rationality: Requirements of practical rationality are not necessary, *a priori* requirements that apply to all possible worlds. This is because what these requirements prescribe are the intentional state patterns that are constitutively involved in the proper function of agency. But what these patterns amount to has to be determined by empirical features of the world that we are in (Chapter 4).

I then put forward the functionalist view of normative force to account for the normative force involved in normative reasons for action. On this view, the normative force of a normative reason for a particular attitude or emotional response lies in the former's capacity to engage the latter's function. This view successfully captures our ordinary intuitions about the extension of normative force. For instance, only truth-related considerations can boast the normative force in calling for a belief, because only such considerations are able to properly engage the function of belief, which is to represent what is actually the case. This functionalist account also explains why the normative force of reason can have both a justificatory and a motivating dimension. Indeed, truth-related considerations can both justify our beliefs and exert pressure on us to form a certain belief just in virtue of their capacity to engage the function of belief. Now, given that the function of

action is to fulfill an intention of ours, it follows from this functionalist view that the normative force of normative reason for action lies in its capacity to engage this central function of action in intention-fulfillment. And this kind of normative force, as I will explain, is the practical normative force that we experience from our deliberative perspective and that philosophers have tried to capture. After all, our intention carries our all-out commitment and settles our practical question. Owing to this, considerations capable of engaging the function of our action in intention-fulfillment can typically have the same sort of psychological grip on us and are thus capable of successfully addressing our normative question (Chapter 5).

As you can see here, the central philosophical significance of the functionalist view about practical rationality and normativity lies in how it reveals a new way for us to conceptualize the nature of practical rationality and normativity. In fact, it provides support for a neo-Humean account of practical rationality and normativity, an anti-rationalist picture of practical rationality, and also a volition-based view about normative reasons for action. This functionalist account is also of interest to a broader audience, for it shows that our judgments about people's practical rationality are contingent on empirical features of the psychological mechanism that realizes their agency. Moreover, the functionalist account renders people's desires and life projects answerable to rational critique on grounds other than instrumental irrationality and internal incoherence. Finally, it suggests that our justification for morality should best invoke, not normative reasons for action, but normative reasons for other types of response whose functions are constitutively less sensitive to instrumentalist considerations.

Although this thesis contains a continuous line of argument, each chapter is self-contained and deals with an individual philosophical issue: Chapter 1

investigates the nature of the instrumental requirement and argues in favor of an executive conception of instrumental rationality. Chapter 2 considers the issue of practical disintegration and contends that its problem lies in how it prevents our intention from having its required volitional-all-out-ness and settling force. Chapter 3 in turn deals with the issue of misinformed desire and argues that the problem of misinformation rests in how it renders our intention infeasible and defective in light of its executive function. Chapter 4 examines three competing conceptions of practical rationality and thereby defends a functionalist understanding of practical rationality. Chapter 5 proposes a functionalist account of normative force and outlines a novel argument for a volition-based account of normative reasons for action.

Chapter 1 In Defense of the Executive Conception of Instrumental Rationality

1. Introduction

Instrumental rationality (henceforth: IR) has long been taken to be the most minimal and the least controversial conception of practical rationality, and it provides the very foundation for the influential Humean picture of practical rationality and normativity. Recent studies on the scope of IR, however, have suggested two quite different ways to understand its nature: Wide-scopers have regarded IR as being derived from the consistency requirement on our telic and instrumental intentions (Bratman 2009a; Broome 2013b; Way 2010a, 2011; Wedgwood 2011), or on the beliefs entailed by these intentions (Setiya 2007; Wallace 2001). Narrow-scopers, on the contrary, have contended that IR is partly grounded in the Enkratic requirement, which prescribes us to intend the means to the end that we desire and thereby believe to be valuable under the guise of the good (Scanlon 2007; Schroeder 2009; also see Kolodny 2005).

Virtually no one, however, has taken IR to be a rational executive capacity for realizing our desires. Although Michael Bratman has highlighted how IR encompasses a dimension of effective intention-fulfilment that can't be adequately captured by the consistency requirement on beliefs, he still construes the instrumental requirement as a wide-scope consistency requirement on our telic and instrumental intentions, which is supported by our need in reaching a clear,

unambiguous answer as to whether we really intend to realize a particular end in the future (2009a: 97-8).[1] Joseph Raz has also pointed out that IR may be characterized as the ability for realizing our desires. However, he rejects this conception of IR in the end by claiming that this capacity does not really present a form of rationality distinct from our general rational capacity for responding to normative reasons, including the facilitative reasons to promote what we have normative reason to do (2005: 23-6; also see his 1998: 110).

This is strange. After all, if we ask non-philosophers what IR consists in, most of them will say it's the capacity for realizing our desires effectively. So why is this simple and intuitive view so unpopular among contemporary philosophers? I presume this is because most of them have embraced the general coherentist view about the nature of rationality and thereby committed themselves to the idea that IR must also consist in a kind of means-end coherence in our attitudes.

But in this chapter I will resist this coherentist trend and defend the executive view instead. I begin by explaining how the narrow- and the wide-scope reading of the instrumental requirement can be construed as being grounded in the *executive* and the *means-end coherence* conception of IR respectively (Section 1). I proceed to offer two arguments to support the executive conception of IR and to thereby debunk the means-end coherence conception. As I argue, the means-end coherence conception is problematic since it fails to capture the distinctive executive dimension of IR: It entails that one can still be perfectly rational on instrumentalist terms by always abandoning one's telic intentions—even if one simply lacks the capacity for forming instrumental intentions in the first place, or even if one's

[1] In Bratman's earlier work, he regards this kind of consistency requirement on our intentions as being supported by the contribution of means-end consistency to "our getting what we (rationally) want" (1987: 35).

instrumental intentions cannot even trigger the psycho-physiological process that leads one to act and to realize one's ends (Section 2).

More importantly, the idea of means-end coherence seems capable of grounding IR only because it maps the psychological steps involved in the proper function of our executive system for fulfilling our intentions. But the very idea of means-end incoherence, I argue, fails to apply to some worlds where we have vastly different executive systems for intention-fulfillment, world where our instrumental failures do not manifest themselves in the form of internal means-end incoherence. I in turn consider the proposals to ground instrumental irrationality in the idea of higher-order means-end incoherence or cognitive incoherence. Neither of these incoherent patterns, I contend, can ground instrumental irrationality. I conclude that IR should be thought of as prescribing *not* a particular psychological pattern among our intentional states, but certain psychological steps that are constitutively involved in our intention-fulfillment process (Section 3).

I then explain why the executive conception of IR is philosophically interesting while addressing its attendant shortcomings. *Contra* coherentists about the nature of rationality, the executive conception undermines the received view that rationality strongly supervenes on our intentional state patterns. In fact, it reveals that coherence—at least when it comes to means-end coherence— is not a relation among intentional states that holds necessarily across different possible worlds. *Contra* Humeans about normative reasons for action, the executive conception actually imposes certain internal rational constraints on our intentions, which are the triggers of the instrumental requirement and the main intentional states that can potentially boot-strap normative reasons for action. This, I argue, in turn largely mitigates the boot-strapping problem associated with the executive conception of IR

(Section 4).

2. Two Conceptions of Instrumental Rationality

Rational requirements admit of different readings depending on how we construe their scope (Broome: 1999: 399-403; 2007: 359-60). In the case of the instrumental requirement, the following two main readings are available:

> Wide-Scope Reading (IR-W): it is rationally required of you that [**if you intend to pursue an end E, and if you believe that M is the necessary means to E, then you intend to M**].

> Narrow-Scope Reading (IR-N): if you intend to pursue an end E, and if you believe that M is the necessary means to E, then it is rationally required of you that [**you intend to M**].

The main difference between IR-W and IR-N is this. What IR-W requires of you is to have the intentional state pattern that falls within its scope. So if you intend to pursue an end E, but lack the intention to take the subjective necessary means M (i.e. the means that you believe to be necessary for getting E), then IR-W will require you to acquire the intentional state pattern that it prescribes. This can be achieved either by forming the instrumental intention to M, or by abandoning the telic intention to pursue E (or even by revising the means-end belief about how M is the necessary means to E). In contrast, what IR-N requires of you is just to form the intentional state that falls within its scope, namely, the instrumental intention. Thus,

if you intend to pursue E but lack the intention to take the subjective necessary means M, then IR-N will simply require you to form the instrumental intention to M. Unlike IR-W, IR-N cannot be satisfied by abandoning the telic intention or revising the means-end belief. This is because, as you can see from the formulation of IR-N above, the possession of these intentional states is only the condition for triggering the requirement of IR-N. However, these states do not fall directly within its scope. So to abandon these states is only to prevent the condition for triggering the requirement of IR-N from obtaining; it is not a way of satisfying the very requirement itself (Schroeder 2004: 337-48).

The wide- and narrow-scope distinction has sparked a lively debate in the literature. Many philosophers, however, have implicitly assumed that only one of these readings can present a valid rational requirement deserving the name of the instrumental requirement, while judging the other reading to be confused and lacking rational grounding. So it appears we are forced to choose between the two (see Brunero 2012; Schroeder 2004; Way 2011). This assumption, however, is not entirely true. IR-W and IR-N, I think, present two mutually compatible and equally valid rational requirements, requirements that can be construed as springing from two quite different underlying conceptions of practical rationality.

As many have noted already (Bratman 2009a; Brunero 2012; Schroeder 2004; Way 2011: 227), IR-W is grounded in the *means-end coherence conception of practical rationality*. On this account, being practically rational consists in being coherent in one's attitudes about an end and the means necessary for its fulfillment. I will later examine the idea of means-end coherence under its different construals (Section 4), but the rough idea is as follows. Your telic intention to bring about a prospective state of affairs S constitutively involves your commitment to S's

realization, and it typically entails your belief that you are going to bring about S (or at least your belief that S's realization is possible). However, if you do not correspondingly intend the means that are necessary for S's realization, then you will be inconsistent in your commitment to S's realization, or inconsistent in your beliefs about whether you are going to bring S about (or whether the realization of S is possible). Therefore, you will fall into inconsistency in your attitudes if you intend an end but not the necessary means to it. So the means-end coherence conception of rationality would then require you to resolve such inconsistency by either acquiring the instrumental intention or abandoning your telic intention in the first place (or even by revising your means-end belief). This conception of rationality is embodied in IR-W, which in effect prescribes through its wide scope the coherent combination of these intentional states, and which can be satisfied by any of the three ways aforementioned.

Although all narrow-scopers have modelled narrow-scope requirements after the Enkratic requirement (Scanlon 2007; Schroeder 2009; cf. Kolodny 2005), IR-N, I propose, can be sensibly construed as being founded on the *executive conception of practical rationality*. According to this conception, being practically rational simply consists in taking the psychological steps that are constitutively involved in the proper function of our executive psychological capacity for fulfilling our intentions. It is important to note that, on this executive conception, IR only consists in taking the *psychological steps* that are *constitutively* involved in the proper function of our rational executive system for fulfilling our intentions. In our actual world, these presumably include finding and selecting the means to fulfill our telic intentions,[2]

[2] For a more detailed account about how our instrumental capacity also involves the selection of means, see Wedgwood (2011).

forming instrumental intentions for such means, and then initiating the psycho-physiological process leading to action through the instrumental intentions we have formed. I will elaborate on this in the following section.

Now this more careful characterization forestalls the worry that the executive conception might wrongly identify certain types of non-rational failure in intention-fulfillment with instrumental irrationality (Raz 2005). So, failures in intention-realization resulting from lapses in bodily movement (such as not being sufficiently dexterous or agile when we take the means) do not constitute instrumental irrationality, for they are not failures in taking the *psychological steps* constitutively involved in the proper function of our rational executive system. Nor do other more general psychological lapses (such as inattention or sudden loss of consciousness) constitute instrumental irrationality, unless they actually prevent us from taking the psychological steps that are constitutively indispensable for our intention-fulfillment.

So why can IR-N be satisfied only by forming the instrumental intention, but not by abandoning the telic intention? The leading proposal from narrow-scopers is that IR-N is partly grounded in the Enkratic requirement (Scanlon 2007; Schroeder 2009; cf. Kolodny 2005): We intend things under the guise of the good. Our telic intention to pursue an end typically entails a corresponding normative belief that certain aspects of this end are valuable. This entailed normative belief will in turn trigger the Enkratic requirement, which, according to narrow-scopers, can only be satisfied by intending the necessary means to fulfill our normative belief conscientiously, and not by abandoning our normative belief. This is why there is a corresponding narrow-scope requirement for us to only intend the necessary means to the end that we intend and believe to be valuable—a requirement embodied in

IR-N.[3]

However, this Enkratic view presents an overly intellectualistic and evaluatively-laden conception of IR, which is supposed to be something also shared by agents lacking the capacity to desire things under the guise of the good. Instead of presupposing the controversial guise-of-the-good thesis and invoking the Enkratic requirement, I believe that the executive conception of rationality that I just described actually provides a much simpler explanation for the properties of IR-N, one that better captures the simple and intuitive idea that IR is primarily a capacity for effective desire-realization: So we can say that IR-N is a requirement grounded in our rational executive capacity for fulfilling our intentions. And the reason why IR-N can be satisfied only by forming instrumental intentions but not by abandoning our telic intentions, we can then say, is that the former is a constituent step in a successful intention-realization process, while the latter merely amounts to the abortion or failure in such a process. That is why the executive conception prescribes *only* the formation of instrumental intentions and correspondingly supports a narrow-scope requirement like IR-N.

The two conceptions of rationality sketched here are compatible. In fact, I think they are equally valid as conceptions of practical rationality—being practically rational presumably includes both being means-end coherent and having a rational executive ability for intention-fulfillment. So, the real question here is: Which of

[3] Note that Scanlon accepts a more qualified thesis concerning the rational asymmetry between forming instrumental intentions and abandoning telic intentions (2007: 94-5). According to him, in cases where we have a telic intention for an end but lack an instrumental intention for its necessary means, rationality speaks more on the side of forming an instrumental intention than on the side of abandoning our telic intention. However, if we believe that the necessary means are costly and judge ourselves to lack conclusive reason to pursue the end (given the cost involved), Scanlon claims, then abandoning our telic intention would be the rational thing to do. So, for Scanlon, rational asymmetry holds only when we hold a normative belief concerning how there is a reason for us to pursue the intended end.

these two conceptions best captures our ordinary understanding of IR?[4] It is fair to say that most contemporary philosophers have embraced the means-end coherence conception and attempted to ground IR in a kind of internal consistency, while virtually no one accepts the executive conception. In fact, even narrow-scopers have largely shunned the executive conception in their defence of the narrow-scope reading, and they have instead attempted to ground IR-N (at least partly) in the Enkratic requirement. The main reason why the executive conception is so unpopular, I presume, is that most contemporary philosophers have subscribed to a coherentist picture of rationality, according to which being rational simply consists in having intentional states standing in proper and coherent relations. That is why most of them have endorsed the wide-scope reading of the instrumental requirement and believe that IR is similarly grounded in a kind of means-end coherence between our attitudes.[5]

Nevertheless, I think the executive conception actually presents a more plausible candidate conception of IR. This is because only this conception adequately captures the distinctive executive dimension of IR. More importantly, the idea of means-end coherence seems plausible only because it maps the psychological steps involved in the proper function of our rational executive system. However, it only does so imperfectly, for sometimes there isn't even a distinctive form of consistency after which we can model IR. I will first develop the former, simpler line of argument in order to build an initial case for the executive

[4] Strictly speaking, what I am going to argue in this chapter is not that the means-end coherence conception is not a valid conception of practical rationality at all, but that it is not a conception of practical rationality that can properly ground IR.

[5] Even narrow-scopers can be said to accept the general coherentist picture. It is just the proper and coherent relation that they have been focusing on is the conformity relation between our instrumental intentions and the normative beliefs entailed by our telic intentions.

conception of IR; I will then move to the latter, more radical argument to debunk the means-end coherence conception.

3. The Executive Dimension of Instrumental Rationality

Suppose Porter intends to go to Hawaii this summer. In order to do this, he at least needs to book the flight from London to Honolulu, for it is the only feasible means for him to get to Hawaii. He knows this perfectly well. But now suppose he lacks the instrumental intention to book the flight ticket. How can he be instrumentally rational?

On the coherence conception of IR, Porter can be instrumentally rational by giving up his telic intention to go to Hawaii in the first place. I agree that this is indeed a rational response for him to make from the perspective of coherence. But is that also the case from the perspective of IR? Let us suppose this is indeed the case, for IR just consists in means-end coherence. Now if IR as means-end coherence permits Porter to give up his telic intention in this particular case, then surely it would also allow him to do so in all other cases, since the requirement of IR is supposed to remain constant across different cases. It follows that the coherence conception would license Porter to always give up his telic intentions and never form any instrumental intention.[6] In fact, even if Porter simply lacks the capacity for forming instrumental intentions in the first place, he can still be an instrumentally rational agent on the coherence conception, as long as he can cope with his potential means-end incoherence by always giving up his telic intentions.

[6] Incidentally, the coherence conception of IR also licenses people to always revise their means-end beliefs as a way of being instrumentally rational. For this objection, see Schroeder (2004: 346) and Setiya (2007: 661).

So the means-end coherence conception of IR actually entails that someone who does not even have the capacity for forming instrumental intentions can still be an instrumentally rational agent.

Of course, coherentists have tried to explain why it is irrational to always give up our telic intentions. Some of them have invoked the so-called "basing prohibition" against the revision of our telic intention on the mere basis of our difficulty in intending its necessary means (see Broome 2013a: 140-1; Way 2011: 324-5). This kind of move, however, does not adequately address the central problem here. This is because these coherentists regard the basing prohibition as a requirement that is distinct from IR itself. So, as far as IR goes, always giving up one's telic intention while never forming any instrumental intention is still sanctioned by the means-end coherence conception as one of the possible forms of being an instrumentally rational agent. In other words, the precise problem here is that the coherence conception does not really allow us to rationally condemn this kind of disposition as *instrumentally defective*. And the reply that there are other rational prohibitions forbidding this kind of disposition seems a red-herring.

Furthermore, let us suppose Porter first forms the instrumental intention to book the flight and then forms some further instrumental intentions to carry out this instrumental intention, up until he reaches the last instrumental intention that he could possibly form. This final instrumental intention would presumably be an immediate intention to perform a basic action to which he can find no further means, and to which he can form no further instrumental intentions (say, the immediate intention to move his fingers and dial the number of his travel agent).[7] Now while

[7] Some philosophers of action think that basic actions are mental acts that we can just do, say, the acts of will (Ginet 1990: ch.1-2; Pietroski 2000). This kind of basic mental act, however, is often

the process of carrying out this final, immediate instrumental intention certainly extends beyond one's mind, there is still a gap between the formation of this final instrumental intention and the initiation of the psycho-physiological process leading to action. And ample empirical evidence shows that one's instrumental intention can sometimes fail to trigger the latter process due to interfering psychological factors, such as inattention, pressure, or temptation.[8]

Now, let us suppose that due to sheer pressure, Porter fails to initiate the psycho-physiological process leading to action and resultantly fails to carry out his very last instrumental intention. He repeatedly taps on the screen of his smartphone, but cannot bring himself to carry out his final instrumental intention and dial the number. Is he rational on the means-end coherence conception of IR? Presumably so. He intends to go to Hawaii and commits himself to the fulfillment of this project, and his further commitment to its necessary means has reached the furthest point it can reach, i.e. to his basic action. Although his final instrumental intention fails to initiate the psycho-physiological process leading to action, this still does not compromise the consistency in his attitudes: He commits himself to move his fingers and dial the number of his travel agent, and, in doing so, he has already done his best in being consistent in his attitude toward realizing the goal of going to Hawaii. What happens afterwards is not really relevant from the perspective of coherence.[9]

---

not preceded or accompanied by an intention. We often do not need to intend to will something; we just will it. Because of this, I will here focus on our non-mental acts and take our instrumental intentions to terminate at an intention to perform a basic non-mental act.

[8] For comprehensive surveys on recent empirical studies on the psychological factors that interfere with an intention's capacity in initiating action, see Sheeran and Webb (2016), and Wieber, Thürmer, & Gollwitzer (2015).

[9] Of course, one might want to say that Porter's failure in carrying out his final, immediate instrumental intention to dial the numbers signifies the absence of an immediate intention to do so in

This then suggests that the means-end coherence conception of IR licenses Porter to remain completely unmoved or paralyzed in the face of the gap between his final instrumental intention and the initiation of the psycho-physiological process leading to his action, provided he has already reached the last instrumental intention he could possibly form. After all, his job of maintaining internal consistency has already been accomplished with the formation of his final, immediate instrumental intention; whether this intention can further lead him to act is no longer relevant. So the means-end coherence conception also implies that someone whose instrumental intentions are too weak to resist interfering psychological factors and to trigger the psycho-physiological process leading to action can still be an instrumentally rational agent.

We can draw a general lesson here: IR has an executive dimension that has to do with the effective realization of our intention. But one can remain means-end coherent in ways that do not contribute to one's intention realization at all. That is why the means-end coherence conception allows, implausibly, that someone unable to form any instrumental intentions or unable to initiate her actions by her instrumental intentions can still be an instrumentally rational agent, as long as she can still retain her internal coherence as her saving grace. The executive conception of IR, in contrast, captures this executive dimension quite well, for it explicitly construes IR as prescribing the psychological steps that are constitutively and indispensably involved in our intention-fulfillment process. And this presumably encompasses both the formation of our instrumental intention and the triggering of

---

the first place, so there must still be incoherence in his attitudes toward the realization of his going to Hawaii. This move, however, is quite extreme and implausibly entails that one cannot even fail to carry out one's immediate intentions at all. For, on this view, one's failure in carrying out an immediate intention would simply entail one's lack of that intention in the first place. For an argument against this kind of extreme view, see Korsgaard (1997).

the psychological processes leading to action by our instrumental intention.

Notice that the executive conception of IR does not strictly entail that it is instrumentally irrational for us to abandon our telic intentions. As noted earlier, on the narrow-scope reading of IR, having a telic intention is only a condition whose obtaining will trigger the instrumental requirement. So, on the executive conception, to abandon the telic intention is only to prevent the condition for triggering the instrumental requirement from obtaining; it is not a way for us to satisfy this requirement. Nevertheless, this conception does not forbid us from abandoning our telic intention, nor does it entail that it is instrumentally irrational for us to do so. (The question regarding whether it is rational for us to stick with our telic intention or to revise it, I think, is instead determined by other rational requirements, such as those regarding intention stability and reconsideration; IR only presides over the effective realization of our telic intentions).

Thus, strictly speaking, what the executive conception entails is that (1) always abandoning our telic intentions while lacking the capacity for forming instrumental intentions or for initiating our actions is, on the face of it, neither instrumentally rational nor irrational (after all, if we do not have any telic intentions to begin with, then we wouldn't be rationally required to fulfil any intention on the executive view, let alone be instrumentally irrational). Still, the executive conception allows us to say that (2) what underlies this apparent rational neutrality remains a defect in our instrumental capacity, for we simply lack the capacity to actually take the psychological steps constitutively involved in our intention-fulfillment process and to actually fulfill our telic intention. And my complaint here is that this latter verdict regarding the defect in our instrumental capacity is something that the means-end coherence conception cannot adequately accommodate, for it only construes our

instrumental capacity as a capacity for maintaining internal means-end coherence and fails to capture how IR encompasses an executive dimension that has to do with the effective fulfillment of our intentions.

Of course, one might object that the proper conclusion that we should draw here is not that IR has a distinctive executive dimension (which the coherence conception fails to capture), but that the supposed executive dimension of IR, in so far as it goes beyond mere internal coherence, no longer falls within the ambit of rationality. What it really concerns is effective agency, not rationality. So one might argue that in line with the general coherentist view about rationality, we still need to model IR after means-end coherence. However, I think this kind of modelling still has to be done in light of a specification of the psychological steps involved in the proper function of our rational executive system. In some cases, we cannot even find a distinctive coherence pattern with an instrumental character. This then brings us to the second, and the more radical line of argument for the executive conception of IR.

4. Debunking the Means-end Coherence Conception of IR

4.1. Can IR be Grounded in Means-End Coherence?

It is plausible to think that our rational executive ability for fulfilling our intentions is shaped by our distinctive evolutionary history. If the world had been different, then the psychological mechanism for fulfilling our intentions would have also evolved differently. Suppose we are in a world where our bodies are so unwieldy that the psycho-physiological process for initiating our movements takes up a

disproportionate amount of mental resources. Evolutionary history in this world has then favorably selected a kind of executive system with a built-in "confirming mechanism" preventing the unnecessary waste of the mental resources in initiating movement: In order to fulfill a telic intention, we have to first form an instrumental intention to take the necessary means to it. But in order for this instrumental intention to initiate action, we have to form a further, higher-order desire that we really want the telic intention in question to be realized by devoting our limited mental resources to it. Only with the sanction of such a higher-order approving desire can we trigger the costly psycho-physiological process for initiating movement and actually set out the take the means to fulfill our telic intentions.

So when do we commit an instrumental failure in this world? Presumably it's when we have a telic intention without either the instrumental intention or the higher-order approving desire. Indeed, having a telic intention without either kind of intentional state results in inconsistency in our attitudes. In intending to bring about a state of affairs S, we commit ourselves to S's realization. Yet if we do not also intend the subjective necessary means to S, or form a higher-order approving desire sanctioning the initiation of our actual movement in taking the necessary means, then we would fail to commit ourselves to S's eventual realization. As this case suggests, the most basic form of inconsistency with an instrumental character is as follows:

> **Basic Instrumental Inconsistency**: One is inconsistent in a basic instrumental sense just when one has a telic intention but nonetheless fails to form the intentional states indispensable for one's general intention-fulfillment.

Basic Instrumental Inconsistency is the most intuitive kind of inconsistency with an instrumental character, and it presents the most promising inconsistency pattern in which instrumental irrationality can be grounded. (Of course, there are other kinds of inconsistency that we can potentially model instrumental irrationality after, which I will turn to shortly.) Basic Instrumental Inconsistency, however, has to be characterized partly in terms of the proper function of our rational executive system: This kind of inconsistency consists in having a telic intention without the intentional states which are constitutively involved in the proper function of our executive system, and which are indispensable for our general intention fulfillment. Indeed, what Basic Instrumental Inconsistency amounts to crucially depends on what the proper function of our executive system involves: for people in our actual world, it's [having a telic intention without an instrumental intention], while for people in the first world that I described, it's [having a telic intention without either an instrumental intention or a higher-order approving desire].

But Basic Instrumental Inconsistency fails to find a foothold in some cases. Suppose we are in a world where the formation of a further, distinct instrumental intention (in addition to the telic intention that we already have) takes up a disproportionate amount of mental resources. Evolutionary history in this second world has then favorably selected a kind of executive system with a "specification mechanism": So, in order to fulfill a telic intention, what we need to do is not to form a further, distinct instrumental intention and then act on it, for this is simply too costly. Rather, what we need to do is to turn the content of our telic intention into a more specified overall action plan that encompasses the necessary means we are to take, and then directly act on this plan. To illustrate, suppose you intend to go to an island resort. What you need to do to fulfill this intention is to specify its content

into a more detailed plan (say, that you go to an island resort by travelling on a flight that you booked via contacting your travel agent by dialing his numbers), and then to carry out this action plan directly. In fact, as Bratman has pointed out, we do sometimes fulfill our telic intentions through this kind of specification mechanism in our actual world (1987: 29-31). So I presume the scenario presented here is not completely preposterous.

So when do we commit an instrumental failure in this second world? It's when we have a telic intention but fail to turn its content into a sufficiently specified action plan that incorporates its necessary means (this is presumably equivalent to the failure in having a telic intention but failing to form further instrumental intentions for its necessary means in our actual world). So this happens when the content of our telic intention remains pretty general and underspecified (say, when we merely intend to go an island resort without incorporating further necessary means). But can our under-specified telic intention here constitute a kind of inconsistency? No. Inconsistency is supposed to reside in the relation between more than one intentional state, and having a single telic intention that is not sufficiently detailed isn't a form of inconsistency. Can our under-specified telic intention be said to have an inconsistent content in itself (in the same way that a self-contradictory belief can be said to have an inconsistent content)? No. Although the content of our telic intention fails to incorporate the necessary means we are to take, it doesn't include an explicit contradictory element of not taking the means. Strictly speaking, its content does not involve elements that cannot be brought about at the same time, and saying that its content is inconsistent would surely stretch the meaning of inconsistency.

What this then suggests is that, in the second world, our instrumental failure

(namely, that of having an under-specified telic intention) cannot be captured by Basic Instrumental Inconsistency. There is no such a thing as Basic Instrumental Inconsistency in the second world. After all, this kind of inconsistency consists in having a telic intention while lacking some other distinct intentional states that are indispensable for our general intention-fulfillment. But in the second world, our general intention-fulfillment only requires us to specify the content of our telic intention, but not to form another distinct intentional state.

Now consider a more playful scenario. Suppose we are in a world where we achieved great advances in technology long ago, owing to which we have "desire-realization machines". These machines, like powerful 3D printers, can automatically scan for the prospective states of affairs or objects that we intend to bring about, and then materialize them from scratch. These desire-realization machines will proceed to materialize our intended objects, however, only when they detect that the strength of our telic intentions has reached a certain threshold. Now further suppose that our capacity for forming instrumental intentions has long become useless and degenerate over time (perhaps the desire-realization machines are perfectly reliable), such that forming any further instrumental intentions will just lead to conative overload in us. So, in order to fulfill a telic intention, what we need to do is not to form further instrumental intentions, but to intend something strongly enough, for then the desire-realization machine will pick up on this and help us bring about our intended object. In a way, this third world is the sci-fi version of a world in which we have the God-like ability to bring about what we intend, provided we intend it strongly enough.

The question remains: When do we commit an instrumental failure in this world? Presumably it's when we have a telic intention to bring something about yet

fail to intend it strongly enough. But having an insufficiently strong telic intention surely isn't a form of inconsistency, nor does the content of such an intention involve any inconsistency. So, again, our instrumental failure in this third world cannot be modelled after Basic Instrumental Inconsistency. This idea also fails to have applicability in this world.

The general lesson here is this. We commit an instrumental failure just when we have a telic intention yet *fail to take the psychological steps that are constitutively involved in the proper function of our executive system* and that are indispensable for our general intention-fulfillment. (For the sake of simplicity, in what follows I will call such psychological steps *indispensable psychological steps*.) When such indispensable psychological steps involve the formation of some further intentional states, we can say that [having a telic intention while lacking these further intentional states] constitutes a pattern of inconsistency with an instrumental character. We can then model instrumental irrationality after it. This is how the idea of Basic Instrumental Inconsistency gets a foothold. However, when such psychological steps involve only the working on a single intentional state, as in the second and third worlds that we just considered, we will no longer be able to capture our instrumental failure by the idea of Basic Instrumental Inconsistency. After all, inconsistency is supposed to hold between more than one intentional state. But in the worlds just described, our instrumental failure only consists in our failure in working on our telic intentions in a particular way.

4.2. Can IR be Grounded in Other Kinds of Coherence Pattern?

One might object that although the indispensable psychological steps in the second

and third worlds only involve the working on our telic intentions, the instrumental failures in these worlds can still be captured by other kinds of inconsistency. Indeed, as you might think, the differences in the indispensable psychological steps in the worlds that we have considered only constitute the differences in the *general mental means* that we need to take in order to fulfill our telic intentions. And once we bring in our means-end beliefs about how these psychological steps are the general mental means for our intention-fulfillment, the following candidate inconsistency pattern seems able to ground our instrumental failures in these worlds:

> **Higher-Order-Instrumental Inconsistency**: One is inconsistent in a higher-order-instrumental sense just when one has a telic intention and a higher-order means-end belief that certain indispensable psychological steps constitute the general mental means for one's intention-fulfillment (say, forming further instrumental intentions, specifying the content of one's telic intention, or intending one's end strongly enough), but then fails to form the intention to take these steps.

Can we ground instrumental irrationality in this kind of inconsistency? I don't think so. Suppose there is an agent in our actual world with a rational executive system just like ours: So, in order to fulfill her telic intention, she needs to form further instrumental intentions for its necessary means just like us. But suppose this agent has a massively mistaken view about the indispensable psychological step for her general intention-fulfillment (say, she wrongly believes she can fulfill her intention simply by intending something strongly enough, without the need of forming

further instrumental intentions). Now suppose this agent intends to bring about S but fails to form a further instrumental intention for its necessary means. Will she thereby fall into Higher-Order-Instrumental-Inconsistency? No. Although she fails to form the instrumental intention, she does not really fail to take the psychological steps *that she believes* to be indispensable for her general intention-fulfillment—for she believes that what she needs to do is simply to intend something strongly enough. But should we therefore say that this agent doesn't commit an instrumental failure in this case? Probably not. Given that she has a rational executive system like ours, we should say that she simply commits an instrumental failure when she intends S without forming an instrumental intention, and we should not exonerate her failure here on account of her false belief regarding the indispensable psychological steps for her general intention-fulfillment.

More generally, as John Broome (2013a: 92) and Clayton Littlejohn (2018: 269-70) have pointed out, rational requirements impose "strict liability" on an agent: An agent is simply irrational when she fails to conform to a rational requirement, and her irrationality cannot be excused or rationalized by her false belief about, or ignorance of, what rationality requires of her. Otherwise, we would fall into a form of anarchism about rationality by allowing people to be rational simply by conforming to their idiosyncratic and crazy views about what rationality prescribes us to do. But the idea of Higher-Order-Instrumental-Consistency seems to negate strict liability and entail this form of anarchism about IR. After all, even if the agent fails to take the psychological steps that are *in fact* indispensable for her general intention-fulfillment (say, forming instrumental intentions), she can still excuse her instrumental failure and claim to be rational in a higher-order-instrumental sense by appealing to her crazy higher-order means-end beliefs about what other

psychological steps are the general mental means for her intention fulfillment.

The source of the problem here is this. IR is supposed to prescribe us to take the psychological steps that are constitutively involved in the proper function of our executive system and that are indispensable for our general intention-fulfillment. Yet the idea of Higher-Order-Instrumental-Inconsistency construes theses indispensable psychological steps as a general *mental means* that we can have higher-order means-end beliefs about and that we can intend to take. It thereby allows our own idiosyncratic means-end beliefs (about the indispensable psychological steps we need to take) to trump what IR actually prescribes. That is why instrumental irrationality, I think, should not be grounded in Higher-Order-Instrumental-Inconsistency, for that would license an overly permissive, anarchic view about IR.

One might alternatively suggest that, although our having a telic intention while failing to take the indispensable psychological steps doesn't constitute a form of inconsistency in the second and third worlds, it can still lead to inconsistency in our beliefs. So the following inconsistency pattern in our beliefs might still be able to capture our instrumental failure in those worlds:

> **Cognitive-Instrumental Inconsistency**: One is inconsistent in a cognitive-instrumental sense just when one has a telic intention to bring about a certain prospective state of affairs S and thereby believes that the realization of S is still possible, yet at the same time fails to take the indispensable psychological steps for one's general intention-fulfillment and thereby believes that the realization of S is not really possible.

Can we instead ground instrumental irrationality in this kind of cognitive inconsistency, as cognitivists about IR have suggested (Setiya 2007; Wallace 2001)? The answer is, again, no. Various philosophers have already argued that cognitive inconsistency fails to provide a suitable ground for instrumental irrationality (Bratman 2009a; Broome 2013b). Suppose there is an agent in our actual world with a rational executive system just like ours. And suppose she has a telic intention to bring about a state of affairs S, but then fails to form the instrumental intentions for the necessary means to it. Now further suppose that this agent doesn't notice her failure in doing so. In fact, suppose she falsely believes that she has already formed the instrumental intentions. Now although this agent fails to form the relevant instrumental intentions, there is still no inconsistency in her beliefs: Due to her ignorance or false belief regarding her failure, she still believes that she will bring about S in the future (or that the realization of S is still possible). Again, in this case, I think we should say that the agent simply commits an instrumental failure when she has a telic intention while failing to take the indispensable psychological steps for her general intention-fulfillment (i.e. forming the instrumental intentions), and that her ignorance or false belief about this cannot exonerate such a failure.

More importantly, why is having a telic intention without taking the indispensable psychological steps supposed to result in inconsistency in our beliefs? Cognitivists have argued that in doing so we would end up with inconsistent beliefs about our future actions (or about the possibility of the fulfillment of our intended end). But I don't think this is precisely the right way to characterize the case. First consider the belief constraint on intention. It is quite controversial whether our intention to bring about a state of affairs S entails our belief that we will bring about S in the future (or whether it excludes the explicit belief that we will not bring about

S in the future). After all, we can perfectly intend to bring about S while at the same time doubting whether we will really act on this intention in the future (for we might simply forget it), or whether we will really succeed in bringing about S in the end (perhaps S is something quite difficult for us to realize).[10] That said, I think there is something right about the more moderate idea that our intention to bring about S entails our belief about how S is possible for us to bring about (or at least it excludes the explicit belief about how S is impossible for us to bring about (Bratman 1987: 31)). The rationale for this weaker, and more plausible, construal of the belief constraint, I think, is that our intention involves a commitment to a state of affair S's eventual realization, so it is constitutively sensitive to our belief about whether S's realization is possible for us to bring about. After all, if we know that S is simply impossible for us to realize in the first place, then it would be pointless to intend it and commit ourselves to its realization.

On this more plausible understanding of the belief constraint, the belief entailed by our telic intention should be the belief that the realization of S is *possible and within our power to bring about*. The belief entailed by our failure in taking the indispensable psychological steps for our general intention-fulfillment (say, forming instrumental intentions or specifying the content of our telic intentions), on the other hand, should be the belief that the realization of S *is not possible given that we fail to take the indispensable psychological steps*—for our failure in doing so doesn't make the realization of S impossible *tout court*. But, as we can see here, these two beliefs are not really inconsistent. Rather, they only reflect our poignant awareness that while we commit ourselves to bringing about something within our power, we

[10] For arguments against the strong belief condition on intention, see Bratman (1987: 37-41); Davidson (1978); Kolodny (2008a: 372-3); and Wallace (2001).

will eventually fail to do so, given that we neglect to take the indispensable psychological steps. These, I believe, are the cognitive states that we are caught up in when we commit an instrumental failure and when we become aware of it. In a way, we can say that these cognitive states are the bearers of the bad news of our instrumental failure. But they do not involve any inconsistency in themselves. Even if they did, the instrumental failure still lies in our failure in taking the indispensable psychological steps for our general intention-fulfillment, not in the beliefs resulting from our awareness of such a failure. Indeed, if we insist that instrumental irrationality has to consist in the alleged inconsistency among these beliefs, then we will be in effect labelling the messenger of the bad news of our instrumental failure as instrumentally bad in itself. That is why I believe that instrumental irrationality cannot be modelled after Cognitive-Instrumental Inconsistency. Otherwise we would wrongly identify the cognitive awareness of our own instrumental failure as inconsistent and instrumentally bad in itself, when in fact it involves neither problem.

4.3. Why IR Cannot be Reduced to Inner Coherence

I think it is safe to conclude that the project of modelling IR after a kind of coherence in our attitudes is not promising. As we have seen, Higher-Order-Instrumental Inconsistency or Cognitive-Instrumental Inconsistency cannot provide good grounding for instrumental irrationality. This is because they either entail an overly permissive view about IR or locate our instrumental failure in the wrong kind of inconsistency pattern. The most plausible inconsistency pattern that may alternatively ground instrumental irrationality is Basic Instrumental Inconsistency,

that is, the inconsistency in [having a telic intention without the intentional states that are indispensable for one's general intention-fulfillment]. But what this kind of inconsistency amounts to has to be determined by what is involved in the proper function of our executive system. More importantly, this kind of inconsistency simply fails to have a foothold in the second and third worlds we considered earlier, that is, in worlds where we fulfill our telic intentions not by forming further intentional states, but by working directly on our telic intentions.

More generally, the means-end coherence conception of IR appears to be a viable conception of IR only because it maps (albeit imperfectly) the psychological steps constitutively involved in the proper function of our rational executive system. When such psychological steps involve the formation of further intentional states, we can say that [having a telic intention without these further intentional states] constitutes a pattern of means-end inconsistency. However, in worlds where our executive systems for intention-fulfillment do not work in this way, this kind of mapping will fail to hold—for our instrumental failure in these worlds no longer manifest itself the form of a kind of inconsistency. And if we want to stick with the general coherentist view of rationality and identify irrationality solely with inconsistency, then we will have to conclude that in these worlds there isn't even such a thing as *instrumental irrationality/inconsistency*. But this is implausible. After all, our counterparts in these worlds can perfectly commit instrumental failures just as we can. And to say that the idea of instrumental irrationality can have no applicability to their worlds is, in effect, to force the notion of IR to fit into the Procrustean bed of coherence. The executive conception of IR, in contrast, provides a more theoretically -unified and intuitive answer as to what instrumental irrationality consists in, i.e. the failure in taking the psychological steps

constitutively involved in the proper function of our rational executive system. This answer applies neatly to all the worlds we have considered and better carves IR at its joints.

5. The Supervenience and Boot-Strapping Worry

5.1. The Supervenience of IR on the Mind

Given the two lines of argument that I have put forward here, I think we have good reason to accept the executive conception of IR in lieu of the dogmatic coherence conception. In fact, the executive conception suggests some interesting and novel ways for us to reconceptualize the supervenience of rationality on the mind as well as the boot-strapping problem. This in fact allows us to address two main worries about the executive conception: The first worry is that the executive conception seems incompatible with the thesis of the supervenience of rationality on the mind. After all, if IR is about effective intention-fulfillment, and if the fulfillment of our intentions depends on factors outside our mind, then how can IR still supervene on the mind (Broome 2013a: 151-2)? The other worry is that the executive conception allows for a form of boot-strapping. This is because on this conception of IR, a mere telic intention can just trigger a narrow-scope requirement for us to (only) intend the necessary means to it. So if we really have a normative reason to follow this narrow-scope instrumental requirement, then we would also seem to have a further normative reason to intend the necessary means to our telic intention, even when it is wholly foolish or immoral.

Let us begin with the supervenience thesis. I think the executive conception of

IR is actually compatible with this idea under some of its weaker readings, while it still casts serious doubt on stronger versions of this thesis.

To begin, on the executive conception, two agents in two different possible worlds with the same intentional state patterns and engaging in the same mental processes can still differ with respect to their IR. This is because, owing to the differences in the evolutionary history in different possible worlds, the exact same psychological patterns and processes can be indispensable for people's general intention-fulfillment in one world, but not so in another. In the first world described in the foregoing section, it is instrumentally irrational for us to hold a telic intention without a higher-order approving desire, whereas this is not the case in our actual world. Therefore, IR does not strongly supervene on the *mental patterns* that one exhibits and the *mental processes* that one engages in. But there is nothing controversial about this idea. What is controversial is rather the strong supervenience thesis and the thought that once a particular mental pattern and mental process can be called instrumentally rational in one world, it must also be so in all other worlds, regardless of whether it can really contribute to people's general intention-fulfillment in some of these worlds.[11]

In fact, even if we retreat to the idea of means-end coherence, IR still does not strongly supervene on one's intentional state patterns. Two agents in two different possible worlds with exactly the same patterns of intentional states might still differ with respect to their means-end coherence. This is because what means-end coherence amounts to, as we saw earlier, is in fact determined by what kinds of intentional state are indispensable for one's general intention-fulfillment, and this

[11] This thesis about strong supervenience is endorsed by Broome, who formulates wide-scope requirements as coherence requirements that apply to all possible worlds (2013a). Also see Way (2010b) and Titelbaum (2015). I will discuss this view in more detail in Chapter 4.

may vary in different worlds with different evolutionary histories. So [having a telic intention without a higher-order approving desire] is means-end incoherent in the first world we just considered, but it is perfectly means-end coherent in our actual world. Therefore, means-end coherence also does not strongly supervene on one's mental patterns.

This in turn challenges the traditional way of conceptualizing the idea of coherence. Almost all philosophers have taken coherence to be a relation among intentional states that holds across different possible worlds: After all, whether a particular intentional state pattern is coherent is supposed to be determined solely by the internal relations between the intentional states involved in such a pattern; therefore, once a particular pattern of intentional states is coherent in one world, it must also be so in all other worlds (see Broome 2013; Way 2010b). But this idea is actually not true, at least when it comes to means-end coherence. In fact, as I shall argue in Chapter 5, whether a pattern of intentional states is proper and prescribed by practical rationality is determined by whether it is constitutively involved in the proper function of our agency. But whether this is so is contingent on empirical features of the psychological mechanism that realizes our agency. So what constitutes a practically coherent intentional state pattern can change from world to world.

That being said, I think there is still some truth to the idea that rationality strongly supervenes on all of our mental properties, when these are taken to include our *mental capacities*. The executive conception is perfectly compatible with this idea. If two agents have the same kind executive system for fulfilling their intentions, and if they have followed the same psychological processes and exhibited the same psychological patterns, then they will be indeed be indistinguishable with respect to

the proper functioning of their executive system, and hence their instrumental rationality.

What about the weak supervenience thesis? Is it compatible with the executive conception of IR? I think the answer is both yes and no: It actually depends on how we understand the idea of proper function. According to the executive conception, IR requires us to take the psychological steps that are constitutively involved in the proper function of our rational executive system. But the idea of proper function can be understood in either an *information-sensitive* or an *environment-sensitive* way: We can say that the proper function of our executive system involves intending the means that *we believe to be able* to fulfill our telic intentions, in light of the information given by our means-end belief. Alternatively, we can say that its proper function involves detecting and intending the means that are *in fact able to* fulfill our intentions. Correspondingly, IR can be taken to prescribe either of these psychological steps.

Consider two agents in the same possible world, say, in our actual world. Suppose they are mentally indiscernible: They have exactly the same type of rational executive system, are indistinguishable with respect to their intentional states, and have taken exactly the same psychological steps. Suppose both of them intend to cool down their room and have then formed an instrumental intention to open the window in response to their means-end beliefs. But suppose, owing to differences in their external environment (say, one of them is in Bangkok), opening the window is actually a wrong measure for one of them to take. Now, both agents have formed their instrumental intentions in light of their means-end beliefs, so both would indeed be indistinguishable with respect to the proper function of their executive system, when it is understood in an internalist, information-sensitive way.

However, only one of them has successfully detected the right measures and formed the instrumental intention for the means that can in fact cool down the room. So they can still be said to be distinguishable in terms of the proper functioning of their executive systems, when it is understood in an environment-sensitive way.

I personally prefer the verdict that these two mentally indistinguishable agents are actually distinguishable with respect to their instrumental rationality. However, I realize that most philosophers would prefer to say the opposite out of their intuition in favor of the weak supervenience thesis (I am less sure about which side non-philosophers would take on this matter). So here I shall only make the more modest claim that the executive conception of IR at least is compatible with both views, for they can be accommodated by the two different conceptions of proper function that I have highlighted here. This kind of ecumenism, I take it, is a strength of the executive conception of IR. And this in turn explains why this conception need not compel us to wholly discard the supervenience thesis, thereby addressing the concerns that some philosophers might have.

5.2. The Problem of Boot-Strapping

Now let us consider the worry of boot-strapping. As noted above, the narrow-scope reading of the instrumental requirement potentially leads to a form of boot-strapping. Most philosophers have tried to avoid this either by embracing the means-end coherence conception of IR instead (Bratman 2009a; Broome 2013a; Way 2011), or by denying the normativity of IR (Kolodny 2005). But neither option is ideal: The means-end coherence conception is not a plausible conception of IR to begin with, and, as many have already pointed out, it also entails the same kind of

boot-strapping consequence in cases where our telic intentions are impossible for us to abandon or modify (Raz 2005; Setiya 2007). Denying the normativity of IR, on the other hand, seems a bit extreme. After all, what the executive conception of IR requires us to do is to take the psychological steps indispensable for our general intention-fulfillment, rather than to remain means-end coherent just for coherence's own sake. So the executive conception of IR still seems to serve a point and is arguably normative for us.

The most promising alternative way to address the boot-strapping worry, I think, is to instead reconsider whether the boot-strapping consequence, together with the subjectivist picture of normative reason that it leads to, is really as unpalatable as is often assumed. And I think we have good reason to doubt this. Perhaps somewhat paradoxically, the executive conception of IR actually allows for a more sensible subjectivist picture of normative reason. This is because once we arrive at a more detailed account of the proper function of IR, we will actually see that IR is only supposed to take as its input a particular kind of pro-attitude that arguably has some reason-giving force.

To begin, misinformed desires or intentions based on false means-end beliefs thwart our intention-fulfillment and have no place in the proper function of IR, if we construe its proper function in an environment-sensitive way. So our IR as a rational executive system, if it functions properly, will only take telic intentions and correct instrumental intentions as its main inputs, correspondingly imposing on us the instrumental requirement and the normative reasons to intend the means to them. This then echoes the Humean tenet that it is our intrinsic desires that determine what normative reasons we have through a kind of objective means-end extension.

What is more, our rational executive system seems to take desires with volitional character as its primary inputs, which include intentions, decisions, commitments, volitions as its subspecies (For the ease of discussion, here I will take intention as a placeholder for desire with volitional character).[12] So why is intention distinctive? Intention is distinctive because of its intermediary causal role between the decision-making process and the executive process in the production of action: An agent's potentially conflicting desires are first weighed against one another, with some of them being rejected or outweighed in her decision-making process. The agent's decision-making process then concludes with an intention, which settles her practical question concerning what to do and carries her commitment to a particular course of action. This intention in turn initiates the executive process. What is left for the agent to do is to set out to fulfill that intention.[13] This intermediary role played by intention is useful for the well-functioning of the agent's rational executive system. So instead of having conflicting desires that drive her executive system to different or even incompatible courses of action, the agent has intentions that arise from the partial resolution of the conflict in her desires and that direct her rational executive system in a less disarrayed manner. Intentions thus bring more structure and stability to the inputs to the agent's rational executive system.

Another distinctive feature of intention lies in its distinctive satisfaction condition. A desire is satisfied just when the desired states of affairs is realized, even

[12] This should not be taken to suggest that there are no important differences between these volitional states. See Ginet (1990: ch.2), Chang (2013a), and Bratman (1987, 2018) for detailed discussions on the distinctive nature of volition, commitment, and intention respectively.

[13] This idea of "internal command" is also invoked in Locke's characterization of volition, see Locke (1690, Book II, Chapter XXI, Section X). For a similar characterization of this functional role of intention, see Bratman (1987: 15-6).

when its realization is not brought about by the agent herself, but by other people, by sheer luck, or through a deviant causal chain. In contrast, an intention can be satisfied only when the agent brings about the intended state of affairs because of that very intention. Suppose you intend to paint your table white. This intention is not fulfilled if some fairies appear at night and paint the table white for you. Nor is this intention fulfilled if the thought of that intention becomes so overwhelming that it causes you to spill the paint all over the table and to accidentally make it white. Instead, you can fulfill your intention only if you paint the table white by yourself because of that very intention. So intention is a type of desire that has an internal claim to its own efficacy in moving the agent to act and to change the world in accordance with its content. It is directed at the agent and meant to be realized only by the agent. Intention therefore figures as the internal command to the agent's own rational executive system.[14]

This integral role that intention plays in our executive system then explains why most philosophers take the instrumental requirement to be a rational requirement on our intentions or desires with volitional character (e.g. will, commitment) (Bratman 2009a: 16-7; Broome 2013a; Brunero 2012; Korsgaard 1997; Schroeder 2009; Wallace 2001; Way 2010a; Wedgwood 2011). It also explains why the failure in fulfilling one's intention constitutes a more distinctive kind of executive failure (as compared to the failure in fulfilling one's mere hope or wish). Now given the integral role of intention in our agency —that is, given how it settles

[14] This idea of "internal command" is also invoked in Locke's characterization of volition, see Locke (1690, Book II, Chapter XXI, Section X). For a similar characterization of this functional role of intention, see Bratman (1987: 15-6). The idea of intention as an internal command then explains why, as I have claimed in the Introduction, the content of intention as a motivating state can often be reformulated in an infinitive phrase. Suppose you intend *that you paint the table white*. Your intention here, as an internal command to yourself, is already directed to you as actor who is supposed to bring about this prospective state of affairs. That is why it will make sense to leave out the subject/actor in the content of your intention here and to say that you simply intend *to paint the table white*.

our practical question, carries our commitment to a course of action, and figures as an internal command that we give to ourselves—wouldn't intention be a good candidate for a subjective source for normative reason for action? Indeed, this seems to be what many subjectivists have in mind when they propose their accounts of normative reasons for action, for the committed character of intention or volitional state does seem to make a difference to what we have a reason to do as agents.[15]

What is more, our intention, unlike other kinds of desire, is subject to certain internal rational constraints that arise from its role in the proper function of our rational executive system. To begin, the rational requirement of non-contradiction applies more stringently to intention than to desire. If you consciously intend to φ and intend to not φ at the same time, then you would seem to have failed to intend anything in the first place. This is because your contradictory intentions undermine each other's committed volitional character, cancel out each other's force in settling your practical question, and ultimately makes it indeterminate what course of action your rational executive system is supposed to carry out. Also, there seems to be an additional feasibility requirement on our intention: If you intend to do something that is impossible for you to do in the first place, then there would be something defective with your intention. This is because an intention is supposed to provide the standard input for our rational executive system to carry out, and it has an internal claim to its efficacy in moving us to act and to bring about the intended state of affairs. But an intention to do something that is infeasible in the first place

[15] Of course, I am not saying that volitional state alone can generate a normative reason for action. Perhaps it also needs to be maximally influenced by rationality or to further involve a positive evaluative attitude on our part. But the committed character of volitional state does seem to help it account for part of the normative force involved in reason for action. For the further defense for this idea, see Chang (2013), Korsgaard (1996) and Wallace (2004, 2012).

cannot really be fit for our rational executive system to carry out. Therefore, we can say that the proper function of our rational executive system will exclude contradictory and infeasible intentions. They cannot be the appropriate inputs to our rational executive system and therefore cannot boot-strap any normative reasons for us to engage in contradictory or infeasible projects (cf. Chang 2013a: 102, n 46).

What is more, as I will argue in the next chapter, given the settling function and all-out committed character of intention, intention can only be a practical attitude of you as a whole, and not just an attitude of a part of you. So if there is any subjective state that is able to ground a normative reason for action for *you* as an integrated agent, it should minimally include your intention or volitional state in general.

I am not able to give a full defense of subjectivism in this chapter. I will explain later in Chapters 3 and 5 why intentions present the most promising subjective sources of normative reasons for action. But my point here is that the proper function of our rational executive system can already impose some internal constraints on the kind of pro-attitude that it is supposed to carry out and therefore on the kind of desire that can "boot-strap" normative reasons. Indeed, although I have been sticking with an austere Humean picture of agency here in this chapter, the executive conception of IR is actually compatible with other more substantive accounts of rational executive agency. With these accounts at hand, we can even argue that our rational executive ability is supposed to only take certain types of rational volitional states as its inputs. Perhaps they should be volitional states that arise from our critical reflection (Korsgaard 1996), those that are backed by our normative belief or valuing (Wallace 2004, 2012), or those that best represent where we stand as agents (Bratman 2018: ch.4). Thus, I do not think the executive

conception of IR will entail a hugely implausible picture of normativity. And if you do not like any form of normative boot-strapping, there is always a way out—denying the normativity of the executive conception of IR.[16] The boot-strapping worry therefore does not warrant a wholesale rejection of the otherwise plausible executive conception of IR.

6. Conclusion

We can draw several lessons from our discussion here. First, the means-end coherence conception cannot adequately capture the executive dimension of IR. Second, the means-end coherence conception seems plausible only because it invokes the psychological steps constitutively involved in the proper function of our rational executive system. But in some worlds we cannot even find a distinctive coherence pattern to ground IR. Third, IR does not strongly supervene on our intentional state patterns. What constitutes an instrumentally-rational (or means-end coherent) mental pattern can change from world to world. Fourth, the executive conception of IR is supposed to only take volitional states as its typical inputs, which are subject to some internal rational constraints derived from the proper function of our agency. And other more evaluatively-loaded accounts of rational executive agency can impose even more substantive constraints on the volitional states that our rational executive system is supposed to carry out. So the picture of practical normativity entailed by the executive conception is not a wholly absurd one.

[16] But as I have said, I don't think this is a good move. I will explain in Chapter 5 why, and in what sense, the executive conception of IR is normative.

My aim in this chapter was to make a case for the executive conception of IR and thereby pave the way for the views that I shall develop later, namely, a Neo-Humean view about the rationality of intention, a functionalist view about practical rationality and normativity, and a volition-based account of normative reasons for action. In Chapters 2 and 3, I shall proceed from this conception of IR and attempt to show that we can actually derive some other rational principles for our intention from a characterization of the proper function of our executive system. In Chapter 4 I will then provide a more general explanation as to why we should conceptualize practical rationality in terms of the proper function of our agency. I will eventually explain in chapter 5 why our intention is the sort of thing that can best capture the normativity involved in normative reasons for action.

Chapter 2 The Problem of Practical Disunity and the Integration Principle

1. Introduction

It's bad to have conflicting practical attitudes and be conatively disunified. This is something that most people agree on. But what does the badness of conative disunity consist in? One possible answer is that it's bad because it's irrational: Having desires that cannot be fulfilled at the same time amounts to having inconsistent attitudes. Indeed, sometimes our conflicting practical attitudes will even result in direct inconsistency in our beliefs about our future actions (Setiya 2007; Wallace 2001). Another possible answer is that conative disunity is bad because it leads to substantively bad consequences: Our conflicting desires will vie for our general resources for desire-fulfillment (e.g. time, energy, money) and lead us to engage in conflicting or even mutually-undermining projects, and this in turn may very well prevent us from doing something we have reason to do (Kolodny 2008a: 373-7), or at least frustrate some of our desires and cause in us the feeling of disappointment.

Some philosophers, however, have a bolder answer to the question about the badness of conative disunity. According to them, conative disunity is bad not only because it makes us irrational and disposes us to perform some actions that lead to bad consequences; it's bad primarily because it undermines our agential capacity for doing things in the first place. That is, practical disintegration not only makes us irrational or substantively worse off, but also threatens our existence as agents in

the first place.[1] For instance, Michael Bratman has contended that if you intend an end but not the necessary means to it, then your inconsistent practical attitudes will make it indeterminate where you really stand with respect to whether you really want to bring about the end or not, which can in turn undermine your planning agency (2009a: 97-8). Christine Korsgaard has also argued that if you do not act on some general, universalizable principles that decide which conflicting desires you should endorse and that represent you as occupying a deliberative space over and above these conflicting desires, then there isn't really a *you* that is distinct from that conflicting desires within you in the production of action. So the resulting action will only be attributable to the conflicting forces within your psychology, but not to your agency as a whole (2009: ch.4).

However, the claim that conative integration is indispensable for agency might seem like an exaggeration to many. Indeed, as those who embrace a more austere Humean picture of agency might think, our agency simply consists in our rational executive capacity for fulfilling our intentions. So a certain bodily movement of us can be seen as the product of our agency just when it is actively taken to fulfill an intention that we have, even if we might still be disintegrated in our underlying desires, and even if we might fail to act from any universalizable principles of action. Sartre's pupil, for instance, might be torn between his desires to join the Free French Forces and to stay with his mother. But if he can form an intention, not on the basis of some general principles of action, but through a sheer act of the will and then set out to do something to fulfill it, then perhaps he can still be said to be the author of his action. In fact, one might even think that his conative disunity actually makes

[1] Note that these three views about the badness of conative disunity are not mutually exclusive. For instance, you might think the irrationality involved in practical disunity will lead to substantively bad results for us, or that the loss of agency that practical disunity leads to is something substantively bad.

him more of an author of his action, for his action is determined by an act of his own will, and not just by the overall balance of the conflicting forces within his psychology (cf. Chang 2009). So, as you might think, perhaps the claim that conative disunity threatens our agency can only hold on more substantive conceptions of agency like the ones adopted by Bratman and Korsgaard.

This way of thinking, however, is wrong. As I shall argue in this chapter, practical disunity does undermine our agency, even when agency is understood minimally as our rational executive capacity for fulfilling our intentions. In fact, it does so precisely by way of posing an obstacle for us in forming an intention and in having a will in the first place. Suppose Sartre's pupil still remains torn and cries out, "I intend to leave home tonight to join the Free French Forces. But you know what? I also intend not to leave home tonight!" Presumably, his disunity, as expressed in his claim here, would lead us to think that he hasn't really succeeded in forming a serious intention to do anything tonight. And we would tend to respond to his disunity by saying to him: "Then what do you really intend to do tonight? Make up your mind now!" The underlying reason why his disunity would prevent him from forming a serious intention and having a distinct will, as I shall argue, lies in the very functional nature of our volitional state. Our intention, will, or volitional state functions to settle our practical question concerning what we are to do with its all-out, committed volitional character. Yet a disunified mind cannot normally succeed in forming a mental state with such a character. So conative disunity undercuts our agency, according to *the functional account* that I shall defend, just because the very defining function of intention presupposes psychological integration.

This functional account, I contend, offers a less controversial story about the

indispensability of psychological integration for agency, one that can be accepted even by proponents of the austere Humean picture of agency. It also provides a more illuminating explanation about the functional and volitional character of intention, which is the most central kind of practical state in the general functionalist picture of practical rationality and normativity that I will develop in this thesis. More importantly, and perhaps somewhat surprisingly, it will show that our intention is constitutively governed by the principle of continence. This is because following this principle, as I shall also argue in this chapter, is actually constitutive of reaching the kind of psychological integration presupposed by intention. So the functional account will also reveal how our intentions are not just governed by mere instrumental and formal coherence principles. It will thereby lead us to make the first step toward a neo-Humean view about the rationality of intention.

This more general constitutivist story actually has its origin in Donald Davidson. In his later writings, Davidson suggests that following the principle of continence is constitutive of being an agent with intentions (1985a). This principle, according to him, prescribes us to form an all-things-considered normative judgment and then intend in accordance with it. Unfortunately, he fails to offer any explanation for his bold claim about the constitutive relation between forming an intention and following this principle, and it has largely gone unnoticed in the literature. But in my view Davidson actually has already provided all the required materials for making sense of his claim about the constitutive relation between the principle of continence and intention. It's just that he fails to see the crucial missing link here—psychological integration.

Our story will therefore start from Davidson's constitutivist claim, which I will outline in Section 2. We will then come up with the crucial missing link to make

sense of his claim in Section 3 and 4. I will explain first how the functional nature of intention presupposes a fair degree of psychological integration, and then how this kind of integration is precisely the type that the principle of continence makes for. Section 5 examines intentions formed under conditions of psychological disunity and explains why these intentions are defective in their central function. Section 6 concludes by highlighting how the constitutivist story points out that there is a rational *integration principle* governing our intention.

2. The Constitutive Status of the Principle of Continence

2.1. Davidson on Practical Rationality

Let us begin here with Davidson's view on practical rationality and the origin of the principle of continence, since it will later on provide the necessary materials for developing our functional account. The picture of practical rationality that underlies his early paper "Actions, Reasons, and Causes" (1963) is undoubtedly an instrumentalist one. As he argues there, a bodily movement φ-ing is an intentional action in virtue of its being both caused and rationalized by a suitably connected belief-desire pair, namely, a desire to perform a certain type of action (say, a desire to perform actions with the property of bringing pleasure) and a belief that φ-ing is a token of the desired action type in question (say, a belief that having one more piece of cake is an action with the property of bringing pleasure) (687). As we can see from how Davidson formulates the content of the rationalizing belief-desire pair, for him an intentional action is rendered rational primarily in an instrumental sense—it is rendered rational in virtue of being shown to be a product of the agent's

instrumental rationality in fulfilling her desire to perform a particular type of action.

Given this, every intentional action is rational in a broadly instrumentalist sense on Davidson's account of intentional action. That is why he claims that there is a rational aspect in all intentional actions (1963: 9; 1982: 169-70). Davidson, however, does not think that instrumental rationality is all that there is to practical rationality. Rather, he believes that there is a further sense in which we can judge an intentional action to be practically irrational. It is when an intentional action is *incontinent*, i.e. when it is done against one's evaluative judgment concerning what action would be best for one to do, all things considered. Incontinent action is irrational, Davidson contends, mainly because our all-things-considered judgment (henceforth: ATC judgment) has a rationally privileged status in our mental economy, such that any intention or action contrary to it can be judged to be incontinent and practically irrational.

But why is our ATC judgment rationally privileged? According to Davidson, this is because there is a rational *principle of continence* that prescribes us to intend and perform the action "judged best on the basis of all available relevant reasons (1969: 41)." This principle confers a rationally privileged status to our ATC judgment and renders the intentions and actions contrary to it practically irrational. So what underlies Davidson's treatment of incontinence is a view of practical rationality that clearly moves beyond the instrumentalist one in his earlier work. But how can the principle of continence have a special place in our picture of practical rationality in the first place?

In "Paradoxes of Irrationality" (1982), Davidson answers this question by appealing to the agent's attitude toward this principle itself. As he argues,

> A purely formal description of what is irrational in an akratic act is, then, that the agent goes against his own second-order principle that he ought to act on what he holds to be best, everything considered. It is only when we can describe his action in just this way that there is a puzzle about explaining it. If the agent does not have the principle that he ought to act on what he holds to be best, everything considered, then though his action may be irrational from our point of view, it need not be irrational from his point of view (178).

According to Davidson, it is in virtue of our accepting the principle of continence as our second-order principle that it is rational for us to form intentions in conformity with our ATC judgment. In a way, it is because of our endorsement that the demand of the principle of continence is a demand of rationality for us.

Now what implicitly underlies Davidson's view here is a version of a radical internalist view about rationality, according to which a certain way of forming intentional states is rational (say, forming intentions in conformity with one's ATC judgement) just in virtue of how the agent has a second-order principle about how she ought to form her intentional states in this way. This radical internalist view, however, is not very plausible. This is because it is overly permissive and gives in too much to the agent's own subjective perspective. It implies that almost every irrational way of forming intentional states can be potentially exonerated as a rational one, provided that the agent just has a crazy second-order principle about how she ought to form intentional states in this way. Indeed, as we saw in the foregoing chapter, this view entails a form of anarchism about rationality, for on this view each agent can be rational simply by conforming to his or her own idiosyncratic second-order principle.

In fact, Davidson later on changes his mind about the origin of the principle of continence in his paper "Incoherence and Irrationality" (1985a). There he argues, correctly I think, that the demand of the principle of continence is the demand of rationality for us not because of our acceptance of it, but because of its constitutive status:

> Rationality […] is a condition of having thoughts at all. The question whether a creature "subscribes" to the principle of continence, or to the logic of the sentential calculus, or to the principle of total evidence for inductive reasoning, is not an empirical question. For it is only by interpreting a creature as largely in accord with these principles that we can intelligibly attribute propositional attitudes to it, or that we can raise the question whether it is in some respect irrational. We see then that my word "subscribe" is misleading. Agents can't decide whether or not to accept the fundamental attributes of rationality: if they are in a position to decide anything, they have those attributes (196-7).

As long as we have any intentional states at all, we must be *ipso facto* following basic rational constitutive principles and can be seen as accepting them in our practice. And these principles cannot be subject to our acceptance and rejection at all. In so far as we are in a position to accept or reject them, we must have already been following them and accepting them in our practice in the first place. So the principle of continence has a place in practical rationality not because we accept it as our second-order principle, but rather because following it is constitutive of having intentional states.

Thus, Davidson ends up employing a constitutivist strategy to show how

principles of rationality simply apply to us in virtue of their constitutive status, regardless of whether we accept them or not. This then qualifies him as one of the earliest precursors of constitutivism. What is interesting here, however, is that Davidson also takes the principle of continence to be one of rational constitutive principles. This is a very interesting and daring claim. How can following this principle be constitutive of being an agent with intentional states? What type of intentional state is constitutively linked to it?

2.2. Toward a Constitutivist Account of Intention

The most promising candidate here is intention. Recall that on Davidson's view, what the principle of continence requires us to do is to form our *intentions* in accordance with our ATC judgment. So if we take Davidson's constitutivist suggestion seriously, then a natural claim to make on his behalf is that following the principle of continence is constitutive of forming an intention. In fact, Davidson gives the following description about how different types of rational principles are constitutive of different kinds of mental items:

> I would say the same about the basic principles of logic, the principle of total evidence for inductive reasoning, or the analogous principle of continence. These are principles shared by all creatures that have propositional attitudes or act intentionally; and since I am (I hope) one of those creatures, I can put it this way: all thinking creatures subscribe to my basic standards or norms of rationality. This sounds sweeping, even authoritarian, but it comes to no more than this, that it is a condition of having thoughts, judgments, and intentions

that the basic standards of rationality have application (1985a: 195).

As the parallel structure here suggests, having thoughts is conditioned on following basic principles of logic; making judgments on following the principle of total evidence; and forming intentions on following the principle of continence. This then strengthens my conjecture that Davidson takes the principle of continence to bear a constitutive relation to intention. But why is this so? Unfortunately, Davidson does not give any explanation for this as far as I am aware.

It might be suggested that to form an intention is just to settle the practical question of what to do. But in our practical deliberation, the practical question of "what to do" figures in the form of the question of "what would be the best to do (Raz 1998: 52-3)". So if an agent is to form an intention to settle her practical question, then she must *ipso facto* follow the principle of continence and consider what action would be the best for her to do.[2] In fact, according to one of the most popular constitutivist account of belief, our belief formation constitutively involves following the norm of truth just because in our doxastic deliberation, the doxastic question of "what to believe" figures in the form of the question of "what is true" (Shah 2003; Velleman 2000; Wedgwood 2002). So one might think that a plausible constitutivist account of intention would also share a similar structure. Perhaps given the way that our doxastic and practical question figures in our first-person deliberative perspective, belief has to aim at what we take to be true, while intention at what we take to be the best. That is why intention formation constitutively involves following the principle of continence.

This guise-of-the-good account, however, fails to explain why there is a

[2] This view has been developed by Alfred Mele (1992: ch.12).

constitutive relation between intention formation and the principle of continence as it is formulated in Davidson's own terms. On Davidson's view, what the principle of continence prescribes us to do is not to just intend in accordance with our evaluative judgment, but to intend in accordance with our ATC evaluative judgment concerning what action is the best *in light of all relevant considerations* that we have brought together. And the guise-of-the-good account doesn't explain why intention formation also constitutively involves such an all-things-considered process of "bringing all relevant considerations together", a process that is distinctively incorporated into Davidson's conception of continence.

Indeed, according to Davidson, our intention to φ still reflects our *all-out* evaluative judgment that φ-ing is the best *tout court*, which has an unconditional logical form and expresses a conclusive attitude that leads to action (this action is just the best!). On Davidson's view, it is possible for us to intend incontinently only because our intention and all-out evaluative judgment can diverge from our ATC judgment, which instead has a conditional logical form (this action is the best, given how all the relevant considerations, taken together, speak in favor of it) (1969: 39-40). Yet the guise-of-the-good account only brings back the problem of the impossibility of incontinence that Davidson actively seeks to avoid: If intention formation is constitutively tied to an evaluative judgment concerning what would be best for us to do, then how could it be possible for us to intend and act incontinently?

There is, however, another account that we can propose on Davidson's behalf to explain the constitutive relation between intention formation and the principle of continence, an account that reveals how (1) intention formation presupposes a fair degree of psychological integration, and how (2) such psychological integration is

precisely made possible by the all-things-considered process involved Davidson's principle of continence. I will spell out these two ideas in the next two sections, which will then provide the substance for our constitutivist story. (To avoid confusion, in what follows I will use the term "the C-principle" to refer to the principle of continence as it is formulated in Davidson's terms, while the term "the principle of Enkrasia" will be used to denote the continence principle as it is usually understood.)

3. Intention and Integration

3.1. The Incompatibility of Intention and Mental Partitioning

In order to see how the formation of intention presupposes psychological integration, let us start with an idea that is antithetical to mental unity, namely, mental partitioning. As we have seen earlier, Davidson believes that following basic rational principles is constitutive of being an agent with intentional states. So, like contemporary constitutivists, he is faced with the problem of explaining how we can have irrational, and in particular, inconsistent intentional states at all. Davidson's famous solution to this, as proposed in his "Paradoxes of Irrationality" (1982), is that the inconsistency in our intentional states can be accommodated through a kind of mental partitioning: If two inconsistent attitudes are located in semi-autonomous partitions of the mind, then the inconsistency between them can be mitigated by the isolation between the mental compartments they are in, and they can still retain each of their rational connections with other intentional states in the same compartment.

What is interesting is that Davidson's proposal about mental partitioning seems to apply more readily to desire than to intention. In the case of desire, we can easily imagine how you can consciously have synchronic but contradictory desires. This naturally leads us to think of these desires as residing in different parts of your mind. Indeed, this is how moral philosophers have conceptualized the conflict of desires throughout history. That is why we have psychological partitionings in moral psychology like boulesis versus appetite, reason versus passion, or will versus inclination, etc.

The same thing, however, does not seem to hold for intention or volitional state. Of course you can have two *conflicting* intentions at the same time, say, the intention to lose weight and the intention to have one more piece of cake. This is because one of your intentions here might only be conditional (perhaps you intend to avoid high calories food only under most, but not all, circumstances). Or sometimes you might not realize that there is a conflict between your intentions (perhaps you fail to know how this tiny piece of cake is actually quite high in calories. Or maybe you just forget your intention to lose weight).

But let us bracket these cases of seeming conflict and consider the more extreme case of *conscious*, *synchronic, and direct contradiction*: It is quite easy to see how you can consciously desire to φ and desire to not φ at the same time. The statement that "I desire to go to Mark's party tonight but I also desire to not go to Mark's party tonight" makes perfect sense—you can desire and resist the same course of action for different considerations. But it seems a lot more difficult to conceive how you can consciously intend to φ and intend to not φ at the same time, where the act φ-ing here is intended and resisted under the same description. The statement that "I intend to go to Mark's party tonight but I also intend to not go to

Mark's party tonight" seems to involve a more serious kind of tension that is not present in our preceding statement about contradictory desires. It does not seem to make very good sense, for in intending and resisting the very same course of action under the same description, you seem to have failed to intend anything in the first place. That is why we would feel the need to avoid the conscious, synchronic, and direct contradiction here by asking, "So what do you really intend to do?" It is only after the contradiction here is resolved that we would feel confident in saying that you have really intended to do something.

This difficulty in attributing two conscious, synchronic, yet contradictory intentions is not explained by the fact that forming an intention is a mental act, and that we cannot perform two mental acts at the same time. Even if we concede that your contradictory intentions can only be formed at two different points in time (say, the intention to φ at T1 and the intention to not φ at T2), it still seems a bit difficult to see how you can consciously *hold* or *retain* these two directly contradictory intentions at the same time. The statement that "I have both the intention to go to Mark's party tonight and the intention to not go to Mark's party tonight" still creates an air of confusion in a way that our statement about contradictory desires doesn't.[3] These two contradictory intentions still seem to cancel each other out, and we still feel the need to settle the inconsistency in your attitudes by asking, "So which one is your real intention?"

I think our difficulty in attributing two conscious, synchronic, yet contradictory intentions is rather explained by the fact that we cannot accommodate the

[3] Of course, if we view our contradictory intentions from a third-person point of view, then the tension involved in this statement can be somewhat mitigated. This is because in adopting a third-person point of view, we are also detaching ourselves from the inconsistent commitments involved in our contradictory intentions. But what I want to focus on here is the kind of statement that *expresses* (and not just *quotes*) our contradictory intentions and the inconsistent commitments involved in them.

contradiction here by confining these two intentions to different mental subdivisions in the way suggested by Davidson. Indeed, the idea of "partitioned intentions" seems a bit strange in the first place, for it sounds rather unnatural to say, "One part of me intends to go to Mark's party tonight, while another part intends to not go to Mark's party tonight." This statement about partitioned intentions sounds peculiar not because it involves the presupposition of some "homunculi" or small agents in our mind. The same presupposition can also be said to be invoked in the statement "One part of me desires to φ but another part desires to not φ." But this statement about how different parts of you desire different things sounds a lot more natural than the previous statement about partitioned intentions. Neither does the unnaturalness of the idea of partitioned intentions result from the fact that intention formation is a more sophisticated mental process that incorporates the functions of different modules of our mind. Two contradictory desires can also be formed through rather sophisticated mental processes that incorporate the workings of, say, our visual, linguistic, and reasoning capacities. But once they are formed, we would usually have no problem in conceptually locating them in two different subdivisions of the mind. The same thing, however, does not seem to hold for intention.

The unnaturalness of idea of partitioned intentions, I believe, rather lies in the fact that mental partitioning is simply incompatible with the specific function that intention has. Let me explain this.

3.2. Two Defining Features of Intention

One of the defining features of intention or volitional state is its *settling function*: The formation of your intention signifies that you have at least settled the practical

question of what you are about to do by reaching a verdict in favor of a particular course of action. What is left for you to do is to set out to execute this verdict. An intention can play this settling function in turn because it possesses an *all-out* or *committed* volitional quality: The formation of your intention signifies that you have taken a stand in the face of the practical question by committing yourself to a particular course of action, as opposed to remaining ambivalent between various courses of action that you can potentially perform. Notice that the settling function and the all-out volitional quality that I am canvassing here are the properties that an intention has *at the moment of intention formation*. It is still perfectly possible for an intention to lose its settling power and its all-out volitional character later on. But it seems that at least at the moment of intention formation, your intention has to carry your commitment to a particular course of action and thereby settle the practical question for you. It is this kind of *synchronic* settling force and all-out volitional quality, I believe, that marks the difference between intention and other kinds of pro-attitude.

The settling function and the all-out volitional character of intention have been featured in many philosophers' characterization of intention. So some of them have focused on the aspect of settling function. Bratman, for instance, characterizes intention as a kind of pro-attitude that exerts more direct control over our future actions, that provides the framework for our future practical deliberation, and that disposes us not to redeliberate on the practical question (1987: 16-7). And it is the settling function of intention that explains why it can play the roles that Bratman describes here. Indeed, Alfred Mele even explicitly claims that "intending entails being settled, or at least being thinly settled (1992: 169)." Other philosophers, on the other hand, have paid more attention to the volitional character of intention.

Scanlon suggests that our intention involves our commitment to take certain considerations (say, considerations related to its fulfillment) as reasons in one's deliberation (2007: 92-3). Also, Davidson, as we have just seen, takes intention to express our all-out evaluative judgment, which is unconditional in its logical form. Of course Davidson's identification of intention with all-out evaluative judgment is controversial, but it seems to me that he at least correctly notes how our intention has a distinctive committed volitional character, which makes it bear a closer resemblance to our all-out, unconditional evaluative attitude (1969: 40; 1978: 97-8). So I think the two features of intention highlighted here are perhaps the most basic defining features of intention, features that explain other derived properties of intention that have been pointed out in the literature. In fact, I would say that the all-out volitional quality of intention is more fundamental than its settling function, for it is precisely in virtue of its volitional all-out-ness that intention is able to settle the practical question for us.[4]

In any case, it is these two defining features of intention, I believe, that explain why the idea of partitioned intentions is unnatural. If an alleged intention is only an attitude of a part of you, then it cannot normally possess the required all-out volitional quality and play its defining settling function well in light of the potentially conflicting attitudes in other parts of you. Suppose you say, "One part of me intends to go to Mark's party tonight but another intends to not go to Mark's party tonight." Our natural response to this statement about your partitioned intentions is to say, "Then you haven't really committed to anything and settled the

[4] Although Harry Frankfurt does not explicitly talk about the intention, he correctly points out that a state with wholehearted volitional character can have a special place in our agency in virtue of how its volitional quality enables it to settle our practical question, for "making a commitment without reservation means that the person who makes it does so in the belief that no further accurate inquiry would require him to change his mind (1987: 168-9)."

matter concerning whether you are to go to Mark's party tonight. What do you really intend to do? Make up your mind now!" The underlying explanation for our natural response here is that given the other conflicting intention in another part of you, none of your intentions here can really carry your all-out commitment and settle your practical question. Indeed, it is only after you bring together these conflicting attitudes within you, settle the matter again, and form an attitude with genuine all-out commitment to a particular course of action that a final verdict is reached and a genuine intention is formed.

In other words, the functional nature of intention or volitional state sets a constraint on its metaphysical constitution: Given its defining settling function and all-out volitional character, intention can normally only be a practical attitude of you *as a whole*, and not just a mere attitude of a part of you—for a partial attitude cannot normally have the volitional all-out-ness and settling force required for intention. That is why intention formation has to presuppose a fair degree of psychological integration. Thus, it is indeed the case that psychological disunity undermines our agency. This seemingly controversial and substantive idea actually holds true even on an austere Humean characterization of agency, for psychic disunity simply threatens our capacity for forming an intention with the required committed volitional quality and settling force in the first place.

This then completes my functional account about the indispensability of psychic unity for agency. Of course, I am not saying that it is impossible for us to have partitioned intentions or to form intentions under mental disunity. It's just that these intentions will normally be defective in their volitional character and setting function. If they are defective enough, then it would sometimes be hard to say whether they represent genuine intentions or not. I will deal with these defective

intentions later on in Section 5. But here I shall continue with my general constitutivist story and explain why the kind of psychological integration required by intention is precisely the sort made possible by the C-principle.

4. Integration and Continence

4.1. All-Things-Considered Deliberation and Integration

So how can one form an intention *as an attitude of a whole* in light of the various considerations for and against different courses of action? The answer, which sounds trivial but perhaps formally true of every process of intention formation, is just to bring together all the relevant considerations that one has in mind, and then to reach a single, overall practical attitude based on them. But this kind of integration process is just what Davidson's C-principle, with its distinctive all-things-considered aspect, prescribes us to do: It first prescribes us to form a unified practical attitude—as represented by our ATC judgment—in light of the various considerations pulling us to different directions; it then asks us to form an intention on the basis of the psychological unity that we have reached through the formation of the ATC attitude. That is why following the C-principle can be said to be constitutively involved in forming an intention.

Indeed, recall that Davidson suggests that forming a judgment is conditioned on following the principle of total evidence, while forming an intention on following the C-principle (1985a:195). The general picture behind this claim, I think, is that our all-out attitudes like judgment and intention typically need to be formed through an integrative ATC process, which takes all the relevant evidence or desirability

considerations as its input, and which gives rise to single attitude as its output. In a way the C-principle is the practical analogue of the principle of total evidence for Davidson. They are meant to funnel the relevant evidence or desirability considerations that we have in mind into a single attitude with an all-out character.

Two things are worth noting. First, on Davidson's view, each desirability consideration is underwritten by a corresponding desire. This is because he believes that the logical form of the expression of desire is an evaluative judgment about how a certain type of action is desirable in so far as it has a particular feature F (say, the property of bringing pleasure), a feature that figures as the desirability consideration in favor of that action type (1985b: 202). Conversely, he also contends that our evaluative judgment about the desirability of different courses of action comes down to a kind of conative attitude or pro-attitude (1984: 20). These claims about the intimate connection between evaluative judgment and desire suggest that, on Davidson's account, each desirability consideration is underpinned by a corresponding desire or motivating state. So on his view, the very process of putting together all desirability considerations and then forming a single ATC judgment is also a process of integrating the desires that underwrite these considerations into a single practical attitude of us as a whole.

Second, Davidson's C-principle has a formal character. All that it requires of us is to form a unified ATC attitude by "putting together" all the relevant desirability considerations. But it does not really prescribe us to put them together in any particular way. In fact, we can accord different kinds of considerations different weights, or even exclude some kinds of considerations from our deliberation

altogether by certain general principles that we endorse.[5] All these different ways of deliberating are compatible with following the C-Principle, for it only formally prescribes us to engage in an integration process of reaching an ATC attitude. This formal character, I believe, is what makes the constitutivist view reconstructed here more plausible than its Kantian variant (Korsgaard 2009). This is because it does not have to appeal to other more specific rational principles that we should follow (e.g. the categorical imperative) in order to reach integration. All that it invokes is a general formal rational principle that requires us to reach integration, in whatever ways we like.[6]

I think this proposal about how the C-principle is meant to bring about mental integration is already implicitly hinted by a metaphor that Davidson offers. Davidson suggests that we should think of weakness of the will in terms of the "judicial" instead of the "gladiatorial" model: The will is not one of the conflicting forces within one's mind. Rather, it is the "agent's representative" that "judge[s] the strength of the arguments on both sides (1969: 37)", adds them up, and reaches a single verdict. Now the formation of *a* will or *an* intention to do something signifies that one has settled one's practical question and reached volitional all-out-ness. As the judicial model here suggests, the settlement here is reached not through the clash of the desires and motivating forces within our mind, but rather through an integrative process of bringing together all the relevant considerations. Indeed, in one of his papers, Davidson even explicitly says that "a judgement that, all things

[5] For a more detailed account of how our practical deliberation also relies on general principles for excluding certain kinds of considerations or desires in an outright manner, see (Korsgaard 2009: 49-52; Scanlon 1998: 52-5).

[6] That being said, my constitutivist story here is still compatible with the Kantian view. It's just that Kantians appeal to rational principles that prescribe more specific ways of reaching integration, and these principles will be undoubtedly more controversial than the formal C-Principle. For this kind of Kantian view, see Korsgaard (2009: ch.4).

considered, one ought to act in a certain way presupposes that the competing factors have been *brought within the same division of the mind* (my emphasis, 1982: 182)". Such a single, unified division of mind here, I believe, is just what Davidson's C-principle makes for and what the formation of intention presupposes.

4.2. Two Kinds of Mental Integration

One might doubt whether this constitutivist story can really work. After all, although intention formation requires a fair degree of mental integration, forming an ATC judgment and then intending in accordance with it doesn't always seem to be the best way for us to reach integration. In fact, Nomy Arpaly has provided a good case in support of this view (2002: ch.2). Given that some of our desires (and the desirability considerations that they support) might fail to explicitly figure in our first-person deliberative perspective, our ATC judgment can sometimes come out of touch with our underlying desires. When this happens, Arpaly argues, it would actually be more rational for us to intend and act *against* our ATC judgment, for this will let us better cohere with our underlying desires and better follow our desire-based reasons. Suppose I have a mild desire to visit Bangkok but a strong dislike for hot weather and a strong dislike for humidity. But suppose I am simply unaware of these strong dislikes because I have been living in a country with dry, temperate weather for a very long time. So I simply failed to take these underlying desires into account when I formed an ATC judgment in favor of going to Bangkok for holiday. Now despite my ATC judgment, I might still have a gut reaction to cancel my plan whenever I see the pictures of Bangkok's splendid golden temples under bright sun, even though I cannot formulate a good, explicit reason in support of this

gut reaction. According to Arpaly, in this case it would be actually more rational for me to intend against my ATC judgment and follow my gut reaction, for this can in fact allow my intention to better cohere with my underlying desires, the overall balance of which disfavors my going to Bangkok. So you might follow Arpaly's suggestion and doubt whether following the C-principle can always contribute to psychological integration, since, as our case here suggests, sometimes forming intentions in accordance with our ATC judgment can in fact make us more disunited.

I think Arpaly is right in saying that sometimes following our ATC judgment can make us worse off with respect to our integration with our underlying desires. And I agree that sometimes it is indeed irrational for us to follow our ATC judgment—irrational in the sense of failing to cohere with our underlying desires and conform to our desire-based reasons. All these points are correct, but they do not contradict our constitutivist story here, for Arpaly and I simply have two different kinds of psychological integration and rational problem in mind. What she talks about is the unity between our intention and our underlying desires, while what I have been focusing on is the unity in our first-person deliberative perspective. The former kind of unity consists in intending in a way that coheres well with the overall balance of our underlying desires/motivating forces, while the latter kind of unity resides in reaching an all-out commitment to remove the ambivalence in our first-person deliberative perspective about what courses of action we are to take.

The first thing to notice is that the kind of integration that intention presupposes is the latter kind of integration, the integration in our first-person deliberative perspective. The support for this can be found in Mele's insightful suggestion about how intention cannot be identified with preponderant motivation (1992: 154-71). As he points out, even if the balance of one's underlying desires

preponderantly motivates one to φ, this still does not mean one has an intention to φ: Given my underlying desires, I might be preponderantly motivated to insult my friend and believe that I will do so in the future. Despite this, I can still regard the practical question concerning what I am about to do as unsettled from my first-person deliberative perspective, and I can still try my best to prevent myself from insulting my friend. In this case, it would be counterintuitive to say that I actually have an intention to insult my friend. Conversely, one might intend to φ but lacks overall predominant motivation to φ in the first place: An addict might form an intention to quit and regard the practical question as settled from his first-person deliberative perspective, even if, as a matter of fact, he is not really able to bring himself to do so given the overall balance of his underlying desires.[7] As this break between intention and overall preponderant motivation shows, the type of mental integration that intention formation presupposes is the integration in our first-person deliberative perspective, and not the integration between our intention and our underlying desires/motivating forces.

The next thing to notice is that the former kind of first-person integration is exactly what the C-principle makes for. After all, as we have seen, to follow this principle, we need to first engage in an all-things-considered deliberative process and then form an ATC judgment. But to engage in this kind of deliberative process is just to consider, from our first-person perspective, what action is the best in light of the various desirability considerations pulling us to different directions. And if we successfully reach a single ATC attitude in favor of a particular action in light of

[7] In fact, given that addiction works primarily by over-sensitizing one's dopamine system (the system regulating one's motivation) to the addictive substance, it's plausible to think that the addict can still have overall preponderant motivation to continue to get the addictive substance when he forms the intention to quit (Robinson and Berridge 2003).

these conflicting considerations, then it will indeed signify that a fair degree of integration has been achieved at least in our first-person deliberative perspective in the face of the practical question. If you cannot even decide which course of action is the best in light of the conflicting considerations that you have covered, then you would indeed be disunified in your first-person deliberative perspective concerning what you are to do. That is why forming an ATC judgment and then following it, I think, is still one of the most general and formal ways for reaching the kind of psychological integration required by our intention.[8]

Therefore, Davidson's seemingly controversial claim actually makes good sense once we see how following the C-principle prepares for the kind of mental unity required by intention. This then completes our constitutivist story about the relation between intention, integration, and the C-principle. Of course, this story here does not entail that we can never form intentions without following the C-principle and without reaching integration. Rather, what it suggests is only that intentions formed in this way will be defective. So I will try to strengthen this story here by moving to the other end of the spectrum and considering defective intentions formed without proper integration.

5. Defective Intentions

[8] What if you form an ATC judgment not through an all-things-considered deliberative process, but on the basis of the testimony of other people? Wouldn't it be counter-intuitive to say that this kind of ATC judgment can also contribute to your mental integration and hence intention formation? I would tend to say that in this kind of case, what you have is not a genuine ATC judgment, for you haven't really engaged in the all-things-considered deliberative process in the first place. In other words, what makes an ATC judgment an all-things-*considered* judgment is not its content (for one can capriciously form a judgment with the content that something is all-things-considered the best, even if one hasn't actually considered anything), but the distinctive type of deliberative process that gives rise to it.

5.1. Intentions without ATC Deliberation

As I have just said, the C-principle requires us to cover all the relevant considerations, form an ATC judgment based on them, and then intend accordingly. So there are three ways in which we might fail to follow this principle. The first is when we fail to cover all relevant considerations in our deliberation in the first place; the second is when we do bring together all relevant considerations but then fail to reach a single ATC judgment; the third is when we do form an ATC judgment but then fail to intend in accordance with it. So let us consider why these three ways of forming intention will result in defective intentions.

To begin, we can often form an intention without actively seeking and covering all relevant desirability considerations. Perhaps most of the considerations presently available in our deliberative perspective all point to the same course of action; perhaps some of these considerations are already quite salient and weighty. The concurrence of the presently available considerations or the salience of some of them might then directly lead us to commit ourselves to a particular action and to regard the practical question as settled by these considerations alone. That is why we often seem capable of forming an intention in a quick, non-all-things-considered way without an ATC judgment. I would even say that this is the usual way for us to form an immediate, present-directed intention to perform a basic action.

But I think intentions formed in this way would still have a potential functional defect. After all, if you haven't taken into account some other desirability considerations when you form an intention, then its all-out volitional quality and settling power will often be *fragile*. This is because some of the concurrent considerations that were available but not taken into account at the very moment of

intention formation might still be able to break your integration and overturn your intention later on. For instance, you might first intend to make a mean joke about your friend, since your desire to make a witty yet mean remark might present the most salient consideration for you at the moment. But when words come out of your mouth, a more important consideration that was already there when you first formed the intention and that you failed to take into account comes to your mind. You realize that you actually have a stronger desire not to hurt your friend's feelings. So you change your mind even before you finish the joke, leaving everyone utterly confused. As this case suggests, if your intention is only based on a kind of integration reached in a quick and dirty way by covering only some of the relevant considerations, then it can only have a fragile all-out volitional character and settling force that can be easily overturned.

Notice that here I am focusing on the defect in an intention's *synchronic* volitional quality and settling force. My claim here is that if you form an intention only on the basis of some desirability considerations, then its volitional all-out-ness and settling force will be fragile just *at the very moment of its formation*, given other concurrent desires (and the desirability considerations they support) that you failed to cover when you formed your intention. It is different from cases where your intention loses its volitional all-out-ness and settling force because you acquire other new desires later on. The judicial picture that Davidson appeals to in fact provides a good analogy here. If the will as the judge does not take all the presently available evidence into consideration, then its verdict is more likely to be overturned, and it will hence have a less safe settling power just at the very moment when the verdict is made.

This case also suggests that the two kinds of mental integration that Arpaly and

I have highlighted are not wholly separated. If you do not intend in accordance with the overall balance of your underlying desires, then the integration that you reached in your first-person deliberative perspective will often be fragile as well. After all, as soon as some of these underlying desires (e.g. your desire to not hurt your friend's feelings) surface in your first-person deliberative perspective and make you aware of the conflicting desirability considerations that they support and that you failed to cover when you first formed your intention, you might very well be swayed and become unsettled again in your first-person deliberative perspective.

5.2. Intentions without ATC Judgment

Now come to the second way in which we might fail to follow the C-principle. Sometimes we might actively bring together all relevant considerations but then fail to reach an ATC judgment about which action is the best, all-things-considered. This might happen because the competing options are qualitatively identical and are therefore supported by the exact same set of considerations. For instance, standing in front of the supermarket shelf filled with Doritos, I might be unable to reach an ATC judgment concerning which exact pack of Doritos is the best all-things-considered. Or this might happen because the competing options are supported by comparable but equally weighty considerations, or by incomparable considerations that cannot be weighed against one another in the first place.[9] For example, after a comprehensive survey of all relevant considerations, I might still be unable to reach an ATC judgment about whether it would be all-things-considered

[9] Another possibility is that the competing options are supported by considerations that are comparable but put these options "on a par", see Chang (2002).

best for me to go to Bangkok or the Grand Canyon for holiday. Bangkok has an exciting food culture, while the Grand Canyon boasts scenic natural landscapes. For me these two considerations are just equally weighty or perhaps even incomparable in the first place. Still, in a scenario like this, we seem perfectly able to form an intention in the absence of an ATC judgment. So how can our constitutivist story explain this kind of intention?

Intentions formed in the way are actually still often backed by a tacit ATC judgment and a fair degree of mental integration, so they do not constitute counterexamples to our story here. Start with cases where two or more qualitatively identical options are supported by the very same set of considerations. In this kind of case, we can still be said to have an ATC judgment in favor of an action *type* (e.g. taking a pack of Doritos) and is still fairly integrated with respect to the question concerning what we are about to do. It's just that the course of action that we commit ourselves to and settle on is only an action *type*, and we have the executive discretion in determining how we shall carry out that intention at the moment of action by a kind of improvisation (e.g. by taking this or that pack of Doritos).

Next consider cases where two or more options are supported by equally weighty or incomparable considerations. Similarly, although we cannot form a strict ATC judgment about which option is the best, in this kind of case we often have a tacit ATC judgment that what is best for us to do, all-thing-considered, is to *choose* one option instead of wavering between various options and being torn with indecision like Buridan's ass. It is this kind of background ATC judgment that supports the exercise of our will to break the tie and to resolve the quandary of

evaluative equality or incomparability (see Chang 2009).[10] In fact, we would indeed have difficulty in forming an intention without such a background ATC judgment. Suppose you are in a tragic case where you are faced with two options that are bad for equally weighty or incomparable considerations, say, an autocrat forces you to sacrifice either your son or your daughter (and let's suppose that you don't have a favorite). Would you have the ATC judgment that what is best for you to do now, given your situation, is to *make a choice* to sacrifice one of them and to break the tie resolutely rather than wavering around? Perhaps rationally you should, for you know that the autocrat will force you to make a choice in the end. But I suppose you would probably have difficulty in sincerely holding this kind of ATC judgment (or at least you will have to temporarily suspend it). This is because your choice itself has important symbolic meaning and will forever change your relationship with your children, so in your mind the act of choice itself cannot be said to be strictly the best. "How can I choose one of them over another? What will they think about my choice?", you would say to yourself. So would it be more difficult for you to form an intention to sacrifice one of your children over the other, as compared to forming an intention to go to Bangkok over The Great Canyon? Certainly. And what predicts your greater difficulty in forming an intention in the former case, I think, is precisely the absence of your tacit ATC judgment that making a choice to break the tie resolutely is the best thing to do—for in the former tragic case it's often quite

[10] Alternatively, our tacit ATC judgment can be taken to support, not the mental act of choosing, but that of picking, intending, or willing any of the options in a tie. Despite the difference in names, I think these all denote the same kind of exercise of our volitional power (see Chang 2009). In other words, to form an intention or a volitional state, we need to first reach first-person integration concerning what we are about to do. But sometimes the integration we reach might actually concern how we ought to exercise our volitional power to resolve the quandary of evaluative equality, incomparability, or parity. It is just that people have different names for the exercise of this kind of power.

difficult for people to sincerely hold onto this kind of judgment.[11]

Still, even if we have an implicit ATC judgment in favor of the act of choosing, sometime psychological unity will still be hard to come by. After all, in cases of evaluative equality or incomparability, all the substantive considerations still do not strictly favor any particular option. The very option that you settle on is only opted for because of the formal consideration in favor of the act of choosing. That is why intentions formed in this way will often have a quite fragile settling force and all-out volition character, for there is another set of considerations and desires that equally pull us to another option at the very moment of intention formation. Often integration has to be reached by bracketing this contrary set of considerations and desires in our deliberative perspective. But the general lesson still holds. The more difficult it is for us to reach an ATC judgment and the accompanying mental integration, the harder it would be us to form an intention.

5.3. Intentions against ATC Judgment

Finally consider the third way in which we might fail to follow the C-principle. Sometime we might survey all relevant considerations, reach an ATC judgment successfully, but still fail to form an intention in accordance with it in the end. In fact, this is precisely what incontinence consists in on Davidson's analysis. Yet we seem perfectly able to form intentions against our ATC judgment in an incontinent way. So

[11] Of course, how difficult it is for one to sincerely form an ATC judgment in favor of the act of choice and then to actually exercise one's volitional power has to depend on one's personality. Some people might just be more resolute in making a choice by temperament. More generally, I think whether one can successfully follow the C-principle and reach integration depends on both one's character traits and the situation that one is in (such as the choice situations that we are considering here). The same thing also holds for the success and failure in following other rational principles for intention. People's making a rational response, like their performing a virtuous action, can be explained by both personal and situational factors.

how can our constitutivist story accommodate this kind of intention?

Again, I think intentions formed in this way are functionally defective. Presumably this is what happens when you intend in an incontinent way: You first reach integration in your deliberative perspective by integrating all the relevant considerations that you have in mind into a single ATC attitude in favor of a certain action. But then some of the considerations become more salient from your deliberative perspective and lead you to commit to a contrary course of action and settle the practical question in an outright manner. That is how you can come to have an incontinent intention. However, although the incontinent intention here seems to carry your commitment and settle the practical question concerning what you are about to do, it can only do so in a defective way. This is because at the very moment of intention formation, you are just in a good position to retract your commitment and reopen the practical question, given that there is at the same time a further, and a more comprehensive, perspective from which you can view the practical question in a different light—a perspective that is represented by your ATC judgment.

Indeed, I would say intentions formed in this third way are functionally the most defective. After all, the previous two kinds of defective intention are still grounded in a kind of fragile psychological integration, which is supported either by some of the relevant considerations that we have covered, or by an implicit ATC judgment in favor of an action type or the act of choosing. But incontinent intentions are based on an extreme sort of first-person disintegration, for they explicitly break away from the integrated practical perspective that we just reached through the formation of the ATC judgment. That is why the standard image of an incontinent person is not someone who is staunchly committed to the action that she judges to

be less than the best. Rather, it is someone who wavers between different courses of action and different points of views, or, in simpler words, someone who cannot reach a single mind in the first place.

Richard Holton (1999) argues that we should replace the traditional conception of weakness of will. Weakness of will, he contends, does not consist in incontinence, i.e. intending against one's better judgment. Rather, it consists in irresolution, or over-readiness in changing the intention that one has already formed. But now we can see how these two conceptions of weakness of will are not entirely separated. An incontinent intention is disintegrated from one's more comprehensive practical perspective and is hence defective in its volitional character and settling force. As a result, an incontinent intention will typically be an irresolute intention that one has difficulty in sticking with.

6. Conclusion

This completes my constitutive story about the relation between agency, integration, and continence. Every story has some morals, so does mine. It teaches us four main lessons.

First, intention is best defined by its all-out volitional character and settling function. Other properties of intention can be plausibly construed as being derived from these two basic defining features. These two features in turn entail that an intention, if it is to retain its defining volitional character and play its defining settling function well, has to be a practical attitude of us as a whole. That is why agency, even when it is characterized minimally as the capacity for fulfilling our intentions, has to presuppose a fair degree of psychological integration.

Second, the C-principle prescribes the most general way for us to reach integration in our first-person practical perspective, which is the very type of psychological integration presupposed by our intention. That is why there is a constitutive relation between following the C-principle and forming an intention.

Third, there is actually a rational *integration principle* governing our intentions, since intentions rested on disintegration in our first-person deliberative perspective will typically be defective in their defining volitional character and settling function.[12] As I suggested in Chapter 1, our intentions are the primary triggers of the instrumental requirement and the main subjective sources of normative reasons for action. So our constitutive story here reveals how the main subjective sources of reasons for action are actually subject to a rational integration principle. In fact, as I will argue in Chapter 5, it is in virtue of how an intention typically represents a practical attitude of us as a whole that it is able to ground a normative reason with the required normative force to address our normative question. So in a way the general adherence to the integration principle is what confers reason-giving power to an intention.

Finally, the C-principle has a place in practical rationality not because it is accepted by us as our second-order principle, but because it constitutively contributes to the proper function of our intention. So our constitutivist story here actually points to a more general *functionalist conception of practical rationality*. According to it, a certain psychological pattern or a particular psychological step is proper and prescribed by practical rationality in virtue of how it is constitutively involved in the proper function of the different aspects of our agency. We have

[12] Of course, I am not denying that there can be other standards for rationally evaluating our intentions. For instance, intentions whose formation takes up too much time or cognitive resources might still be irrational from a pragmatic perspective.

already seen in Chapter 1 that the instrumental principle prescribes the psychological patterns or steps that are indispensable for the proper function of our executive mechanism for fulfilling our intentions. Our constitutivist story here further shows that the principle of non-contradiction and integration principle in fact prescribe another psychological pattern that is indispensable for the proper function of our decision-making mechanism for forming intentions (i.e. the lack of conscious, synchronic, and direct contradictory intentions). This then suggests that the origins of the requirements of practical rationality can actually be explained by their constitutive involvement in the proper function of different aspects of agency. I shall further defend this view later on in Chapter 4.

Chapter 3 Misinformation, Subjectivism, and the Feasibility Principle

1. Introduction

On the Humean view of practical normativity, it is our desires that ultimately determine what actions we should perform. Since the ultimate sources of the "should" here are our desires, Humeans believe that there are no further rational principles that prescribe us to desire any particular action. Instead, there are only non-substantive rational principles governing the effective realization of our desire or the internal coherent relations between them. After all, even if these principles tell us that we should form a particular instrumental desire or that we should not hold a desire whose realization is incompatible with that of another, the "should" here is still, in line with the Humean view, generated by one of our desires. This line of thought has led Humeans to conclude that the rational criticism of desire can never be made against a desire's substantive content; rather, it can only be made on the grounds of instrumental irrationality or internal incoherence.

But misinformed desire presents one major exception to such a Humean restraint on content criticism. It is widely agreed among philosophers that any plausible subjectivist account of normativity should exclude misinformed desires from the eligible subjective sources of normative reason (Brandt 1979: 13; Railton 1986a: 11, 1986b:173; Rosati 1995: 229-32; Sobel: 1994, 2009; Smith 1994: 151-6; Williams 1982: 103-4, 1995: 36). Almost all contemporary Humeans have followed suit. The rejection of misinformed desires, however, is precisely targeted at a

desire's content, for being misinformed is actually a content property. This then raises the suspicion that Humeans, in rejecting misinformed desires, might have to willy-nilly engage in a kind of content criticism of desire.

In fact, many normative realists have recently argued that the rejection of misinformed desire can only make good sense within the framework of normative realism (Enoch 2005: 759-87; Lillehammer 2000: 173-190; Parfit 2011a: ch.13; Ripstein 2001: 37-55). According to the *tracking account* that they proposed, misinformed desire lacks reason-giving power because its misinformed content fails to properly track the relevant reason-giving features of its object. In Williams's famous gin/petrol case, for example, the agent's misinformed desire to get the liquid in the bottle (which she takes to be gin but is in fact petrol) cannot ground a normative reason for her just because its content fails to track one crucial reason-giving feature of the liquid, namely, its being petrol. Indeed, if that objective feature cannot give her any normative reason by itself, per the suggestion of subjectivists, then it would be hard to see why her misinformation about it should matter and why her misinformed desire should have no reason-giving power. So, as these realists claim, in order to reject misinformed desire, Humeans have to go substantive and presuppose that there are reason-giving features in the world that the content of our desire must track if it is to have any reason-giving power at all.

Some philosophers have instead contended that the rejection of misinformed desire can make perfect sense on a subjectivist or Humean picture of normativity. According to the *genuineness account* suggested by those sympathetic to the subjectivist program, misinformed desire lacks reason-giving power because it fails to represent our genuine favoring attitude toward the object of desire in the first place (Sobel 2009; Railton 2012). In the gin/petrol case, for instance, the agent's

misinformed desire to get the liquid in the bottle cannot give her a normative reason to get it just because her misinformed desire can only represent her favoring attitude toward the liquid *as it seems to her* in light of her misinformation, and not toward the liquid *as it really is*. Given that she lacks a genuine favoring attitude for the liquid as it really is, for sure Humeans can justifiably claim that there is no normative reason for her to get it, without the need to invoke some pre-existing reason-giving features. So these philosophers think that the Humean rejection of misinformed desire is only made on the basis of its failure in reflecting a genuine, full-fledged desire on an agent's part; Humeans actually don't have to go substantive for this move.

In this chapter, however, I propose an alternative *feasibility account* for the problem of misinformed desire. According to this account, misinformation actually makes the content of misinformed desire infeasible, and this kind of desire lacks reason-giving power in turn because its infeasibility renders it unlikely to be fulfilled and hence rationally defective in light of its internal aim. For example, some European explorers in the late 16th century desired that they discover the golden kingdom of El Dorado in South America. Their desire could be said to be misinformed because its content contained a representational element that purported to be true of something but failed to be true of any (i.e. the golden kingdom in South America). Given this, there was simply no existing object that corresponded to this misinformed representational element in its content, and, *a fortiori*, no existing object that could constitute the realization thereof. So their desire that they discover the golden kingdom in South America was simply infeasible due to their misinformation. The infeasibility of their misinformed desire in turn made it rationally problematic: When they embarked on their journey of

fulfilling their infeasible desire for finding the golden kingdom, they were only engaging in a futile attempt that was bound to fail. But such a destined failure in desire fulfillment, as I will argue, actually constitutes a defect in light of both the instrumentalist picture of rationality and the internal aim of desire itself. That is why the European explorers' misinformed desire can be said to lack reason-giving power.

This feasibility account, as I contend, offers a rationale for rejecting misinformed desires that proves to be more helpful for subjectivists who aim to ground normativity on motivating or volitional states. It will also lead us to see how there are actually two very different kinds of subjectivist accounts of normativity, i.e. motivation-based and liking-based subjectivism, which aim to ground normativity in our agential and receptive capacities respectively. More importantly, and perhaps somewhat surprisingly, the feasibility account will show how the very conceptions of practical rationality and desire that underwrite the orthodox Humean view about normative reasons for action already implicate a rational *feasibility principle*, a semi-substantive rational principle according to which any desire whose content is not minimally feasible and is destined to failure in its fulfillment is rationally problematic. So it turns out that Humeans have to go semi-substantive in the rational criticism of desire given their theoretical commitments.

Indeed, this feasibility principle is most plausible when it comes to intention or desires with volitional character, for this kind of desire actually has its own efficacy in leading an agent to fulfill her intended state of affairs as its internal aim. As such, the feasibility principle complements the integration principle that was derived in the foregoing chapter. Together they make for a neo-Humean picture of practical rationality that moves beyond mere instrumental and formal coherence principles.

In Section 2 I examine the existing accounts for the problem of misinformed desire and expose their limitations by distinguishing between two kinds of subjectivism. I propose the alternative feasibility account in Section 3 and explain why our desire and intention are subject to the rational feasibility principle in Section 4.

2. The Rationales for Rejecting Misinformed Desire

2.1. Why Is Misinformed Desire Problematic?

Misinformed desire comes in two types. We can have a *misinformed telic desire* for something X as an end because of our ignorance or false beliefs about its relevant properties.[1] Or we can have a *misinformed instrumental desire* for something X as the means to another end Y because of our ignorance or false beliefs about the means-end relation between X and Y. It is quite easy to come up with a subjectivist rationale for rejecting misinformed instrumental desire. After all, misinformed instrumental desire motivates us to devote our resources to the false measures that cannot in fact fulfill our telic desires. As such, this kind of desire is problematic simply on the instrumentalist picture of rationality, which is taken by most subjectivists to be the most minimal and the least controversial conception of practical rationality.[2] This is why almost all subjectivists are united in their rejection of misinformed instrumental desire.

[1] I borrow the term "telic desire" from Parfit, see his (2011a: 44).

[2] Of course, many subjectivists have embraced a more substantive picture of practical rationality that goes beyond mere instrumental rationality, see Korsgaard (1996: ch. 3, 2009: ch. 4) and Smith (1994: ch. 5, 2015). Also see Williams (1982: 104-5).

It is more difficult, on the other hand, to come up with a subjectivist rationale for rejecting misinformed telic desire. In fact, the rejection of this kind of misinformed desire is particularly vulnerable to the attack from realists. After all, if the relevant features of an object do not provide us with any pre-existing normative reason to desire it or not, then it would be difficult to see how true beliefs about these features could ever imbue our desire with more justificatory force, and how misinformation about them could ever deprive misinformed desire of its reason-giving power. Derek Parfit (2011a), for instance, contends that it is inconsistent for subjectivists to privilege informed desires based on true beliefs about the relevant facts about an object and to claim at the same time that these relevant facts do not provide any normative reasons for us to desire it. This "incoherence argument" constitutes one of his main arguments against subjectivism (2011a: ch. 4). Similarly, David Enoch (2005) argues that the move of privileging desires formed under rational or idealized conditions makes the best sense if the rational or idealized conditions here are thought to help these desires better track pre-existing normative reasons (Also see Lillehammer 2000 and Ripstein 2001: 37-55). Thus, realists think that the problem of misinformed telic desire lies in how misinformation prevents it from tracking pre-existing normative reasons properly and from having adequate justificatory force. It is only in light of this kind of *tracking account*, realists claim, that it makes good sense for subjectivists to disqualify misinformed telic desire from the eligible subjective source of normative reason. (For the sake of simplicity, in what follows I will use the term "misinformed desire" to refer to misinformed telic desire alone, unless otherwise specified.)

There is, however, one natural subjectivist explanation for the problem of misinformed desire. David Sobel (2009) has recently argued that the main reason

for subjectivists to reject misinformed desire is not that it gets the relevant reason-giving features wrong, but that it gets the non-normative features of its object wrong. To illustrate this, consider an example he offers: There are no pre-existing normative facts concerning whether a certain ice cream flavor is good for you or whether you have a normative reason to get it. But a desire formed in light of full information about its relevant features is indeed the very kind of desire that subjectivists should privilege, because "only desires formed in light of an accurate understanding, phenomenological and otherwise, of what an option would be like are responsive to the true nature of the option under consideration (345)." And a misinformed desire for something X, Sobel contends, can only represent our favoring attitude toward the X *as we conceive of it in light of our misinformation about it*, but it does not really reflect a genuine favoring attitude on our part toward the X *as it really is*.

Peter Railton (2012) has also proposed a similar argument. According to him, our desire constitutively involves a favorable affective attitude toward a representation of an object X to be acquired (24-5), and such an affective attitude is sensitive to the feedback from the experiences that we have when coming in touch with the real thing X (34-6). And the reason why our misinformed desire for something X is problematic, he contends, is that it is based on inadequate experiential feedback about the real thing X—which can turn out to be something that we do not really like or enjoy.

Thus, according to both Sobel and Railton, the problem of misinformed desire lies in how it fails to reflect our genuine favoring attitude toward the real thing due to our misinformation about, or lack of experience of it. This *genuineness account* then explains why subjectivists can justifiably reject misinformed desires. After all,

this kind of desire simply fails to reflect in the first place our genuine favoring attitude for an object as-it-really-is, an attitude that is supposed to determine whether we truly have a normative reason to acquire the very thing in question or whether it would really be good for us.

2.2. Two Kinds of Subjectivism

But further reflection reveals the limitation of the genuineness account. This is because there are two ways to understand the favoring attitude that is supposed to be the subjectivist ground for normativity, which then give rise to two kinds of subjectivist accounts of normativity:

> (1) Motivation-based Subjectivism: Desire is the source of normativity in virtue of how it typically reflects how we have a favoring attitude toward the realization of a prospective object or state of affairs that we see under a favorable light and want to realize.

> (2) Liking-based Subjectivism: Desire is the source of normativity in virtue of how it typically reflects how we have a liking for an actual object as-it-is, or for a prospective object or state of affairs as it-will-in-fact-be when it is realized.

The genuineness account is well-suited to help liking-based subjectivism: Only if you are fully informed of the features of an actual object as-it-really-is (say, what a particular ice-cream flavor is really like (Sobel 2009: 343-44)) or the features of a prospective object or state of affairs as it-will-in-fact-be (say, what it will be really

like if you become a top athlete (Railton 2012: 39-4)) can you be sure that you will have a genuine liking for it in the end. Your misinformed desire presumably fails to reflect this kind of genuine liking for the actual thing or state of affairs that you will get, and that is why this kind of desire cannot be the proper subjective source of normativity on liking-based subjectivism. In fact, both Railton and Sobel have stressed that an informed desire for an object X has to be based on adequate phenomenological or experiential information about what it will be really like for us to have X.[3] And the examples that they appeal to all involve an agent with a misinformed desire for something that she will not really like or enjoy in the end. So the version of subjectivism that they had in mind and that their accounts are supposed to help, I presume, is one that is more akin to liking-based subjectivism.[4]

The genuineness account seems less helpful, however, for motivation-based subjectivism. Suppose you have in mind a mental representation of a prospective state of affairs (say, that you get a bottle of gin that is in front of you) and you desire to bring it about as an end in itself. But suppose this mental representation is actually based on a misrepresentational element - the bottle of liquid that you take

[3] Sobel claims that "idealization is an attempt to get those facts clearly and accurately into view...It is important for my purposes that the notion of facts here includes an accurate phenomenological impression of what an option would be like for one (2009: 344)". Railton also explicitly claims that one's informed desire for a particular kind of life needs to be based on full experience about what it is really like to lead that particular kind of life (2012: 40). This stress on phenomenological information can also be seen in their earlier work. In his earlier paper, Sobel (1994) distinguishes between the report and the experiential model of full information and contends that the latter kind of information is indispensable if one's informed desire is to determine what is good for one. Railton's suggestion that the full information must be vivid (1986a: 21-22, 30-1) also indicates how he understands full information primarily as a kind of phenomenological information.

[4] Notice that liking-based subjectivism is distinct from the view defended by Chris Heathwood (2005) and Eden Lin (2019), which is a version of actual-desire subjectivism. According to them, all actual telic desires, even including misinformed ones, are valid subjective sources of normativity; it is just that misinformed telic desires can often be outweighed because they are more likely to frustrate other actual telic desires or likings that we also have. The liking-based version of subjectivism that we are considering here, in contrast, is a version of idealized-desire subjectivism. This view simply excludes our actual misinformed telic desires from the potential subjective grounds of normativity, and the rationale for excluding these desires is that they fail to reflect our genuine likings for different types of things.

to be gin is actually petrol. Now despite your misinformation about the liquid as-it-really-is, you still genuinely want it to be true that you get a bottle of gin in front of you, and you still see this prospective state of affairs in an attractive light and commit yourself to its realization. So why can't this attitude give you a normative reason to make it the case that you get a bottle of gin in front of you, in line with the spirit of motivation-based subjectivism? More generally, the standard objects of our likings are things-as-they-really-are, so it makes sense to say that our genuine likings can only be reflected by desires formed under full information about the objective features of actual things. The same point, however, seems more controversial when it comes to our motivating state. Indeed, according to the motivating conception of desire, the standard object of our motivating state is a mental representation of a prospective state of affairs to be realized (Smith 1987). So despite our misinformation about the actual things, our misinformed desire *qua* a motivating state still constitutively involves a genuine favoring attitude toward the realization of a prospective state of affairs.[5] And on motivation-based subjectivism, this is precisely the kind of attitude that is supposed to give us a normative reason to realize it (for the sake of simplicity, in what follows I will call this kind of attitude *favoring attitude toward the prospective*).

I am not saying that our mere motivation to bring about a prospective state of affairs alone can generate a normative reason for action. Perhaps our favoring attitude toward the prospective needs to be maximally influenced by rationality; perhaps it has to carry our personal commitment to its realization (see Wallace 2004: 459-62); perhaps it also needs to further involve our positive evaluation

[5] This point has been recognized both by Sobel (2009: 347) and Dorsey (2017: 208).

about how the prospective state of affairs in question is valuable (Street 2012: 43-5), or our willing such a state of affairs to be reasoning-giving for us (Chang 2013a: 176-80, 2013b: 92-7). Regardless, the spirit of motivation-based subjectivism lies in the recognition of the normative importance of our favorable attitude toward the prospective, yet our misinformation about the actual things doesn't seem to impinge on the genuineness of this kind of attitude. So the genuineness account doesn't seem to explain why our misinformed desire *qua* a motivating state really fails to reflect a genuine favoring attitude on our part, and why it should consequently lack reason-giving power on motivation-based subjectivism.

Of course, one might object that we still need full information about what the prospective thing that we desire will be really like when it is realized. Even if your misinformed telic desire involves your favoring attitude for making it the case that you get a bottle of gin in front of you, it might still turn out that this is not what you really want; perhaps you just won't enjoy the taste of gin anyway. So, as you might think, only if we have full information about what a prospective state of affairs will be really like when it is realized can our favoring attitude for its realization be genuine in a strict sense and give us a normative reason to fulfill it. Indeed, Dale Dorsey (2017) has suggested that one necessary condition for our genuinely valuing an object X is that we would still have a favoring attitude toward X when we have full information about the X as-it-really-is and consider the different ways that X might be. So any version of value-based subjectivism, he contends, should invoke desires or values formed under full information and imaginative consideration just because full information is a condition for genuine valuing. And you might think that the same rationale also applies to our favoring attitude toward the prospective.

However, is it a condition for having a genuine favoring attitude of this kind that

we have full information about what the prospective state of affairs will be really like? Should the reason-giving power of our favoring attitude toward the prospective be *always* conditioned on this kind of full information on motivation-based subjectivism? I think we have some reason to doubt this. A pregnant woman might desire that she becomes a full time mother in the future, sincerely believe that this is the best life option for her, and strongly commit herself to it. Although it might turn out that this is not what she really wants because of her misinformation about what that life is actually like, she simply doesn't care. For her the force of her love and devotion is precisely embodied in her commitment to full-time motherhood, whatever its consequences will be. Can't her favoring attitude for the prospective state of affairs that she has in mind still mean *something* on motivation-based subjectivism and give her *some* normative reason to bring it about? More generally, people commit themselves to various projects that they take to be expressive of who they are; they make an inner vow to themselves, "This is what I am going to commit myself to, for better or worse." Can the mere fact that they will not enjoy or like it in the future completely nullify the reason-giving force of this kind of attitude toward the prospective? I think the answer is no, unless we want to give total normative deference to our likings and passive propensities for enjoyment, and to unduly downplay the normative importance of the attitudes that we can actively form toward prospective things in the future (see Appendix 1 for a more general discussion on these two types of subjectivism).[6]

More importantly, even if we concede that we can have a genuine favoring attitude for the prospective only if we have full information about what the

[6] Both Chang (2013a: 92-4, 2013b: 179-80) and Korsgaard (1996: ch.3) have argued that normative reasons for action are grounded in our active rather than passive capacity. I think this is the right way to think about the matter. I will later argue in Chapter 5 that normative reasons for action are in fact grounded in our agential capacity for action.

prospective state of affairs involved will be really like, the genuineness account is still not able to help motivation-based subjectivism. This is because, unlike our liking, our favoring attitude toward the prospective is actually subject to a distinctive problem of infeasibility. Consider the European explorers' desire that they discover the golden kingdom of El Dorado. Their favoring attitude toward the prospective state of affairs they had in mind could be said to be genuine in a strict sense: They had full information about what its realization would be like, and presumably they would have really liked, enjoyed, or valued it when it was realized. It was actually what they genuinely wanted all along. Their misinformation indeed made their desire problematic, but the problem with their misinformed desire was not that it failed to reflect their genuine favoring attitude. Rather, the problem was that the favoring attitude that it involved was for something that simply didn't exist and couldn't be brought about in the first place. To properly explain why motivation-based subjectivists have a good reason to reject this kind of misinformed desire, we have to move beyond the genuineness account and invoke an alternative *feasibility account* for the problem of misinformed telic desire.

Indeed, it seems to me that the genuineness account explains how misinformation can prevent our desires from reflecting our genuine likings, and why these desires are objectionable for liking-based subjectivism. The feasibility account, on the other hand, shows how misinformation can prevent our favoring attitudes toward the prospective from being feasible, and why our misinformed desires *qua* motivating states are problematic for motivation-based subjectivism. Let me explain it.

3. The Feasibility Account

3.1. Why is Misinformed Telic Desire Infeasible?

Let us begin by first considering how our telic desire *qua* a motivating state can ever be misinformed in the first place. On the face of it, this kind of desire does not even seem to be susceptible to the problem of misinformation at all. After all, its content is a prospective state of affairs that is supposed to be realized or "made true" rather than being true of something in the first place, so it is a bit difficult to see how it can ever be misinformed. Your telic desire that you bake a chocolate cake from scratch, for instance, cannot strictly be misinformed because it only contains a mental blueprint of a cake that is supposed to be made true later instead of being true of an existing cake.[7] But I think our telic desire as a motivating state can still be said to be misinformed if its content has some representational elements that purport to be true of something in the first place. For instance, your telic desire that *you get a bottle of gin that is in front of you* as an end in itself presumably includes a representational element that purports to be true of an object out there, i.e. the liquid in the bottle in front of you. It is in virtue of the failure of a representational element like this, I think, that a telic desire as a motivating state can ever be misinformed.

What is then the problem with a misinformed telic desire whose content contains a representational element which fails to correctly represent things? Consider what motivation-based subjectivists should say about your misinformed telic desire that you get *a bottle of gin that is in front of you* in the gin/petrol case.

[7] Of course, this telic desire can be said to be misinformed if the chocolate cake turns out to be something that you do not genuinely like. But this will only bring us back to liking-based subjectivism.

One thing they should say is that in virtue of your favoring attitude toward the content element involved in your desire, there is a normative reason for you to realize it and make it the case that you get a bottle of gin in front of you. However, there is another thing that motivation-based subjectivists should say. They should also say that since the content element in your desire contains a misrepresentation that fails to be true of any existing object, it is very unlikely to be made true. Your desired prospective state of affairs that you get *a bottle of gin in front of you* is unlikely to be made true because there is simply no bottle of gin out there for you to get.

More generally, if the content of your telic desire involves a misrepresentational element, then it would be *unrealistic*—in both the sense of having no basis in reality and the sense of being infeasible: Its content involves a misrepresentation that purports to be true of an existing object but fails to be true of any (e.g. getting *a bottle of gin that is in front of you*). As a result, such content can hardly be made true by an existing object that corresponds to the misrepresentational element and that constitutes the realization thereof (e.g. there is simply no object that corresponds to *a bottle of gin that is in front of you*). Therefore, a misinformed telic desire like this is often infeasible and likely to lead to failures in your attempts to realize it. You will try to find out or think up the means to realize it, devote your limited and valuable resources to its fulfillment, but only to discover in the end that its content cannot really be made true. So your misinformed telic desire *qua* a motivating state turns out to be problematic in the same way as the European explorers' desire for the golden kingdom in South America: These desires all come out of touch with reality due to misinformation and consequently have realization conditions that can hardly be fulfilled. These desires, as motivating states that aim at their own realization, in

turn set us out on a voyage that is highly likely to result in eventual failure.

Thus, misinformation not only makes our telic desires unable to reflect our genuine likings for things-as-they-really-are; it also leads our telic desires *qua* motivating states to target at prospective states of affairs that are unrealistic and infeasible in the first place. And it is the latter problem that poses a special threat for motivation-based subjectivists, who aim to ground normativity not in our likings, but in our favoring attitudes toward the prospective. That is why the genuineness account alone cannot help motivation-based subjectivism.

Notice I am only claiming here that misinformed telic desire is highly likely to result in failure in realization. This is because sometimes the misrepresentational elements involved in its content might still be accidentally true of something. For instance, if there is, unbeknownst to you, another bottle of gin in front of you along with the petrol, then your misinformed desire might still be feasible in the end. I suppose a lot depends on the fine-grainedness of the content of our telic desire and what it purports to be true of in the first place. If your desire is that you get *this very bottle of gin* that is in front of you, then your desire would be infeasible in the end (because this very bottle of liquid in front of you contains just petrol). But if you only desire that you get *a bottle of gin that is within your reach*, then perhaps your desire wouldn't be infeasible (because there are both bottles of petrol and gin near you). And if what you really want is only that you get *any bottle of gin*, then your desire would not even be misinformed in the first place (because it does not purport to be true of any stuff near you). Real life cases are far more complicated. That is why I am only making the more qualified claim that misinformed telic desires are highly

unlikely to be satisfied due to its misinformed content.[8]

3.2. Why is Infeasible Desire Rationally Defective?

But why is infeasibility a defect for our telic desire *qua* a motivating state? Why is it able to deprive our misinformed desire of its reason-giving power on motivation-based subjectivism? To see this, consider the following passage from Hume:

> ['T]is only in two senses, that any affection can be call'd unreasonable. First, when a passion […] is founded on the supposition of the existence of objects, which do not really exist. Secondly, when in exerting any passion in action, we chuse means insufficient for the design'd end (1739-40: Book II, Part III, Section III).

Of course, Hume thinks that the irrationality involved in these two cases is only a kind of theoretical irrationality. But if you think that the second case here also involves a further kind of practical irrationality, or at least a kind of practical defect, then you would also have good reason to say the same thing about the first case.

[8] Can it be argued that our telic desires must be desires with coarse-grained content and for general kinds of things in the world (say, the agent's telic desire in the gin/petrol case has to be a desire for gin in general rather than a desire for this very bottle of gin), such that they cannot really be misinformed in the end? I don't think so. For one thing, our telic desire can often have fine-grained content and target at something specific as an end in itself. One can, for instance, have in mind an ideal wedding, with all the details thought through, and one can surely want to realize that detailed, fine-grained mental content as an end in itself. For another, even a telic desire with course-grained content has representational purport and can therefore be subject to the problem of misinformation. Your telic desire for gin would be misinformed, for instance, if gin were to become no longer available in the world (because, say, a new plant disease devastated juniper berry harvests, the very kind of berry that is used to flavor gin). In fact, I think almost every desire is subject to the problem of misinformation, as long as it has any representational content at all.

This is because both cases presumably involve the failure in the satisfaction of our passions (if passions can be said to have satisfaction conditions). In fact, we can even say that Hume's diagnosis of the two kinds of unreasonable passion here corresponds to the defects of misinformed telic and instrumental desire respectively. And this actually helps us to see why infeasibility is a kind of rational defect.

Instrumental rationality, as I argued in Chapter 1, is a rational executive ability for fulfilling desires (or more precisely speaking, for fulfilling desires with volitional character). On this executive conception of IR, a failure in rationality is primarily a failure in desire realization. Now misinformed instrumental desire simply leads us to take the measures that cannot in fact fulfill our telic desires, and that is why it is problematic. Misinformed telic desire, on the other hand, also moves us to take the means to realize it. It is just that we will eventually discover that its content cannot really be brought about because it is infeasible—there is simply nothing that corresponds to the misrepresentational elements in its content and that constitutes the realization thereof. Both kinds of misinformed desire lead to failures in desire realization and are therefore rationally defective on the instrumentalist conception of rationality. It is just that in the case of misinformed instrumental desire, the failure pertains to taking the wrong means; whereas in the case of misinformed telic desire, the failure results from having an infeasible end in the first place. That is why the infeasibility of misinformed telic desire can be said to constitute a kind of rational defect (I will explain in the next section what kind of infeasibility constitutes a rational defect for desire in more detail).

The same point can also be made by focusing on the motivating dimension of desire. According to the direction of fit analysis of belief and desire, belief *qua* a representational state aims at representing a state of affairs that already obtains in

the world, whereas desire *qua* a motivating state aims at realizing a prospective state of affairs. Given this structure, belief can be said to be constitutively governed by the aim of truth, while desire by the aim of efficacy. If a desire, however, has as its content a prospective state of affairs that is infeasible and unlikely to be realized, then it would be problematic simply in light of its own constitutive aim of efficacy, just as a belief with false content is problematic in light of its own internal aim of truth. This also explains why the infeasibility of misinformed telic desire constitutes a kind of defect.

One might think that on the direction of fit analysis, it is the world that is supposed to fit the content of desire. So even if a misinformed telic desire has an infeasible content that is very unlikely to be realized, the problem, as you might think, doesn't really lie in its infeasibility, but in the world's not cooperating (by not containing an object that corresponds to the misrepresentational element in misinformed desire).

This understanding of the direction of fit of desire, however, is a bit crude. The idea of "fit" in the direction of fit analysis, I think, has a built-in causal connotation. To say that belief aims to fit the world, strictly speaking, is to say that belief is an intentional state whose content causally tracks (at least in normal circumstances) the states of affairs in the world and the evidence for their obtaining. Similarly, to say that desire aims to make the world fit it is to say that desire is an intentional state whose content causally influences (at least in normal circumstances) an agent's action and leads her to realize the prospective state of affairs contained in it. Indeed, it is only on this causal understanding of direction of fit that we can explain how desire can motivate and regulate an agent's action in a way that is sensitive to its content. So the internal aim of desire is not just the world's fitting its content.

Rather, it is its own causal efficacy in changing the world in a way that fits its content. That is why the infeasibility of misinformed telic desire can be a defect in light of its constitutive aim.

More generally, human beings and higher animals have a distinctive mental capacity to have prospective objects or states of affairs in mind, to form favoring attitudes toward them, and to set out to bring them about. And on motivation-based subjectivism, it is precisely this kind of favoring attitude toward the prospective,[9] together with our executive capacity to bring about the prospective thing that we favor, that ultimately gives rise to normative reasons for action. In fact, the orthodox Humean view of normative reason is exactly premised on this basic line of thought and attempts to ground normative reasons for action in our motivating attitudes and our instrumental capacity to fulfill them. However, as we have just seen, both the instrumentalist conception of rationality and the motivating conception of desire take efficacy or desire-realization as their internal, constitutive aim. And the infeasibility of our misinformed desire simply constitutes a defect in light of this aim. In fact, on motivation-based subjectivism, normative reasons for action are reasons pertaining to the *realization* of the prospective states of affairs that we favor, so for sure misinformed desires for unrealizable states affairs would lack reason-giving power on this view (Indeed, if these desires had reason-giving power, they would implausibly ground normative reasons that ask us to engage in futile attempts to realize the infeasible and to waste our general resources for desire-fulfillment). That is why motivation-based subjectivists—Humeans in particular, have a good rationale to reject misinformed desire as a motivating state toward a prospective state of

[9] Again, in addition to our motivation toward the prospective, our favoring attitude perhaps needs to further involve our positive evaluation, commitment, or our taking something to be a reason.

affairs to be realized. This then completes my *feasibility account* for the problem of misinformed desire.

4. The Feasibility Principle

4.1. A Further Rational Principle for Intention

Although the feasibility account helps motivation-based subjectivists to explain why they have a good rationale for rejecting misinformed desire on the basis of its infeasibility, it also shows, surprisingly, that they cannot remain completely content-neutral with the rational criticism of desire given their own theoretical commitments. Indeed, the very foundations of the orthodox Humean view (i.e. the motivating conception of desire and the instrumentalist conception of rationality) take effective desire-realization as their constitutive aim, an aim which in turn imports a rational *feasibility principle* for our desire. According to it, any telic desire that is not minimally feasible is rationally defective.[10] It is in light of this principle, as we have seen, that it makes sense for subjectivists to reject misinformed telic desire *qua* a motivating state and to divest it of its reason-giving power. This very rational principle, however, is also a semi-substantive rational principle targeted at the content of our telic desire. Although it does not require us to desire any particular end, it does forbid us from desiring a specific type of end, i.e. an infeasible

[10] Nozick (1993: 144) similarly contends that a rational person is someone who "does not have desires that she knows are impossible to fulfill" (Also see Hublin 1991: 25; Wallace 2001: 20-1). My view here is different from Nozick's mainly because he takes desire to be subject to a subjective or belief constraint about the possibility of satisfaction, whereas I think that desire is subject to a broader objective constraint of feasibility given its constitutive aim of efficacy. As I will argue later, the former belief constraint can be plausibly construed as being derived from the latter objective constraint on desire.

end. This then shows that to reject desire with misinformed and infeasible content, Humeans do have to go semi-substantive with the rational criticism of desire.

This feasibility principle is intuitive because we do rationally criticize people's projects by whether they are realistic or feasible enough, regardless of our moral or value judgments about them. We are willing to accept that the kinds of life projects people have normative reasons to pursue are relativized to their own desires and preferences. Still, life projects that are simply infeasible (say, attempting to learn to run at light speed) will often be seen irrational and foolish regardless of their moral or evaluative character.

In fact, the feasibility principle is actually most plausible when it comes to a kind of desire that figures as the best potential subjective source of normative reason for action, namely, intention or volitional state. As I argued in Chapter 1, our intention has a *settling function*—it functions to settle our practical question with its all-out volitional quality. But that is not the end of the story. Our intention also has an *executive function*—it also functions to lead us to act and to realize its content, which is a prospective state of affairs to be realized. It is because of this executive function that intention has a distinctive satisfaction condition (which I highlighted in Chapter 1): It is realized only when the desired state of affairs is brought about by the agent herself *because of it*.[11] This in turn explains why our intention has a stronger built-in claim to its own efficacy and is subject to governance of the instrumental principle as its core rational constitutive principle. Indeed, I think it is the executive function of intention that leads most philosophers to regard our intention (or volitional state in general) to be both the main trigger of the

[11] Setiya (2007: 664) even argues that the content of one's intention to φ, strictly speaking, is that one is going to φ because of that very intention.

instrumental requirement and the best source of a kind of normative reason grounded in our agential capacity to realize the prospective (Bratman: 2009a; Broome 2013a; Brunero 2012; Korsgaard 1997; Schroeder 2009; Wallace 2001; Way 2010a).

But an intention that is not even minimally feasible will simply be defective in light of its executive function. Indeed, given this function, we can say that the point of intention is to make something happen and eventually bring about the prospective state of affairs that we intend. That is why the typical justification for an intention is derived from the justification related to the fulfillment of the intended state of affairs. For instance, if we travel back in time and ask the European explorers to justify their intention, they would presumably tell us how good it would be if they actually found the El Dorado. But the infeasibility of an intention can simply defeat its point and undercut its typical justification. That is why the feasibility principle presents a valid basis for us to rationally criticize people's intention as a kind of mental state with a distinctive executive function.

Indeed, I think the feasibility principle plausibly explains why there is a belief constraint on our intention. Given the executive function of intention, it does seem irrational for an agent to intend to bring about a state of affairs S while at the same time believing that S is not even possible or feasible in the first place—for the precise point and justification for intending so just lies in the realization of S. That is why if an agent is to be rational in intending to bring about S, then she must not believe that S is impossible or infeasible.[12] This belief constraint, I think, is in fact a

[12] Here I go for a weaker reading of belief constraint. So to be rational in intending to φ, what is needed is not our having a further belief that φ-ing is feasible or possible; what is needed is instead *our lacking* the belief that φ-ing is infeasible or impossible in the first place.

subjective, second-best rational requirement derived from the feasibility principle[13]: our intention formation should be rationally sensitive to our beliefs about infeasibility just because infeasibility in the first place constitutes a rational defect for our intention as a kind of intentional state that aims to bring about a prospective state of affairs.

Of course, given the settling function of intention, sometimes the point of intending something might lie more in resolving our practical question and thereby facilitating our inter- and intra-personal coordination. For example, when faced with the difficult choice between Bangkok and the Grand Canyon, I can form an intention to settle the matter and to allow others (including my future self) to anticipate what I am going to do in the future and plan accordingly. The justification for my intention might be grounded more in such settling and coordinative considerations than in the substantive considerations for either option.[14]

But this kind of justification is only secondary. The primary point of intending something is still to make something happen and to bring about a certain end result. When it comes to the rationality and justification of intention, the consideration of objective infeasibility seems to far outweigh settling and coordinative considerations. A teenager wavering between various life options can settle the matter by intending to become the first person to run at light speed, and people can coordinate very well with her thanks to this intention (perhaps she would practice running in her backyard every afternoon). But do such settling and coordinative

[13] It is a second-best rational principle because, in order to follow the feasibility principle, the best that we can do is just to consult our beliefs about what things are infeasible or impossible for us and try to comply with the belief constraint.

[14] This point is precisely what Mark Schroeder relies on in his argument that considerations related to the effects of intention can sometimes be the right kind of reasons for intending something (2012).

considerations rationalize an objectively infeasible intention like this? No. An intention like this is simply foolish, even if it resolves our practical question and contributes to interpersonal coordination.

What is more, subjective infeasibility actually undercuts an intention's settling and coordinative effects in the first place. If we know that a prospective state of affairs S is something that we simply cannot bring about (say, learning how to run at light speed) or something that we won't bring about (say, actually drinking the toxin in the Toxin puzzle (Kavka 1983)), then we will have real difficulty in committing ourselves to bringing about S and thereby settling our practical question in the first place. People who know S's infeasibility will not be able to rely on the obtaining of S in their future projects either.[15] So a subjectively infeasible intention like this often cannot be supported and rationalized sufficiently by settling and coordinative considerations. Feasibility thus presents a weightier consideration in our rational evaluation of intention.

Of course, sometimes an infeasible intention can even be supported by considerations wholly unrelated to its settling and executive function. A rich person might reward you with one million dollars, for instance, if you intend to run at light speed. I would say this kind of infeasible intention is actually an irrational intention supported by the wrong kind of reason, which is much like a false belief that is supported by weighty pragmatic considerations. I will deal with this problem later on in Chapter 5.

4.2. How Feasible?

[15] The most that they can rely on is our taking the means to bring about S.

How feasible must our intention be if it is not to be deemed rationally defective? I am first inclined to say that it depends on the weight of the justification for the prospective state of affairs that we intend to realize. As we have just seen, the typical justification for an intention is derived from the justification for the intended state of affairs to be realized. And the degree of an intention's infeasibility actually signifies how likely its typical justification will be undermined. Accordingly, the rationality of an intention should be determined by both the weight of the justification for the intended state of affairs and its level of feasibility. So an intention that is not even possible or minimally feasible would clearly be rationally defective. But an intention that is highly unlikely to be fulfilled might still be rational if its fulfillment is valuable enough to be worth a try. The feasibility principle, understood in this way, is actually in line with the central insight of rational choice theory. And this further strengthens my claim that the orthodox Humean picture of rationality already incorporates a rational principle that asks us to be sensitive to feasibility considerations and forbids us from having clearly infeasible intentions.[16]

Of course, one might still think that an intention that is not even minimally feasible can still be rational in itself. A smart mathematician, for instance, might intend to find out a difficult proof for a complicated theorem, and her search for it might turn out to be not even minimally feasible. But such potential infeasibility, as we might think, does not really render her intention and attempt irrational.

But I think we have this kind of intuition primarily because of two reasons. The

[16] Still, the feasibility principle is distinct from the rational principles codified in contemporary decision theory. After all, the former is grounded in the internal, constitutive aim of intention (or desire) as an intentional state that aims at its own efficacy, while the latter still largely concern the maximization of expected utility. Indeed, the feasibility principle doesn't seem applicable to preference, which lacks a built-in claim to its own efficacy. And this might explain why many contemporary Humeans have failed to notice how there is a feasibility constraint imposed on our practical states.

first, which I just mentioned, is that the infeasibility of an intention can be compensated by the enormous value involved in its fulfillment (say, the value in finding out the proof). The second reason is that many intentions that are not objectively feasible can still perfectly satisfy the subjective feasibility constraint. Our smart mathematician, for instance, might intend to find out the proof while lacking the belief that the search for it is simply infeasible. Indeed, if she has the explicit belief that it is simply impossible for her to find out the proof in the first place, then we certainly would have some difficulty in understanding the point and the rationality of her intention in doing so. Now given that many of us embrace a version of internalism about rationality, we would indeed tend to withhold our rational criticism when it comes to infeasible intentions that satisfy the subjective feasibility constraint. We have, however, just seen that the feasibility principle is a rational constitutive principle grounded in intention's executive function. And the very idea of proper function, as noted in Chapter 1, can be understood either in an information-sensitive or in an environment-sensitive sense. So if we are warranted to rationally criticize people for intending an end that they believe to be infeasible, then it should also make good sense to criticize them for intending an end that is simply infeasible as a matter of fact. After all, the intentions in both cases can be said to be defective in their executive function. The former aims to bring about a prospective state of affairs despite the information about its infeasibility, while latter simply aims to bring about an infeasible one.

Of course, if you think that objective infeasibility cannot constitute a *rational* defect for our intention, then I am happy to concede here that it only makes for a kind of *normative* defect for our intention in light of its executive function. The feasibility principle still provides a good rationale for Humeans to divest our

misinformed and infeasible intentions of their reason-giving power. And this principle still reveals how the very foundations of the orthodox Humean view already implicate a semi-substantive normative (though not rational) principle for our intention.

4.3. How Substantive is the Feasibility Principle?

You might think that the feasibility principle is still quite content-neutral and formal. After all, this principle only sets a kind of negative feasibility constraint on our intention. And it does not tell us how certain kinds of intention (say, the intention to commit immoral acts) might just be irrational in themselves, apart from infeasible and foolish ones. So how could the feasibility principle have any substantive implications on our first-order normative theories?

The defense of moral rationalism is beyond the scope of this thesis. But I think the feasibility principle does have more substantive implications on our first-order normative theories. First, this principle allows our human nature to set meaningful constraints on what we can rationally intend. One source for the infeasibility of our intention lies in the way things are, say, how there is no existing object that corresponds to the representational element in the content of our intention and that constitutes the realization thereof. And it is this source of infeasibility that makes our misinformed intention rationally defective. But another important source of infeasibility lies in the limitation to our capacities as human beings. For instance, an intention to run at light speed is infeasible given our physical constraint, while an intention to calculate every logical consequence of our beliefs is infeasible due to our psychological constitution. So our limitations as human beings, together with

the feasibility principle, can also determine whether an intention is rational and capable of being a legitimate source of normative reason for action for us. Such rational constraints from the limits of our capacities in turn highlight the importance of self-knowledge to our practical rationality: To avoid infeasible (and hence irrational) intentions, it is not enough to have full information about the external world. A sober understanding of the limits of our own capacities is also required.

Second, as our talk about the limit to our capacities here suggests, the rational defect of infeasibility can actually be dealt with by either revising our infeasible intention or improving our capacities pertaining to its realization. So in a way the feasibility account also shows how there are derivative rational or normative requirements on us to enhance our own capacities, and to rely on one another's agency through cooperation and reciprocity in our venture of changing the world in the ways that we envisage it. And both types of requirement can potentially provide the justification for morality. Perhaps its justification can be derived from the rational requirement on how we should enhance, and refrain from undermining, our own agency and that in other people (Smith 2015). Or perhaps its justification is instead derived from the rational requirements governing the cooperation among us as agents. So the feasibility principle should only be the beginning of our neo-Humean picture about the rationality of intention; it shouldn't be the final word.

5. Conclusion

We can draw three lessons from our discussion here. First, we can distinguish between liking-based and motivation-based subjectivism. And misinformation

presents different threats to these two kinds of subjectivism, for it can either make our telic desire unable to reflect our genuine likings or lead it to aim at states of affairs that are unable to be realized.

Second, the very conceptions of practical rationality and desire that underlie the orthodox Humean view already implicate a rational feasibility principle for desire. And orthodox Humeans have to invoke this principle in order to exclude misinformed and infeasible desires from the eligible subjective sources of normative reason. So they do need to appeal to a semi-substantive rational principle for criticizing a desire's content in the end.

Third, the feasibility principle applies most naturally to our intention or volitional state. This is because intention functions to lead us to act and to realize a prospective state of affairs in virtue of it. So the point and the typical justification of an intention lie in the realization of the prospective state of affairs intended. But infeasibility can simply nullify this very point and justification. That is why an intention that is not even minimally feasible is rationally criticizable.

The feasibility principle then casts doubt on the orthodox Humean picture about the rationality of intention. Indeed, although I argued in the foregoing chapter that there is an integration principle for our intention given its settling function, this rational principle is still quite formal in its nature. It still does not really set any substantive constraint on what we rationally should intend. However, as we have seen, there is also a feasibility principle for our intention given its executive function, which is actually a semi-substantive rational principle that forbids intentions with a particular type of content, i.e. infeasible content. And it can in turn support other derivative rational principles for enhancing our agential capacity and for mutual cooperation. So a more careful examination of the functional nature of our intention

actually shows that the rational principles for our intention are not just limited to the instrumental and coherence principles. We should therefore move beyond the orthodox Humean picture of practical rationality.

Appendix 1: Motivation-based versus Liking-based Subjectivism

You might wonder why subjectivists shouldn't just be content with liking-based subjectivism and the genuineness account. After all, if the most plausible version of subjectivism is one that grounds normativity on our likings for actual things as-they-really-are, then subjectivists just don't need to worry about the infeasibility problem that inflicts motivation-based subjectivism. Consequently, they probably won't have to invoke the feasibility account and the semi-substantive rational principle that I proposed in this chapter. Let me address this concern very quickly here.

To begin, motivation-based subjectivism actually represents the orthodox subjectivist view about normative reasons for action. Most contemporary subjectivists have tended to focus on the motivating dimension of desire and take it to be the very feature in virtue of which desire is the source of normative reason for action. For instance, what Bernard Williams appeals to in his account of normative reason are precisely the states in our subjective motivational set. And most of the motivating states that he considers involve a prospective object to be realized. When he discusses whether our normative beliefs can give rise to motivating states, he mostly formulates the object of our motivating state in terms of a prospective action to be performed in the future (1982: 108-9). Another central state in the subjective motivational set that he appeals to involves "projects, as they may be abstractly

called, embodying commitments of the agent (105)," but the idea of project also involves a prospective mental plan to be carried out. Michael Smith likewise takes desires to be the main source of normative reason, more specifically, the desires of our rational selves about what our actual selves should do. But for him, desire is just a motivating state with a representation of a state of affairs to be realized (1994: ch.4). He briefly considers whether normative reasons are somewhat relativized to our likings for different kinds of things. But he argues in the end that such idiosyncratic tastes or likings can actually be conceived as the features of the circumstances that our actual selves are in, and our idealized selves might not take such features into consideration when they form desires concerning what our actual selves should do (170).

More importantly, I think there are actually independent reasons for subjectivists to take motivation-based subjectivism seriously instead of completely deferring to our likings. First, liking-based subjectivism is in more danger of collapsing into a form of realism. Both Parfit and Thomas Scanlon, for instance, have contended that normative reasons grounded in our likings are actually objective hedonic reasons grounded on facts about what kinds of things are enjoyable for us (Parfit 2011a: 67; Scanlon 1998: 44-5). Parfit has even suggested that our likings admit of a more objectivized understanding, such that reasons grounded in them are more akin to objective rather than desire-based reasons (2011a: 52-6, 67). Of course, it is controversial whether our likings should be understood as subjective states, or whether they can be "objectified" into facts about our dispositions for feeling pleasures.[17] But motivation-based subjectivism is at least less vulnerable to the

[17] Railton also adopts this more objectivist understanding of our idiosyncratic psychological constitution and calls himself as a realist, see (1986a: 25-6; 1986b: 173-6). For an argument

potential problem of collapsing into realism because most philosophers can agree that motivating states are paradigm instances of subjective state.

Second, I think the normativity involved in good or well-being is best captured by liking-based subjectivism, whereas that involved in normative reason for action is better accommodated by motivation-based subjectivism. Well-being involves a receptive kind of normativity—even beings without rational agency can be said to have things that contribute to their well-being or things that are good for them. And it seems that a thing contributes to people's well-being primarily in virtue of how it possesses features that suit their likings and is able to bring about experiences that they enjoy. Indeed, the latter experiential aspect is particularly important when it comes to well-being (Griffin 1986: 16-20). Normative reason for action, on the other hand, involves an active kind of normativity—only beings with agential capacity and capable of effectuating changes in the world can have reasons for action. It seems that their telic desire gives rise to reasons for action primarily in virtue of how it, as a motivating state containing a prospective state of affairs to be realized, provides the typical input to their rational executive ability to carry out. On motivation-based subjectivism, normative reasons for action are grounded in our agential capacity for realizing the prospective thing that we favor. I believe this is the central line of thought that informs the subjectivist program about normative reasons for action and that is worth taking seriously.

---

concerning how this move of "objectifying the subjective" will still land one in a broadly subjectivist position, see Sobel (2012).

Chapter 4 In What Sense is Practical Rationality Internal?

1. Introduction

According to internalism about rationality, whether one is rational or not is solely determined by the internal relations among one's intentional states, and not by anything external to the totality of these states. This internalist view has received overwhelming support from contemporary philosophers despite their disagreement on the nature of rationality and the proper scope of rational requirements. So *Coherentists* believe that being rational consists in having coherent intentional states, and that one can be rational by revising any of the intentional states standing in incoherent relations. Thus, Coherentists typically construe rational requirements as wide-scope requirements governing the coherent patterns of our intentional states (Broome 2013a; Way 2010a, 2010b). *Enkratics*, on the other hand, believe that being rational consists in governing one's own mental life by one's beliefs about reasons. So most Enkratics contend that one can only be rational by revising one's intentional states in accordance with one's reason-beliefs, such that rational requirements should be best understood as narrow-scope requirements prescribing the conformity of our intentional states to our reason-beliefs (Kolodny 2005; Scanlon 1998, 2007; cf. Schroeder 2009)[1]. In spite of their differences, both Coherentists and

[1] Here I am only using the terms "Coherentists" and "Enkratics" as two convenient labels to refer to the proponents of these two general views about rationality. They should not be taken to imply that Coherentists do not accept the Enkratic requirement, or that Enkratics do not care about coherence requirements at all. They do. It is just that they have different views about the nature and the grounding of these requirements.

Enkratics agree that rationality is about having intentional states standing in proper patterns, and both commit themselves to internalism. It's just that for Coherentists, a proper internal pattern consists in the coherence among one's intentional states, while for Enkratics it is the conformity of one's intentional states to one's reason-beliefs. (For the sake of convenience, in what follows I will use "proper relations" and "proper patterns" to cover both kinds of pattern here.)

In fact, internalism about rationality has become the working assumption for almost all contemporary philosophers working on rationality. Indeed, most of them have embraced a thesis directly entailed by internalism, i.e. the supervenience of rationality on the mind. After all, if rationality is only about maintaining proper intentional state patterns, then for sure two agents with indistinguishable mental patterns would also be indistinguishable with respect to their rationality. Most contemporary philosophers have in turn treated this thesis as a guiding principle for theorizing about rationality. For instance, some have appealed to the supervenience thesis and argued that the instrumental requirement can only prescribe us to intend the actions that *we believe* to be able to realize our intended ends, rather than to intend the actions that can *in fact* realize them (Brunero 2012; Broome 2007). In this way, internalism has greatly influenced how philosophers understand the nature and the precise formulation of the instrumental requirement, and, by extension, other more specific rational requirements.

But is being rational only a matter of having intentional states standing in proper relations? I believe we have a good reason to doubt this, at least when it comes to practical rationality. Indeed, once we examine the precise senses in which practical rationality can be said to be internal—as well as the different conceptions of practical rationality that they entail—we will see how practical rationality is not

just about having proper intentional state patterns on the superficial level; it is ultimately about the proper function of our agency.

So in what sense can rationality be called internal? It might be said to be internal in a "source" sense. That is, we might say that truths about what rationality requires of an agent are ultimately grounded in what seems right from her subjective point of view,[2] including which intentional state patterns seem right and proper from her own perspective. However, this *radical internalist conception of rationality*, I argue, is problematic. This is because the requirements of rationality have a nature that is not of the agent's own making, so truths about what they require of her cannot wholly be determined by what the agent herself thinks. Accordingly, rationality cannot be internal in a "source" sense (Section 2).

Alternatively, rationality might be said to be internal in an epistemic sense. We might say that rational requirements are necessary, *a priori* requirements that prescribe the coherent patterns among one's intentional states. And although truths about what rationality requires of the agent don't have their sources in the agent's own subjective perspective, they are still something to which the agent has special internal epistemic access, given their necessary, *a priori* status. However, this *rationalist conception of rationality*, I contend, is implausible when it comes to practical rationality. This is because the requirements of practical rationality are, contrary to what many philosophers have claimed, in fact contingent. What the instrumental requirement prescribes, for instance, is determined by what intentional state pattern is indispensably involved in our general intention-fulfillment process. This is in turn contingent on concrete empirical

[2] Here I follow Chang's understanding of the source question (2013: 164-6). To say that rationality is internal in a "source" sense is to say that facts about what rationality requires of an agent are grounded in facts about what it seems to the agent from her point of view. It is the latter facts that provide a metaphysically necessary explanation for the former.

features of our intentional-fulfillment mechanism, as well as the distinctive evolutionary history that shaped it. So the requirements of practical rationality, I argue, are not really necessary, *a priori* requirements. Practical rationality is hence not internal in an epistemic sense (Section 3).

I instead contend that practical rationality can be said to be internal only in a *constitutive* sense. On my view, being practically rational consists in having the mental patterns that are constitutively involved in the proper function of our agency: So a certain intentional state pattern is instrumentally rational just because it is indispensable for the proper function of our executive mechanism in fulfilling our intentions; a pattern of non-contradictory intentions is rational in virtue of its contribution to the proper function of our decision-making mechanism in settling us on a particular course of action; finally, a conformity pattern of intention to normative judgment is rational just because it realizes a form of accountable normative self-governance. Thus, truths about what mental patterns are proper, and what practical rationality requires of us, are in fact grounded in what the proper function of our agency entails (Section 4).

I will then explain why this functionalist picture presents the most plausible conception of practical rationality and consider its implications on our general understanding of practical rationality (Section 5). As I will elaborate, this conception is significant because it reveals how external environmental factors play a role in determining which mental patterns count as practically rational. It therefore casts doubt on the stronger, rationalist variant of internalism and thereby lays the foundation for a more empirically-minded picture of practical rationality.

2. Rationality and the Internal Perspective

2.1. The Radical Internalist Conception of Rationality

Let us start by considering the following case:

> *Practical Incoherence:* The agent intends an end E, believes that φ-ing is the necessary means to bring about E, but then intends not to φ.

Why is this case problematic and why does it deserve the criticism of irrationality? According to an intuition shared by many internalists, this is because the charge of irrationality involves a kind of harsh, negative criticism, so irrationality is supposed to be an internal problem that the agent would find problematic simply from her own subjective perspective. Now the problem of practical incoherence here indeed has this internally-identifiable character: There is a tension among the agent's intentions and means-end belief. And such a tension is something that she will find problematic if she views the matter from the totality of her intentional states alone. If she reflects carefully on what she intends and believes, then she will inevitably see how her intention to not φ undermines her intention to bring about E, given how she also believes that φ-ing is indispensable for the realization of E.

This intuition about the internally-identifiable character of irrationality is taken seriously by Enkratics in particular, who conceive of irrationality as a glaring problem that is readily identifiable from the inside, i.e. the problem of having intentions and mental states that simply fail to meet one's own normative beliefs about what actions one has reason to do, and which intentional states one has reason to have (Kolodny 2005: 557-60; Scanlon 1998: 25-30, 2007). Other

internalists who don't subscribe to the Enkratic view about the nature of rationality also share this kind of internalist intuition. Clayton Littlejohn, for instance, has remarked that if an intentional state S is irrational, then the formation of S is usually something that cannot make sense to the agent herself given the intentional states she already has (2018: 262-3). Similarly, Kieran Setiya (2007: 660, n.25) has also noted that failures of practical rationality are usually those that the "agent herself can always recognize as such", and that she is "in a position to avoid" (see also Parfit 2011a: ch.16).

This internally-identifiable character of irrationality is also evidenced by how mental compartmentalization can often extenuate the problem of irrationality. For instance, the irrationality involved in an agent's contradictory beliefs can be mitigated somewhat if they belong to separate subdivisions in her mind, with different clusters of intentional states supporting each of them in each subdivision. Now the best explanation as to why an agent's irrationality can be extenuated in this way, I think, is that mental compartmentalization confines her contradictory beliefs to each of these separate inner-divisions and thereby renders their contradiction less internally-identifiable.[3] This in turn provides support for the idea that the problem of irrationality has an internally identifiable character, a character which makes practical incoherence a paradigm case of irrationality.

So far, so good. But now suppose our agent with practical incoherence comes to form a radical belief about rational requirements. Suppose her guru tells her that she can get whatever she wants if she "wants it bad enough". Not fully understanding what this means, our agent is misled into believing that what instrumental

[3] As we saw in Chapter 2, this is Davidson's proposal for accommodating irrationality within an overall rational framework, see Davidson (1985a).

rationality prescribes us to do is to intend our ends very strongly, and that it doesn't really forbid us from both intending an end and intending not to take the necessary means to it—after all, what matters is wanting it badly enough.

> *Practical Incoherence II:* The agent intends an end E, believes that φ-ing is the necessary means to bring about E, but then intends not to φ. However, she also believes that instrumental rationality only requires us to intend our ends very strongly, and that it doesn't forbid us from both intending an end E and intending not to take the necessary means to it. So she doesn't perceive any tension between her intentions.

Is this case also problematic? Does it also deserve the criticism of irrationality? I take it that there are two possible verdicts we can give. One verdict is that the agent in this case is irrational because her conflicting intentions violate the instrumental requirement. The other verdict is that this agent is in fact rational despite her intentions, given how she believes that rationality doesn't actually forbid this kind of intentional state pattern, and how she simply doesn't regard these two intentions as standing in a conflicting relation.

The latter verdict, I think, is what we would give if we take internalism to its extreme: If being rational is simply a matter of having intentional states standing in proper relations, then we might think that rationality in the end only consists in having intentional states standing in the relations that *one believes* to be proper. After all, in doing so, one is holding intentional states in a way that conforms to, and coheres with, one's idiosyncratic beliefs about the proper relations that our intentional states are supposed to stand in. So one's intentional state pattern as a

whole can still be said to be internally coherent and rational. As we saw in Chapter 2, this *radical internalist conception of rationality* is actually endorsed by Davidson in one of his later papers (1982). Indeed, this conception is something that the Enkratic view about the nature of rationality would in particular lend strong support to: After all, if being rational consists in conforming to one's normative beliefs about what individual intentional states one has reason to have, then it would be a natural further step to claim that being rational also consists in conforming to one's normative beliefs about what intentional state *patterns* one has reason to maintain.

2.2. The Independence Thesis

The radical internalist conception of rationality, however, is too permissive. This is because it implies that almost every incoherent pattern of intentional states can be potentially exonerated as proper and rational, provided that one has a bizarre normative belief about how the pattern in question is in fact proper, coherent, and thereby required by rationality. Indeed, as Littlejohn (2018: 264) has pointed out, this kind of view entails a form of anarchism about rationality, for, on this account, people can simply be rational by conforming to their own idiosyncratic beliefs about what coherence amounts to or what rationality requires of us. That is why many Coherentists also find radical internalism overly permissive.

Another problem with radical internalism is that even this view has to presuppose that there is one particular rational requirement that applies to the agent independently of what she believes. This is the higher-order Enkratic requirement on the match between the agent's intentional state patterns and her

normative beliefs about which intentional state patterns she has reason to maintain. That is, even those who embrace radical internalism have to say that it is rational for an agent's intentional state patterns to conform to, or to cohere with, her normative beliefs about proper patterns—and this is so regardless of whether she endorses or rejects the higher-order Enkratic requirement itself (Broome 2013a: 93-4). Once we admit this, however, there will be no strong reason for us to deny that there are other rational requirements that apply to the agent independently of what she believes, and that some mental patterns count as rational just in virtue of their meeting these independent requirements, regardless of how the agent views such patterns from her subjective perspective.

In fact, due to these problems with radical internalism, some Coherentists have even embraced the so-called "fixed-point thesis", according to which our false beliefs about rational requirements are themselves irrational from the perspective of rationality, even if they might be otherwise well-supported by our beliefs (Littlejohn 2018; Titelbaum 2015). Here I am less interested in the rational status of our false beliefs about rational requirements. Rather, my central concern is the rational status of an intentional state pattern that is forbidden by a rational requirement yet still deemed to be rational by the agent due to her false beliefs about rationality (which is a distinct issue that has not been paid enough attention to). I am more interested in this because I think this issue has important implications on our understanding of the nature of rationality. Indeed, in light of our earlier arguments against radical internalism, we can draw the following general thesis about the nature of rationality:

The Independence of Rationality: Whether a particular intentional state

pattern is proper and rational is not determined by what seems right from the agent's subjective perspective, but by certain requirements of rationality that hold independently of what she thinks about them.

Given the independence thesis, we should say that our agent in the second case is irrational simply because her intentions violate the instrumental requirement, and that she cannot let herself off the hook by her bizarre and mistaken belief about instrumental rationality.[4]

The chief lesson to be drawn from the independence thesis is that truths about what rationality prescribes the agent to do are not wholly grounded in her subjective perspective, as constituted by the totality of her intentional states (which importantly includes her idiosyncratic beliefs about which intentional state patterns are proper and rational). To put this in more metaphorical terms, truths about what rationality requires of us don't have their sources within the totality of our intentional states. Rather, these truths hold independently of what we think and have a nature that is not of our own making. This then discredits the radical internalist conception of rationality and shows that rationality is not internal in the source sense.

3. Rationality and Coherence

3.1. The Rationalist Conception of Rationality

[4] As we saw in Chapter 2, Davidson eventually abandons radical internalism and instead embraces the independence thesis (1985a: 196-7). Broome (2013a: 92) and Littlejohn (2018: 269-70) also accept this thesis, but they formulate the idea in terms of how rational requirements impose "strict liability" on us.

Of course, the independence thesis does not entirely undermine the internalist doctrine. Indeed, those who offer arguments in support of the independence thesis, such as Donald Davidson (1985a), John Broome (2013a), and Michael Titelbaum (2015), still wholeheartedly embrace internalism about rationality even though they think that we cannot always give deference to the agent's subjective perspective. But why would they think that rationality still has an internal character despite the independence thesis? This is because they seem to believe that although truths about what rationality requires of us hold independently of what we think, such truths are still more internally accessible to us. In fact, many of philosophers working on rationality have followed rationalists in the Early Modern period and conceptualized such truths as necessary, *a priori* truths that we have special internal access to. So, for instance, the modal operator "necessarily" appears in Broome's formulations of all the rational requirements that he examines in his work (2013a), suggesting that he thinks of these requirements as necessary requirements applying to all possible worlds. Jonathan Way similarly contends that truths about what rational requirements demand of us are necessary, *a priori* truths (2010b: 1065). Titelbaum even explicitly argues that we have "a priori, propositional justification for true beliefs about the requirements of rationality (2015: 276)." (Indeed, it is partly because of this that Titelbaum takes false beliefs about rational requirements to be irrational in themselves.)

This *rationalist conception of rationality* seems a lot more plausible than the radical internalist alternative. Most rational requirements, such as the requirement of non-contradiction, Modus Ponens, and means-end coherence, seem to be requirements applying to all agents in all possible worlds, requirements that they can come to know without recourse to their empirical knowledge about the world

they are in. Indeed, this conception comports particularly well with the Coherentist view about the nature of rationality. This is because for most Coherentists, coherent intentional state patterns are supposed to be coherent solely in virtue of the internal relations among the intentional states involved. So presumably, rational requirements would prescribe all agents in all possible worlds to maintain these patterns (which are just coherent in and of themselves), regardless of the differences in their external environments. That is why rational requirements are necessary requirements to which we have special, *a priori* access. So I presume most internalists would still hold that the problem of irrationality has an ineliminable internal epistemic character, and that it is this epistemic character that explains why the cases of practical incoherence that we considered earlier deserve the damning label of irrationality. After all, our agent with practical incoherence is still in a good position to know, in an *a priori* manner, how her conflicting intentions are in fact forbidden by rationality.

3.2. The Contingency of Practical Rationality

But I believe rationality can neither be said to have an internal character on epistemic grounds, for the rationalist conception of rationality is not very plausible when it comes to practical rationality. Why? This is because the most central requirements of practical rationality are actually contingent on empirical features of our underlying psychological mechanisms and the world we are in. Moreover, our knowledge of the precise content and the stringency of these requirements also needs to presuppose our knowledge of such empirical features. Let me elaborate.

(1) The Contingency of the Means-End Coherence Requirement

In Chapter 1, I argued in favor of the executive conception of IR. There, we saw that on the executive conception, what the instrumental requirement prescribes is contingent on empirical features of our rational executive system for intention-fulfillment, that is, on what psychological steps are constitutively involved in the proper function of our rational executive system. Of course, most internalists about rationality probably will not find this conception of IR congenial. So let us focus instead on the requirement of means-end coherence, which, as I recognized earlier, is still a valid requirement of practical coherence (though as I suggested in Chapter 1, such a means-end coherence requirement cannot be identified with the instrumental requirement, which has a distinctive executive dimension that cannot be captured by the idea of coherence). Now the rationale for the means-end coherence requirement is that intending an end but not the necessary means to it leads to inconsistency in our attitudes: Our intending to bring about a prospective state of affairs S carries our commitment to its realization, but our not intending the necessary means to S involves our non-commitment to its eventual realization. So in order to restore our coherence, rationality will require us to either form the instrumental intention for the necessary means to S, or to abandon the telic intention for the realization of S in the first place.

However, as explained in Chapter 1, there might be a possible world in which our bodies are so cumbersome that the psycho-physiological process for initiating our movements becomes very costly in terms of the time and the general psychological resources required. Evolution in this world then favorably selected a kind of executive system with a "confirming mechanism" to avoid the waste of

resources: To fulfill a telic intention, we have to first form the instrumental intentions to take the necessary means. But in order for these instrumental intentions to lead us to act, we have to form a higher-order approving desire that we want this very telic intention to be realized by triggering the costly process for initiating our movement. In this world, it is means-end incoherent for us to intend to bring about a state of affairs S without forming a higher-order approving desire: In intending to bring about S, we commit ourselves to its realization; yet in failing to form the higher-order approving desire, we also fail to commit ourselves to its eventual realization. The very same pattern of intentional states, however, is not means-end incoherent in our actual world—even if we fail to form a relevant higher-order approving desire, this still won't compromise our commitment to S's realization. So presumably the means-end coherence requirement would forbid [having a telic intention without a higher-order approving desire] only in this other possible world but not in our actual world.

More generally, the means-end coherence requirement forbids us from [having a telic intention while lacking the intentional states that are constitutively indispensable for our general intention-fulfillment]. Yet which intentional states are indispensable in this way depends on the empirical features of our psychological mechanism for fulfilling our intentions, as well as its external evolutionary history. That is why the precise content of this requirement is contingent.

Can we say that what the means-end coherent requirement forbids is [having a telic intention while failing to intend the necessary means to one's intention-fulfillment], and that the contingency that I just highlighted only constitutes the contingency in the *mental means* we need to intend? No. This is because we should not construe the intentional states that are indispensable for our

general intention-fulfillment as a kind of mental means. For one thing, the instrumental requirement and the means-end coherence requirement don't prescribe these states *qua* a mental means we need to intend. For instance, if the agents in our actual world are to be instrumentally rational and means-end coherent, then what they need to do is to just form the relevant instrumental intentions. They don't need to have a prior, higher-order intention to form these instrumental intentions as a mental means they need to take. For another, as we saw in Chapter 1, once we construe the intentional states indispensable for our general intention-fulfillment as a mental means, we would allow people's idiosyncratic beliefs about the relevant mental means to have a say in determining whether they are instrumentally rational or means-end coherent. That is, their instrumental rationality and means-end coherence would no longer be determined by whether they in fact have these indispensable intentional states (say, instrumental intentions or higher-order approving desires), but rather by whether they have taken the mental means that they believe to contribute to their intention-fulfillment. So I think we should say that the means-end coherence requirement simply forbids [having a telic intention while lacking the intentional states indispensable for one's general intention fulfillment]. And we should not construe the contingency in these indispensable states as the contingency in the general mental means we need to intend and take, for these are not mere means in the first place.

(2) The Contingency of the Requirement of Practical Non-Contradiction

Now let us consider the requirement of non-contradiction in our practical attitudes. Of course, this requirement is derived from one of the most fundamental logical

truths, so I don't think I can persuade you that its content can ever be contingent. But I think the *stringency* of this requirement can at least be shown to be contingent on our underlying psychology.

When contemporary philosophers talk about the requirement of practical non-contradiction, they have almost exclusively focused on the prohibition against contradictory intentions (Bratman 1987; Broome 2013a: ch. 9; Kolodny 2008a). This, I presume, is because the requirement of practical non-contradiction seems to apply more stringently to intention than to desire. Indeed, for an agent who says, "I desire to go to Mark's party tomorrow but I also desire to not go to Mark's party tomorrow", her contradictory desires can often seem understandable or even rationally uncriticizable. Perhaps there are both pros and cons for going to Mark's party tomorrow, and her contradictory desires here might simply be responding to these conflicting considerations in a right way. However, if our agent says, "I intend to go to Mark's party tomorrow but I also intend to not go to Mark's party tomorrow", then her contradictory intentions would clearly baffle us and constitute a clear case of irrationality.

So why do we have a more stringent rational prohibition against contradictory intentions? One potential explanation is that contradictory intentions entail inconsistent beliefs about our future actions,[5] which constitute a more paradigm case of irrationality than the mere conflict in our desires or practical attitudes. And it is the cognitive requirement against belief inconsistency that explains why we have a stricter rational prohibition against contradictory intentions (see Wallace 2001; Setiya 2007). However, we have already seen in Chapter 1 why this kind of

[5] Or at least they entail inconsistent beliefs about the possibility of the realization of our intended state of affairs.

cognitivist account about practical rationality is implausible. Our agent, for instance, might intend to go to Mark's party and to not go to Mark's party at the same time while believing that she is going to perform only one of these actions in the future—perhaps she just thinks that the conflict between her intentions will be magically resolved by God before she sets out to act. Now despite having no inconsistent beliefs on her part, our agent's contradictory intentions still seem rationally problematic.[6] Indeed, they are still rationally more condemnable than merely conflicting desires are. So the rational problem with her contradictory intentions cannot be wholly explained by the belief inconsistency that they might lead to.

A better explanation as to why we have a more stringent rational prohibition against contradictory intentions, I think, is that these intentions have a mutually-undermining effect. So our agent's contradictory intentions to go to Mark's party and to not go to Mark's party will tend to cancel each other out and make her end up intending nothing at all. Her contradictory desires, in contrast, can often co-exist with each other despite their conflict. And it is presumably the mutually-undermining effect of contradictory intentions—a kind of effect that contradictory beliefs also share—that explains why the requirement of non-contradiction applies more stringently to intentions than to desires.

As I argued in Chapter 2, it is the functional characters of intention and desire that explain the differences in their mutually-undermining effect. Intention functions to settle our practical question and to represent our final answer regarding what we are to do in the future. And contradictory intentions will tend to

[6] For related arguments against cognitivism about practical rationality, see Bratman (2009a) and Broome(2013b: 163-66).

prevent each other from playing this defining settling function, thereby undermining each other's status *qua* intention: If you intend to φ, then this is supposed to signify that you have settled your practical question in favor of φ-ing in the future, However, if you also intend not to φ, then this means that you have not really settled your practical question in favor of φ-ing. In contrast, desire does not function to settle our practical question. Rather, it only acts as a potential influencer in our practical deliberation on what we are to do in the future (Bratman 1987: 15-6). That is why contradictory desires tend not to undercut each other's status *qua* desire.

How, then, can the stringency of the requirement of practical non-contradiction be contingent at all? Consider a possible world in which we settle our practical question by our intentions in quite a different way. Suppose that in this possible world, the general resources for carrying out our intentions are so scarce that the evolutionary history there has favorably selected a more conservative decision-making mechanism, one that prioritizes stability over the ease of changing one's mind: Once we intend to φ and settle ourselves on this course of action, there will be presumptive support in favor of carrying it out. And a subsequent, contradictory intention to not φ will be able to overturn our settledness only if its strength greatly exceeds that of the earlier intention. Now in a world like this, our contradictory intentions are far less prone to the problem of mutual-undermining. We can consciously hold an earlier intention to φ and a later intention to not φ at the same time, and, owing to the presumptive support for stability in this world, only the settling force of our later intention will be compromised. And our earlier intention in the contradictory pair can still act as the default intention and perfectly play its functional role in settling our practical question and in leading us to act. So in this world the contradiction in our intentions has less of a mutually-undermining

effect and won't really leave us intending nothing genuinely in the end. And I would say that holding contradictory intentions in this world will indeed constitute a less serious form of rational failure than it does in our actual world.

More generally, if you accept the idea that the rational prohibition against contradictory intentions is more stringent than the one against contradictory desires due to the former's mutually-undermining effect, then you should also think that the stringency of the requirement of non-contradiction in intention can also be contingent. This is because whether the contradiction in our intentions can really have a mutually-undermining effect depends on our underlying decision-making mechanism, the functional profile of our intention, and the external evolutionary history that shaped it.

(3) The Contingency of the Enkratic Requirement

Finally consider the Enkratic requirement. For ease of discussion, I will focus on the Enkratic requirement regarding the proper relation between our intentions and normative judgments, that is, judgments about what actions we have a reason to take. Still, I think the argument to follow should equally apply to the broader Enkratic requirement regarding the proper relation between our intentional states and our normative judgments about which intentional states we have a reason to possess. The rationale for the Enkratic requirement is that our normative judgment has a special authority among our intentional states, partly owing to how its content is about what we have a normative reason to do, and partly owing to how it is sensitive to the rational influence of evidence and is capable of being criticized as false or unjustified. Indeed, it can be said that those who form and revise their

intentions and intentional states in accordance with their normative judgments realize a distinctive form of accountable normative self-governance, since they are in effect governing their mental life with normative judgments that are answerable to rational criticism.[7]

To capture the authority of normative judgment and the importance of accountable normative self-governance, Enkratics have conceptualized the Enkratic requirement as a narrow-scope requirement, which can only be satisfied by intending the actions that we take to be supported by reason, and not by revising our reasons-judgments (Kolodny 2005; Scanlon 2007; Schroeder 2004, 2009). Wide-scopers, in contrast, have taken the Enkratic requirement to be a wide-scope coherence requirement, which can be equally satisfied by revising our normative judgments. But even wide-scopers have tried to accommodate the special authority of normative judgments by alternatively appealing to the so-called basing-prohibition, that is, the prohibition against revising our normative judgment on the basis of what we intend to do (Broome 2013a: ch. 10; also see Way 2011: 232-4).

So how can the Enkratic requirement be contingent? Suppose we are in a world in which we make normative judgments to ourselves through a kind of overt or inner speech. Further suppose that in this world, our normative judgment always aligns with our motivation without exception; it is less prone to change even with the presence of new information; and the very same brain area is activated when we make a normative judgment and when we entertain a desire for something. It turns out that in this world, our normative judgment that we have normative reason to φ

[7] Notice that this idea of *accountable* normative self-governance is different from the idea of self-governance proposed by Bratman (2018b), since the former form of self-governance further requires that we intend in accordance with evaluative states that are sensitive to the rational influence of evidence and that are answerable to rational criticism.

expresses nothing more than our desire to φ. Now I would say that in this world the Enkratic requirement is more plausibly a wide-scope coherence requirement than a narrow-scope one. After all, in this world, our normative judgment only expresses a mere desire at bottom, which is neither sensitive to the rational influence of evidence nor subject to the rational criticism of being false or unjustified. So intending in conformity with such a judgment cannot really realize a form of accountable normative self-governance. Thus, in this world, it should be perfectly rational for us to revise our normative judgment that we have a reason to φ on the mere basis of our lacking the intention to φ, since in doing so we are in effect restoring the coherence between our intention on the one hand, and our apparent normative judgment/desire on the other hand. But surely things would be different in a world where our normative judgments express genuine cognitive states.

More generally, I think the psychological underpinnings of our normative judgments are able to affect whether following them can truly realize a form of accountable normative self-governance (In fact, this is why non-cognitivism is philosophically important). And this can in turn determine whether the Enkratic requirement is more plausibly a wide-scope or narrow-scope requirement. That is why the character of the Enkratic requirement can also be contingent on empirical features.

One might object that whether our normative judgments in fact express beliefs or desires is an issue irrelevant from the perspective of rationality. After all, what is relevant to rationality, as you might think, is *what it seems to us* from our subjective perspective. So as long as we make a normative judgment, and as long as *it seems to us that we believe* how there is a reason to φ, there will be subjective normative support for φ-ing. Rationality will then just impose a narrow-scope Enkratic

requirement on us—and this is so regardless of whether our inner normative judgment here actually expresses a belief or not. Thus, you might think that the psychological underpinnings of our normative judgments cannot really affect the character of the Enkratic requirement; it is rather our own subjective point of view that matters.

But I think this view only brings us back to radical internalism, and it is subject to the same kind of problem. An agent, for instance, might have an intentional state that functions just like a belief: She would readily assent to its content when asked; she would rely on it in her theoretical reasoning; its revision is sensitive to factual evidence, etc. Now suppose this belief stands in an incoherent relation to other beliefs that she has. Can she then plead innocent by saying that from her subjective perspective it merely seems to her that she is only supposing (rather than believing) that something is the case? Presumably not. As this case suggests, what rationality requires of an agent should hinge on the intentional states that *she in fact has*, as determined by their objective functional character, and not by the intentional states that *she takes herself to have*. This independence from the agent's own identification of her own mental states can be said to follow from the independence thesis considered in the last section: Just as the agent's mistaken views about rational requirements cannot change the rational status of her intentional states, nor can her mistaken views about her own intentional states affect her rationality. Thus, given rationality's independence from our identification of our own intentional states, the character of the Enkratic requirement is also supposed to be sensitive to what intentional states that our normative judgments *in fact* express, and not to what intentional states that we take them to express.

Of course, you might complain that what is at issue here is not our own views

about the psychological nature of our normative judgments, but our judging ourselves to have a reason. Indeed, the central tenet of the Enkratic view about the nature of rationality is that rational requirements are grounded in the demands of *the reasons that we judge ourselves to have* from our first-person perspective (see Kolodny 2005, 2008b; Scanlon 1998, 2007): as long as we judge that there is a reason for us to do something, rationality will just ask us to intend in accordance with the reason that we judge ourselves to have, thereby imposing a narrow-scope Enkratic requirement on us—and this is so regardless of the psychological underpinnings of our reasons-judgments. Thus, as you might think, if we take this Enkratic line of thought seriously, then perhaps we should still say that the Enkratic requirement necessarily has a narrow-scoped character.

However, I find this view dangerous. Does the Enkratic requirement always ask us to conform to the reasons that we judge ourselves to have, even in worlds where our reasons-judgments are grounded in non-cognitive states wholly insensitive to evidence, subject to manipulation, and distorted by prejudice? This would certainly make the Enkratic requirement normatively bankrupt. It would be more sensible, I think, to say that, in these worlds, the Enkratic requirement would be a wide-scope requirement, which also recommends the revision of our reasons-judgments as a way of reaching coherence. In other words, if the Enkratic requirement is to have objective normative authority at all, then it should give deference to the agent's reasons-judgments and have a narrow-scoped character *only in worlds where these judgments are grounded in the right kinds of intentional states*. That is why I think the Enkratic requirement has to be sensitive to, and contingent on, the actual psychological underpinnings of our normative judgments.

3.3. The Contingency Thesis

In light of our discussion thus far, we can derive another thesis about the nature of practical rationality:

> **The Contingency of Practical Rationality**: The content, stringency, and character of the requirements of practical rationality are contingent on empirical features of our psychology and the world we are in.

Although seemingly controversial, what underlies and supports this contingency thesis is the rather innocuous idea that if we had different psychological makeups and were in different environments, then we would be subject to different requirements of practical rationality (henceforth: p-requirements). Indeed, as we have seen, the requirements of instrumental rationality, means-end coherence, practical non-contradiction, and Enkrasia are all—to varying degrees—contingent on empirical features of our underlying psychology. This is quite important, for these requirements are often taken to be the most fundamental p-requirements. Given this, we cannot simply shrug off the contingency thesis as a thesis that only applies to certain peripheral p-requirements.

Of course, one might think that our p-requirements would not really be contingent if they were formulated in more abstract terms. So we can say p-requirements prescribe us to *take the psychological steps indispensable for our intention-fulfillment*, to *avoid self-undermining in our practical attitudes*, and to *revise our practical attitudes in accordance with cognitive states with normative content*. This seems to be the case across all possible worlds. Thus, abstract p-requirements

like these, as you might think, can still claim to have a fixed, non-contingent character.

Now I would be happy to concede that abstract p-requirements are indeed necessary requirements with fixed, non-contingent character. But what about specific p-requirements that we just considered, such as the requirements to form both instrumental intentions and higher-order approving desires, or to follow the normative judgments that we make to ourselves (as opposed to following only our cognitive states with normative content)? If we take these specific p-requirements to be rational requirements as well, then their contingency will suffice to establish the contingency thesis: Our p-requirements, except abstract ones, are still contingent on empirical features of our underlying psychology and the world we are in.

Of course, one could reply that we shouldn't think of these specific p-requirements as rational requirements in and of themselves. Rather, they are merely secondary principles that inform us of the general ways to fulfill abstract p-requirements. And their contingency, as you might contend, only shows how there are different ways for us to satisfy abstract p-requirements in different worlds.

But I think this way of conceptualizing specific p-requirements is misguided. For one thing, this introduces a needlessly complicated structure of rational norms. The requirement of forming instrumental intentions, for instance, is not a secondary principle that tells us how to conform to an even more abstract instrumental requirement. Rather, it is simply the instrumental requirement in our actual world. For another, this construal of specific p-requirements potentially brings us back to radical internalism. After all, if specific p-requirements are only secondary principles that suggest the general ways for us to fulfill some even more abstract

p-requirements, then the agent's idiosyncratic means-end beliefs would be able to intervene. What should we say if the agent believes that there are other better ways for her to fulfill abstract p-requirements? If we say that the rational thing for her to do is to follow her own means-end beliefs, then we will return to radical internalism. Alternatively, if we say that the rational thing to do is to follow specific p-requirements, no matter what the agent believes, then we are in effect recognizing the rational status of these specific requirements. And their contingency would still suffice to establish the contingency of what practical rationality requires of us. Needless to say, I think we should accept the contingency thesis instead of returning to the implausible radical internalist picture.

The contingency thesis then casts doubt on the rationalist conception of rationality and thereby shows that practical rationality is not entirely internal in the epistemic sense: Truths about what p-requirements require of us are not necessary truths. Also, our knowledge of them presupposes knowledge of the empirical features in the world. To know what the instrumental and means-end coherence requirements prescribe us to do, we have to know what our psychological mechanism for intention-fulfillment is like; to know the stringency of the requirement of non-contradiction on a particular type of intentional state, we have to know whether the contradiction in that type of state can really have a mutually-undermining effect; finally, to know whether the Enkratic requirement is a narrow- or wide-scope requirement, we have to know the psychological underpinnings of our normative judgments.

4. The Functionalist Conception of Practical Rationality

4.1. Practical Rationality and the Function of Agency

While the arguments that I just gave are supposed to show how p-requirements are contingent, what they actually suggest is that p-requirements are grounded in the proper function of our rational agency. Let me explain it.

Our rational agency minimally includes a decision-making mechanism, which allows us to settle our practical question in favor of a particular course of action. Now in our actual world, contradictory intentions or volitional states will tend to overturn each other's settling force and make us end up settling on nothing. This, however, is not the case for contradictory desires, which only present competing sources of influence in our decision-making process but don't usually prevent us from settling on a final course of action in the end. That is why the rational requirement of practical non-contradiction applies more stringently to intention than to desire. Indeed, in a possible world where our earlier intention is able to settle us on a particular course of action and is impervious to the undermining effects of a later contradictory intention, our contradictory intentions there would seem rationally more pardonable than the mutually-undermining contradictory intentions in our actual world. This then suggests that we have a more stringent rational prohibition against contradictory intentions not because of the formal inconsistency that it involves (for contradictory desires can also be said to involve a kind of formal inconsistency), nor because of the inconsistent beliefs that they entail (for this doesn't explain why contradictory intentions that do not entail inconsistent beliefs on our part are still rationally problematic), but mainly because of their undermining effects on the proper function of our decision-making mechanism.

Our rational agency further includes an executive mechanism, which helps us to

set out to realize a prospective state of affairs that we have intended to bring about and settled ourselves on. Now in our actual world, the fulfillment of our intentions requires the formation of further instrumental intentions in light of our means-end beliefs, while it furthermore requires the formation of higher-order approving desires in the other possible world just considered. Presumably, the instrumental requirement would prescribe us to form instrumental intentions in our actual world, and to form both instrumental intentions and higher-order approving desires in the other possible world. The means-end coherence requirement would similarly forbid [having a telic intention without instrumental intentions] in one world, and [having a telic intention without either instrumental intentions or higher-order approving desires] in the other. This then shows that we have an instrumental requirement prescribing certain psychological steps (or a means-end coherence requirement calling for certain psychological patterns) just because these steps or patterns are constitutively involved in the proper function of our executive mechanism for fulfilling our intentions.

Finally, on a more substantive picture of rational agency, our agency also involves a form of accountable normative self-governance, which consists in regulating our intentions by a kind of intentional state with normative content and answerable to rational criticism.[8] Now in a world where our normative judgments express intentional states that function just like beliefs, practical rationality would indeed give deference to them, such that the Enkratic requirement would have a narrow-scoped character (or such that there would be a relevant basing-prohibition against changing our normative judgments on the basis of our intentions). But in a

[8] Although Kolodny does not fully accept constitutivism about rationality, he still thinks that a disposition to follow one's reason-beliefs is an indispensable part of our agential capacity for being guided by reasons (2008b: 457-8), also see Scanlon (1998: 22-4).

world where our normative judgments turn out to express mere desires, which are insensitive to factual evidence and unanswerable to rational criticism, the Enkratic requirement would be more plausibly a wide-scope coherence requirement that confers no rational primacy to them. Again, we have a narrow-scope Enkratic requirement (or a corresponding basing prohibition) in the former but not in the latter world, I think, just because only in the former world can following our normative judgments properly realize accountable normative self-governance.

So I think what my earlier arguments actually support is a *functionalist conception of practical rationality*, according to which truths about what practical rationality requires of us are truths grounded in the proper function of our agency.[9] More specifically, on this conception, an intentional state pattern is proper and prescribed by practical rationality just in virtue of how it is constitutively and distinctively involved in the proper function of our agency.[10] This functionalist conception of practical rationality, I believe, is more attractive than the radical internalist and the rationalist conception for several reasons.

4.2. The Attraction of the Functionalist Conception

(1) The Grounds of the Requirements of Practical Rationality

[9] This conception has already been suggested by Bratman (2018) and Raz (2005). The functionalist conception also bears resemblance to Wedgwood's account of rationality as a virtue (2017: ch.6), given the structural similarity between virtue and agency.

[10] The qualification of *constitutive* and *distinctive* involvement here is meant to exclude cases where, say, an evil demon threatens to interfere with the function of our agency unless we maintain a particular intentional state pattern that he would like to see. Only the intentional state patterns that are constitutively and distinctively involved in the proper function of our agency in all cases can be elevated to the status of rational patterns.

The functionalist conception is attractive because it provides a more plausible explanation for the grounds of p-requirements. Principles of practical rationality don't have their sources in the agent's own subjective view about what mental patterns are supported by reason, nor are they grounded in some immutable, necessary requirements on such patterns. The former radical internalist picture simply defers too much to the agent's subjective perspective and potentially rationalizes all kinds of intentional state patterns, while the latter rationalist account implausibly entails that some intentional state patterns would always be prescribed by practical rationality in all possible worlds, despite their serving no point at all in some of these worlds (for instance, the pattern of having both a telic intention and a higher-order approving desire is useless or even counterproductive in our actual world). The functionalist conception, in contrast, offers a more plausible story about the sources of p-requirements. Indeed, as we have just seen, this conception nicely explains the origins of the three most fundamental p-requirements (namely, the requirements of instrumental rationality, practical non-contradiction, and Enkrasia), for the mental patterns that they prescribe are indeed constitutively involved in the proper function of different aspects of our agency.[11] In fact, the functionalist conception can even allow Kantians to ground their preferred principles of practical rationality in more substantive conceptions of the proper function of agency (see Korsgaard 2009).

[11] The p-requirements codified in contemporary decision theory can also be explained in the same way. For instance, a pattern of transitive preferences is proper, as the famous money-pump argument has shown, just because a non-transitive, cyclic preference pattern will lead us to alternate between the fulfillment of the different options involved in the preference cycle and to eventually waste our general resources for preference fulfillment (e.g. time, money). And other p-requirements that are given instrumentalist justification by decision theorists can also be thought of as contributing to the proper use of our general resources for preference fulfillment. And that is why the preference patterns prescribed by these p-requirements can also be said to be involved in the proper or optimal function of our executive agency.

(2) The Character of the Requirements of Practical Rationality

The functionalist conception is appealing also because it accounts for the character of p-requirements very well. So p-requirements have an *independent* character and hold independently of what we think just because what they require is determined by which mental patterns are constitutively involved in the proper function of our agency, which has a nature that is not of our own making. P-requirements are *contingent* just because which mental patterns are constitutively involved in this way are contingent on empirical features of our agential capacity (e.g. what our decision-making and executive mechanisms are actually like), as well as the distinctive evolutionary history that shaped it. Finally, p-requirements have a *seemingly internal* character just because they are derived from the proper function of our agency, which is still something, in a figurative way of speaking, within our head. Finally, the functionalist conception actually shows that p-requirements can be said to internal in a *constitutive* sense: Given that p-requirements prescribe the constitutive mental patterns indispensable for the proper function of our agency, we can say that to be an agent with properly-functioning agency is to more or less exhibit the mental patterns that conform to these requirements.

(3) The Constitutive Status of Practical Rationality

Moreover, I think the functionalist picture gives us a convincing explanation as to why p-requirements can have a constitutive status in the first place (Davidson 1985a; Korsgaard 2009: ch.2 and ch.4; Wedgwood 2017: 147). The idea that

rationality is constitutive of agency has been put into doubt because it seems to entail an overly intellectualistic and demanding picture of agency. After all, many ordinary agents do not actively aim to follow p-requirements at all; some even fail to conform to these requirements from time to time. But this does not seem to undermine their status as agents. So how can p-requirements be constitutive of agency in the first place?

I think the idea of proper function gives the constitutivist thesis a naturalistically-respectable foundation and thereby addresses this central worry about constitutivism about rationality. To begin, on the functionalist picture, to conform to p-requirements is just to exhibit the psychological patterns constitutively involved in the proper function of our agency, and the proper functioning of our agency is often an automatic process. That is why we can often form instrumental intentions, avoid contradictory intentions, and conform to other relevant p-requirements in an automatic manner. Still, reflective beings like us can reflect on the relevant constitutive mental patterns, render them explicit, and then intentionally try to maintain these patterns so as to better ensure the well-functioning of our agency. That is why we can also follow p-requirements in a more conscious, deliberate manner.[12] Therefore, those who do not actively aim to follow p-requirements can still be agents because the conformity to these requirements can often be achieved through an automatic process on the functionalist picture.

What is more, the functionalist conception explains why those who fail to conform to p-requirements in the first place can still be agents. This is because our

[12] Notice that only in the latter case can we be said to follow the relevant p-requirements. This is because to *follow* a norm (as opposed to merely *conforming to* it), we have to do what it requires because of our recognition of the norm itself.

agency encompasses various sub-mechanisms. So someone who fails to conform to the rational constitutive principle for one sub-mechanism might still perfectly follow the principle for a distinct sub-mechanism. For instance, as I argued in Chapter 2, someone who flouts the integration principle will typically have difficulty in forming an intention to successfully settle her practical question. But this kind of person might still conform perfectly to the instrumental principle and be quite efficient in carrying out her intentions once she forms one. So she can still be qualified as an agent on the basis of that.

More importantly, on the functionalist picture, to be practically rational is to conform to p-requirements and to exhibit the relevant constitutive mental patterns. And an agent's practical rationality, like the proper functioning of her agency, can come in degrees, depending on how often she exhibits the relevant constitutive mental patterns (Wedgwood 2017: 148, 212-3).[13] An agent who fails to conform to the instrumental principle 5 times out of 100 can still be called instrumentally rational and be said to have a properly-functioning executive system; she can still be counted as an agent despite her occasional lapses. But what if the failure rate is 50 times out of 100? Or 70 times out of 100? Sometimes it is difficult to pinpoint the loss and the emergence of agency/practical rationality in such a continuous spectrum. This is then another reason why one's failure in conforming to p-requirements does not automatically undermine one's status as an agent on the functionalist conception.

Thus, the functionalist picture presents a plausible construal of the otherwise controversial constitutivist thesis. It therefore provides an attractive picture of

[13] Indeed, it is only when we focus on the rationality of a specific mental pattern or mental step that practical rationality becomes an all-or-none matter (for either this mental pattern conforms to a particular rational requirement or it does not).

practical rationality for those who are drawn to the constitutivist idea.

(4) The Normativity of Practical Rationality

Finally, the functionalist conception also better preserves the normativity of practical rationality than the other two competing conceptions. Some philosophers have recently expressed doubt about the normativity of rational requirements (Broome 2005; Kolodny 2005; cf. Wedgwood 2017: ch.8). Indeed, the rationalist conception of rationality is particularly vulnerable to this attack (see Kolodny 2008a; 2008b). After all, if rationality is only about having coherent and "tidy" intentional state patterns prescribed by certain necessary requirements, then what is the point of being rational? Do we have a normative reason to maintain such patterns just for internal coherence's own sake? Although the radical internalist conception of rationality seems to better preserve the normativity of rationality to a certain extent, the kind of normativity that it saves is actually not the one we want. So, on this conception, the mental states or patterns prescribed by rationality are just those that we believe to be proper and supported by reason. That is why *it would seem to us* that we have a normative reason to follow what rationality prescribes. However, as you can see here, what this establishes is only that the mental patterns prescribed by rationality are normative in an apparent or subjective sense: They are mental patterns which we take ourselves to have a reason to maintain (cf. Kolodny 2005, 2008b: 460-1; Scanlon 2007: 87-8). But there is still no genuine normative reason for us to maintain these patterns in and of themselves. So rationality still cannot be

genuinely normative for us on the radical internalist conception.[14]

The functionalist conception, in contrast, at least provides a more promising and palatable answer as to why practical rationality can at least have a kind of objective normative authority over us. This is because, on this conception, the mental patterns prescribed by practical rationality are just those constitutively involved in the proper function of our agency. So we won't even be an agent if we don't by and large follow the requirements of practical rationality and maintain the mental patterns that they call for. In order to make a decision and settle our practical question in favor of a single course of action, we have to follow the requirement against contradictory intentions (which often cancel each another out). In order to further fulfill a telic intention, we have to follow the instrumental requirement and form relevant instrumental intentions, or other intentional states that are constitutively indispensable for our general intention-fulfillment. So practical rationality is not just about having tidy intentional state patterns; it's about having intentional state patterns that help us to be agents and to do things. That is why requirements of practical rationality serve a point and are normative for us.

Of course, this still does not explain why we have a normative reason to *always* follow p-requirements, for occasional non-conformity for sure won't undermine our agency (Kolodny 2008b: 459; Wedgwood 2017: 205-9). I will address the more radical version of the "why-be-rational" challenge in the next chapter. However, for now, my point is that the functionalist conception can at least explain why there is an objective normative reason for us to *by and large* follow the requirements of practical rationality. And in this respect, it can at least better save the normativity of

[14] Of course, Enkratics can still say that the disposition to follow one's reasons-judgments is a valuable disposition to have, see Kolodny (2008b: 457-8). But this still does not show why we have an objective normative reason to follow what rationality requires of us, especially when we take into account that our reasons-judgments might sometimes go wrong.

practical rationality than the other two competing conceptions of practical rationality.

4.3. The Characterization of Function and Rationality

There are two major worries about the functionalist conception of practical rationality. Both concern how we can properly characterize practical rationality in terms of function. The first worry was raised by Joseph Raz. He argues that we should not understand instrumental rationality in terms of the proper function of our desire-realization capacity, for many types of malfunction in the latter capacity (e.g. "to fumble, drop things, mix them up, mix up left and right (2005: 25)") clearly do not make us instrumentally irrational. Similarly, you might think that many kinds of malfunction in our agency (e.g. lacking agility in bodily movement or lacking attention while acting) likewise do not make us irrational. But wouldn't the functionalist conception entail that all instances of malfunction in agency are cases of practical irrationality?

I think this worry largely stems from a misunderstanding of the precise shape of the functionalist picture of practical rationality. The central idea behind my proposal is that certain psychological patterns (or certain steps in a psychological process, as I will argue later) are proper and prescribed by practical rationality just in virtue of how they are constitutively involved in the proper function of our agency. In line with this, we can say that on the functionalist conception, practical irrationality only consists in the failure in maintaining these relevant constitutive psychological patterns. Now lapses in our bodily movement are not failures in maintaining such *psychological* patterns, and that is why they do not constitute

instances of practical irrationality. Lapses in other general psychological capacities related to our agency (such as the lack of attention), on the other hand, sometimes do result in failures in maintaining the relevant constitutive patterns (e.g. we might fail to form an instrumental intention due to our inattention), but sometimes they do not (e.g. we might fail to realize our telic intention due to our inattention while acting, even though we have already taken all the psychological steps prescribed by the instrumental requirement). Again, psychological lapses in the latter case cannot constitute instances of practical irrationality. So I think the functionalist conception, once it is suitably characterized, won't entail the implausible identification of all instances of agential malfunction with practical irrationality.

The second worry is this. Given that our agential capacity is shaped and selected by our distinctive evolutionary history, it seems that its proper function is also meant to contribute to our fitness in surviving and reproducing in our typical environment. As a result, the psychological patterns constitutively involved in its proper function might sometimes be fitness-promoting yet irrational ones. For instance, the proper-function of our decision-making mechanism may in fact require us to form intentions in a quick and dirty way. This is because instead of spending too much time and mental resources in eliminating contradictory intentions and in reaching firm settledness on a particular course of action, doing so might actually maximize our ability to act and consequently enhance our evolutionary fitness. In light of this, one might say that the proper function of our agency also constitutively involves contradictory intentions formed in a quick and dirty way. And wouldn't the functionalist conception entail that such a pattern of intention is also rational and prescribed by practical rationality?

A related worry can be raised here. My earlier arguments for the contingency

thesis largely relied on how we might have very different psychological mechanisms for decision-making and intention-fulfillment in other possible worlds. So in another possible world our earlier intentions might be able to settle our practical question without being easily overturned by later, contrary intentions. Moreover, in another possible word our intention-fulfillment might require us to form other types of intentional states instead of instrumental intentions. However, as you might think, the intentions in these different possible worlds just have very different causal-functional profiles from the ones in our actual world. After all, they typically have a lower degree of sensitivity to contradictory intentions, and they typically lead to actions in conjunction with other types of intentional state. So perhaps the intentions in these worlds are just very different beasts from our intentions: They are sch-intentions rather than real intentions like the ones we have. And instead of saying that our rational requirements are contingent on how our intentions might have different empirical features in these different possible worlds, why shouldn't we simply say that there is a different set of rational requirements for sch-intentions in these worlds?

I think both worries can be addressed by observing how our understanding of function often involves a degree of abstraction. We can often focus on the core function of agency or that of a particular type of intentional state, and abstract away from its quirky empirical features. One reason for this kind of abstraction lies in our quest for theoretical generality. So even if our counterpart's sch-intentions have a slightly different causal-functional profile from our own intentions, we can still abstract away from such differences and treat them as variations in the realization of the same general kind of intention with the same core function (i.e. the function in settling our practical question and leading us to change the environment in

accordance with its content). With this kind of generalization, we can then formulate a broader theory of the rationality of intention, instead of having fragmented accounts for the rationality of different species-specific intentions (see Dennett 1987: 58-60).

Another reason for abstraction lies in our quest for theoretical idealization. Even if the proper function of our intention requires us to settle our practical questions in a quick and dirty way so as to make the best use of our limited psychological resources, we can still abstract away from this, and treat the function of our intention as if it is unbounded by such psychological constraints. With this kind of idealization at hand, we can then formulate a simpler account for the (unbounded) rationality of intention, instead of one with many provisos that take into account our psychological limitations (cf. Wedgwood 2017: ch. 6). So, for instance, we can simply say that practical rationality requires us to form the instrumental intentions for the relevant means and forbid us from having contradictory intentions, rather than saying that it requires us to form the instrumental intentions for the relevant means, if the cost of considering the potential means and forming the instrumental intentions is not higher than the potential gain from the fulfillment of telic intentions; or that practical rationality forbids us from having contradictory intentions, if the cost of detecting such a contradiction is not higher than the potential loss from its undermining effect in our decision-making.

In fact, we abstract away from quirky, species-specific features and empirical psychological limitations not just for the sake of theoretical generality and idealization. This kind of abstraction in our understanding of function also helps us to capture certain fundamental truths about agency. This is because there are two

most basic ways for any intentional system to interact with its environment: One is to represent the environment with the content of its intentional state; the other is to change the environment in accordance with such content. The capacity to interact with the environment in the latter way constitutes agency in the most fundamental sense, which is naturally understood in abstraction from different intentional systems' species-specific features and their idiosyncratic psychological limitations.

And it is the proper function of agency in this most basic sense, I think, that grounds the most basic principles of practical rationality, which presumably apply to all intentional systems. In fact, we should think of these basic principles as prescribing some *abstract rational mental patterns*, patterns which are constitutively involved in the proper function of agency, when agency is understood in its most fundamental sense. So, for instance, we can say that they require any intentional system to *take the indispensable psychological steps for the fulfillment of its desires (i.e. intentional states with world-to-mind direction of fit)*, to *avoid mutual-undermining in these states*, and to *stick with these states while it carries them out*. Now these basic rational principles, on the one hand, enjoy a fundamental status and have universal applicability. This is because they are grounded in the proper function of agency in the most fundamental sense. On the other hand, these principles still have a contingent character, as I have just argued. This is because the abstract rational mental patterns they prescribe (say, taking the indispensable psychological steps for one's desire-fulfillment) manifest themselves in different specific psychological patterns in different intentional systems. In fact, as I will explain shortly, we should think of these abstract rational patterns as being multiply-realized by different specific psychological patterns in different intentional systems and in different environment.

Therefore, I would say that on the functionalist conception of practical rationality, a rational mental pattern is one that is constitutively involved in the proper function of agency, when agency is understood in its the most fundamental sense. (In fact, this is also the conception of function that I am going to rely on in the next chapter when I defend the functionalist account of normative force.) The mental patterns constitutively involved in species-specific proper function or in proper function within the bounds of our available psychological resources, on the other hand, are rational only from a *species-specific* perspective or from the perspective of *bounded rationality*. And the functionalist conception nicely accommodates these two additional senses of rationality, which are often overlooked by philosophers but still taken seriously in both the empirical sciences and folk psychology. So the problem of the characterizing proper function turns out to add more plausibility to the functionalist conception.

5. The Significance of the Functionalist Conception

In sum, the functionalist conception of practical rationality adequately explains the grounds for the most central p-requirements, including the requirements of instrumental rationality, practical non-contradiction, and Enkrasia. It nicely accounts for the basic characteristics of p-requirements, namely, their independence, contingency, and constitutive status. It also preserves the normativity of practical rationality better than the radical internalist and rationalist conceptions. Finally, it persuasively accommodates how our evaluation of practical rationality can also be made from a species-specific perspective or from the perspective of bounded rationality.

But why should we even care about the functionalist conception of practical rationality? What are its implications for our general understanding of rationality? This account is important, I believe, because it implies that practical rationality is something embedded in the environment: What specific form it takes is determined by our evolutionary history and empirical features of the world we are in. This conception hence undermines the stronger, rationalist variant of internalism and provides the groundwork for a more empirically-minded picture of practical rationality.

(1) Pattern Requirements and Process Requirements

To begin, once we accept the functionalist conception, we will no longer have good reason to take practical rationality to solely consist in having proper *mental patterns*. After all, on this conception, certain mental patterns get to be called rational only because they are constitutively involved in the proper function of our agency, and what practical rationality is ultimately about is the underlying proper function. However, in many cases, what the proper function of our agency requires is that we take certain *mental steps* rather than to simply maintaining certain mental patterns. Accordingly, the requirements of practical rationality should encompass not only pattern requirements but also process requirements on the functionalist conception (cf. Kolodny 2005). Indeed, as we saw in Chapter 1, when it comes to the instrumental requirement, only a kind of process requirement that prescribes us to take certain psychological steps can capture the executive dimension of IR, for a means-end coherent mental pattern alone cannot contribute to our intention-fulfillment in all cases. Furthermore, sometimes the proper function of our

agency might only involve the working on a single intentional state, say, specifying the content of our telic intention into a more detailed action plan, or not reconsidering the telic intention that we have just formed (Bratman 1987: ch.5 -6). When it comes to this kind of case, it is also more plausible to say that practical rationality imposes process requirements rather than pattern requirements on us. (To avoid repetition, in what follows I will understand mental patterns in a broader sense and take them to include both mental processes and mental patterns in a narrow sense.)

(2) The Multiple Realizability of Practical Rationality

More importantly, the functionalist conception of practical rationality entails that practical rationality does not strongly supervene on one's *mental patterns*: Even if two agents in two different possible worlds exhibit the same mental patterns, they can still differ with respect to their practical rationality. This is because, owing to the differences in the evolutionary history and empirical features of their inhabited worlds, these two agents may very well have different kinds of underlying mental mechanisms for decision-making and intention-fulfillment to being with. So the proper function of their agency might simply implicate different mental patterns, making them subject to different p-requirements in the first place. That is why they can still differ with respect to their practical rationality even if they share the same mental patterns. For instance, someone who has a telic intention but lacks a higher-order approving desire is practically irrational in the other possible world that we considered in Section 4. Nevertheless, someone with the exact same mental pattern can still be perfectly rational in our actual world.

Of course, as I mentioned earlier, there are some basic p-requirements prescribing certain abstract rational mental patterns that are constitutively involved in the proper function of agency, when agency is understood in its most fundamental sense. And the abstract patterns that these requirements prescribe (say, *taking the indispensable psychological steps for our general intention-fulfillment*, or *avoiding mutual undermining in our intentions)* can still be said to be practically rational in a necessary manner. This is not surprising, for these are just the formal characterizations of what is involved in the proper function of agency in the most fundamental sense. Still, what the functionalist conception shows is that these abstract rational patterns can in fact be realized by different specific mental patterns in different worlds. And it is contingent environmental factors that determine which specific patterns can be such realizers and become prescribed by practical rationality in a particular world. The idea of multiple realizability therefore also has applicability to practical rationality itself.[15]

This in turn undermines the stronger, rationalist variant of internalism. As we have seen in Section 3, for rationalists, certain mental patterns are rational or irrational in a necessary, *a priori* manner. Consequently, many rationalists believe that rationality is determined solely by the mental patterns that one exhibits, such that rationality can be said to supervene on these patterns strongly. But the functionalist conception shows that this view is actually not plausible when it comes to practical rationality. Certain mental patterns are practically rational not because

[15] Notice that what is involved here is not the multiple realizability of our intentional states themselves, but that of abstract rational patterns. So rationalists can claim that a specific mental pattern (say, having a telic intention with an instrumental intention) is rational in a necessary and *a priori* manner, while at the same time perfectly recognizing that the intentional states involved in this pattern can be multiply realized by different physical states. However, what the functionalist conception shows here is that what constitutes a rational mental pattern in the first place is multiply realizable.

they are coherent in a necessary, *a priori* manner; rather, their rational status depends on whether they are constitutively involved in the proper function of our agency, which is hostage to the fortune of environmental factors. As a result, practical rationality cannot be solely determined by— nor can it strongly supervene on—one's mental patterns. We have to look beyond mental patterns at the superficial level.

(3) Practical Rationality and the External Environment

Nevertheless, on the functionalist conception, practical rationality can still be said to strongly supervene on one's *mental properties*, when these properties are taken to include one's underlying psychological mechanisms. If two agents share the same kind of underlying psychological mechanism to begin with, then the proper function of their agency will presumably implicate the same mental patterns, subjecting them to the same p-requirements. So if they are really indistinguishable with respect to their mental patterns, then presumably they will also be indistinguishable with respect to their practical rationality. Similarly, practical rationality can also be said to *weakly* supervene on one's mental patterns on the functionalist picture. This is because two agents in the same possible world will presumably share the same type of underlying psychological mechanism, unless one of them is from a planet with different external environment and evolutionary history. So their pattern indistinguishability will presumably also guarantee their rational indistinguishability.

This then shows that the functionalist conception is still compatible with a weaker version of internalism, which restricts the object of our rational evaluation

to what happens in one's head while allows the criteria of our rational evaluation (i.e. p-requirements) to be empirically determined (cf. Wedgwood 2017: ch.7).

However, as we have just seen, the guiding spirit of the functionalist account is that practical rationality is ultimately about the proper function of our agency. But the idea of proper function can be understood in either an information-sensitive or an environment-sensitive sense. For instance, we can say that the proper function of our rational executive system consists in forming instrumental intentions in a way that is sensitive to the means that we *believe to be able* to fulfill our telic intentions (i.e. subjective means), given the information we have. But we can also say that its proper function resides in forming instrumental intentions in a way that is sensitive to the means that can *in fact* fulfill our telic intentions (i.e. objective means). In other words, we can take the proper function of agency as consisting in either the proper relations between our intentional states themselves, or between our intentional states and the world. (In fact, we have short-arm and long-arm versions of functionalism in philosophy of mind just because we can understand function in these two ways.)

Given this, it also seems to make good sense to ground p-requirements in the environment-sensitive sense of proper function. After all, it is this sense of proper function that explains the information-sensitive one: We are first supposed to intend the means that are able to bring about our ends and to avoid infeasible ends, and it is in virtue of this that we should be sensitive to our beliefs about these matters. However, once we accept the environment-sensitive sense of proper function and ground p-requirements in it, our evaluation of practical rationality will no longer be restricted to what happens in our head; it will also have to concern whether our intentional states bear the right relations to the world. For instance, we would have

to say that the instrumental requirement also prescribes us to form instrumental intentions for the objective means that can in fact bring about our ends, and that the feasibility requirement that I have defended in Chapter 3 also forbids us from having telic intentions that are in fact infeasible. In other words, once we accept this additional sense of proper function, practical rationality will no longer weakly supervene on our mental properties, for p-requirements and the proper function of our agency can surely also concern the proper relation between our mind and the world.

I am not able to give a full defense for externalism about rationality here. In fact, I think there are still good arguments available in support of the weaker version of internalism. Perhaps rational norms need to be able to guide us directly (Wedgwood 2017: ch.7). Or perhaps these norms are supposed to govern how we should properly respond to our existing information by its very nature (Sobel 2001). So maybe only p-requirements grounded in the information-sensitive sense of proper function can capture these features of rational norms. Nevertheless, what I want to highlight here is that the functionalist conception can at least accommodate people's externalist intuitions[16] and lay the foundation for externalism about practical rationality—all we have to do is to appeal to the environment-sensitive sense of proper function. This compatibility with both internalism and externalism, as I take it, is another virtue of the functionalist account.

6. Conclusion

[16] Indeed, if you ask non-philosophers about the rational status of an intentional system that is wholly insensitive to changes in its external environment and incapable of tracking the measures that can in fact fulfill its desires, many would judge it to be irrational from an externalist perspective.

Here I have examined three different conceptions of the nature of practical rationality. And I have argued that the requirements of practical rationality are grounded neither in the agent's subjective perspective nor in some supposedly necessary, *a priori* requirements that apply to all possible worlds. Rather, they are grounded in the proper function of our agency. This explains why these requirements have an independent and contingent character, yet are still seemingly internal and normative due to its constitutive status.

This functionalist conception is philosophically significant because it provides a more empirically-minded and environmentally-embedded picture of practical rationality: It reveals how the requirements of practical rationality encompass both pattern and process requirements, and how practical rationality is multiply realizable by different mental patterns in different worlds. In fact, it directly undermines the stronger, rationalist variant of internalism. Although it is still compatible with a weaker version of internalism, it also leaves room for a fully externalist picture of practical rationality.

More importantly for our general purposes here, the functionalist conception clarifies the underlying nature of the three central rational principles for intention that I have considered in the foregoing three chapters, i.e. the instrumental principle, the integration principle, and the feasibility principle. These principles are in fact constitutively indispensable for the proper function of our agency, and that is why they are normative for us. Of course, one might still question why the constitutive status of these rational principles for intention makes them normative for us. Why should we care about the proper function of our agency in the first place? And what is the normative significance of intentions? These are then the remaining questions that will be the focus of the final chapter of this thesis.

Chapter 5 Function and Normativity

1. Introduction

I argued in Chapter 1 that the instrumental principle is grounded in our rational executive ability for intention-fulfillment. In Chapter 2 and 3 I contended that our intention, given its settling and executive functions, is also subject to the integration and the feasibility principles as its core rational constitutive principles. I then argued in Chapter 4 that principles of practical rationality are ultimately grounded in the proper function of our rational agency. However, I have postponed two sets of questions about my overall view here: First, are principles of practical rationality (including the rational principles for intention that I have defended) normative for us? Do we have a normative reason to follow them? Second, are intentions normative for us? Can they ground normative reasons for action for us?

In the foregoing chapter I gave a preliminary answer to the first set of questions. I claimed that principles of practical rationality are normative for us because they prescribe the psychological steps and patterns that are constitutively involved in the proper function of our agency. This means that if you don't by and large follow them, then you cannot even be an agent with properly-functioning agency at all. However, this kind of constitutive justification may appear unsatisfactory, and one might even question whether we really have a normative reason to be an agent and to follow the rational constitutive principles of agency, as opposed to being a carefree schm-agent unencumbered by these rules (Enoch 2006).

The questions about the normativity of practical rationality and intention are extremely important. Indeed, I don't think I can provide a satisfactory solution to all of these questions within a single chapter. Nevertheless, I believe we can at least take some steps toward an answer to these questions if we reflect more carefully on the very concept of normative reason and its structure. In fact, if we do this, we will come to see how practical rationality constitutes the pre-condition for our receptiveness to normative reasons for action, and then how our intentions can provide such reasons in virtue of the very function of our action and agency.

My tentative proposal has two parts. The first part concerns the structure of normative reason. As I will argue, the concept of normative reason features a specific kind of response-focused normativity, which is a normative force that a particular consideration has in calling for a certain response on our part (say, a particular attitude, emotional response, or action) (Section 2). The second part of my proposal is about the nature of this kind of response-focused normativity. As I will further contend, the normative force of a normative reason in calling for a particular attitude, emotional response, or action consists in the former's capacity to engage the latter's function. This *functionalist account of normative force* is plausible because it rightly captures the extensions of different kinds of normative reason. Normative reasons for belief, for instance, are provided by truth-related considerations just because they are considerations capable of properly addressing the function of belief in representing what is the case. This functionalist account also explains why normative reason has both a justificatory and a motivating dimension: If a consideration is capable of properly engaging belief's representational function, for example, then it will also be able to show how a certain belief is more likely to be true and thereby justify it by its functionally-given standard of correctness. And this

consideration will also be able to exert some pressure on us in forming such a belief, provided we have a proper-functioning belief system. In fact, the functionalist account, as I shall further argue, also rightly accounts for the phenomenology of different kinds of normative reason, as well as providing an informative distinction between the right and the wrong kinds of normative reason (Section 3).

The functionalist account of normative force then provides an answer to the first set of questions about the normativity of practical rationality. To begin, given that the normative force of a normative reason in calling for a particular type of response R resides in its capacity to engage R's function, those who deny the normativity of R's proper function are often insusceptible to the influence of the normative reasons for R. For instance, if you do not care about the proper function of your belief in representing truth, then no normative reasons for belief will hold sway over you. Similarly, the acceptance of the normativity of the proper function of agency, as I shall contend, is also a precondition for one's susceptibility to the normative force of reason for action. That is why the proper function of our agency—and by extension the principles of practical rationality that contribute to it—can be said to be normative (Section 4).

The functional account also provides a new line of argument for reasons internalism and answers the second set of questions about the normativity of intention. Given that a normative reason exerts its normative force in calling for a type of response by engaging its function, and given that the function of our action is to fulfill an intention of ours, it follows that only considerations related to our intention-fulfillment can have normative force in calling for our action. Indeed, this kind of volition-based reason for action, I contend, actually possesses a kind of

normative force that is able to address our first-person normative question (Section 5).

2. The Structure of Normative Reason

On the most minimal construal, a normative reason is a consideration that counts in favor of a response on the agent's part (Dancy 2000; Parfit 2001; Scanlon 1998: 17). The concept of normative reason thus involves two relata. The first relatum is the very consideration with certain normative significance, a kind of significance that enables it to play a counting-in-favor role. Exactly what kind of consideration can play this counting-in-favor role and constitute a normative reason is determined by our best account about the metaphysical nature of normative reason. For example, a normative reason might consist in a fact (e.g. the fact that the house is on fire is a normative reason for me to believe that it is on fire); a state of affairs (e.g. the state of affairs that the Rohingya people is oppressed by the current regime in Myanmar is a normative reason for us to boycott it); or a mental state that we have (e.g. Mike's desire that he becomes a singer is a normative reason for him to enter the talent competition), and so on. Also, exactly where the normative significance of a particular consideration/normative reason comes from has to depend on our best account about the source of normativity. So the normative significance of a normative reason might come from the very features of a fact or a state of affairs itself (perhaps the fact that I am in pain has inherent normative significance for everyone just in virtue of the intrinsic badness of pain); from some social norms or institutions (perhaps the fact that I am in pain has normative significance for my parents or employer, in virtue of their having a special social or legal relationship

with me); or from people's mental states (perhaps the fact that I am in pain has normative significance for those who desire to help others, in virtue how relieving my pain is a way for them to fulfill their desire). Notice that while many philosophers don't distinguish between them clearly, the metaphysical nature of normative reason and its source of normative significance are two separate issues. A subjectivist, for instance, can take the source of normativity for all normative reasons for action to lie in agents' desires, while at the same time taking these normative reasons to be constituted by facts in the world. For example, a subjectivist can say that the fact that there is going to be a talent competition in London is a normative reason for Mike to book a train ticket to London, but this fact has normative significance and constitutes a normative reason for him only because he desires to be a singer.[1]

The other relatum involved in the concept of normative reason is the response on the agent's part that is spoken in favor of or called for.[2] It is a kind of response that the agent has some control over and that she can form in reaction to the consideration with normative significance. So it can be the formation or the revision of a particular attitude (e.g. belief, intention), the exhibition of an emotional response (e.g. fear, indignation), or the performance of a specific action (e.g. promoting something or refraining from doing something). For ease of discussion, in what follows I will use *response* as a placeholder for the attitude, emotional response, or action that is called for by a consideration with normative significance.

[1] For a full defense of this kind of subjectivist view, see (Schroeder 2007).

[2] Here I leave it open whether the "agent" here denotes agents in general or a specific agent. If you think of a consideration as calling for a response on agents in general, then what you have in mind will be an agent-neutral reason. If you conceptualize it as calling for a response on a particular agent, then it will be an agent-relative reason. For these two formulations of normative reason, see Nagel (1970: 90-5).

As we can see here, the very concept of normative reason already involves an internal structure. It first denotes (1) a particular consideration with certain normative significance, which enables it to play a counting-in-favor role; it then refers to (2) the response on the agent's part that is counted in favor of or called for by that consideration. Given this internal structure, we can make two observations about the idea of normative reason.

First, the concept of normative reason is essentially response-focused, that is, it involves an essential reference to a response that is called for on the agent's part. This sets the idea of normative reason apart from other kinds of normative concept. The concept of value, for instance, can be taken to denote a feature with normative significance without making an essential reference to the responses that an agent is supposed to make in reaction to it. Suppose there is a possible world which is inhabited by sentient beings without the capacity for making any responses to their environment. Perhaps those beings have already evolved a sensory system for feeling pleasure and pain consciously, but not a corresponding motor system for initiating any movements, nor an intentional system for forming any attitudes toward things in their environment, such that those beings can only feel pleasure and pain passively. Suppose now a powerful whirlwind blows these sentient beings away, tossing them onto the hard surface of this world, consequently inflicting extreme pain on them. Now this is a world in which a tragic event with negative value just happened, but this is also a world without normative reasons. After all, the idea of value seems to presuppose only a feature or a consideration with normative significance (e.g. something's being in pain), while that of normative reason has to additionally invoke the response that an agent is supposed to make in reaction to that feature or consideration. Yet in this scenario, there are simply no

beings in this world that have the capacity to respond to this event and to address the pain that is experienced.

Of course, it is controversial whether there can be any value without agents, and whether the idea of value can be understood without reference to certain responses that agents are supposed to make. For instance, you might think that if the tragic event in the world that we just considered is truly bad, then its normative significance would still speak in favor of certain responses on the agents' part, *if there were any agents*. One could even embrace the fitting-attitude analysis of value and explain the badness of this tragic event in terms of how its properties provide normative reasons for agents, if there were any, to respond to it in certain ways. In other words, you might argue that all seemingly non-response-focused normative concepts (e.g. value) can be ultimately reduced to response-focused ones (e.g. normative reasons to form an attitude or to act) (Rabinowicz and Rønnow-Rasmussen 2004; Rønnow-Rasmussen 2011; Scanlon 1998: 96-7; Rowland 2016). Here I will not take a stand on whether there is a kind of normativity that does not presuppose the existence of any agent, and whether there are any irreducible non-response-focused normative concepts. All I want to point out here is that the idea of normative reason is at least uncontroversially response- focused, in virtue of its making an essential reference to a response that an agent is supposed to make. Indeed, it seems difficult to think of a normative reason *simpliciter*. A normative reason always has to be a normative reason for an agent to form a certain response.[3]

[3] Of course sometimes people talk about how there is no reason for something this bad to happen. But I think this kind of expression either uses the term "reason" in its explanatory sense (there is no explanation as to why something this bad came about), or presupposes an anthropomorphic understanding of the world (there is no normative reason for the world to allow something this bad to happen).

Second, given that the concept of normative reason is response-focused, it can be said to involve a distinctive type of normativity that is specifically targeted at a response on the agent's part. Any consideration that is able to play a counting-in-favor role, and to potentially constitute a normative reason, is supposed to carry certain *normative significance*, which is a kind of non-response-focused normativity that is not yet specifically targeted at a response that the agent is supposed to make. However, such normativity, when it is actually applied to and imposed on the agent, will be translated into a *normative force for a certain response*, which, in contrast, is a kind of response-focused normativity that specifically calls for a certain response on the agent's part. We can in principle distinguish between these two types of normativity. Suppose Tom is having an unbearable headache. The fact that he is in severe pain can carry some normative weight by itself. It is a consideration with genuine normative significance that enables it to play a counting-in-favor role. This consideration, however, can only exert a normative force on you to *act* and to assuage his pain, but presumably no normative force at all on you to *feel indignation* about it (let us suppose nobody deliberately caused his unbearable headache). Similarly, suppose a wealthy but vicious man threatens to inflict extreme pain on an innocent person unless you believe that the earth is flat. This would likewise constitute a consideration with genuine normative significance. Nevertheless, its normative significance can only be translated into a practical normative force on you to *want* to bring it about that you have the belief in question, but not an epistemic normative force on you to *believe* that the earth is actually flat. As these cases show, the normative significance of a consideration can sometimes fail to translate into a normative force for a certain response, and that is why it would be fruitful for us to distinguish between non-response-focused and

response-focused normativity. In fact, as we will see later, it is because of the divergence between these two kinds of normativity that the so-called "wrong kind of reason problem" arises.

As I noted earlier, the concept of normative reason is essentially response-focused. So the kind of normativity that is distinctively involved in normative reason is the response-focused kind, i.e. the normative force that a consideration has in calling for a certain response on the agent's part. In what follows I will focus on this kind of normativity. And I will concede for the sake of argument that a realist account about non-response-focused normativity is true, that is, some considerations can just carry normative significance in virtue of some attitude-independent normative facts (say, the consideration that someone is in severe pain can itself possess some normative significance given the objective badness of pain). As I shall contend, even if we grant this realist view about non-response-focused normativity, we still have to understand response-focused one in terms of function.

3. The Functionalist Account of Normative Force

The proposal I would like to make is this:

> **The Functionalist Account of Normative Force**: For any consideration with certain normative significance that enables it to play a counting-in-favor role and to potentially constitute a normative reason, its normative force in calling for a certain response r consists in its capacity to properly engage the function of the type of response that r belongs to.

What is a function? Here I will adopt the idealized conception of function that I proposed in the preceding chapter, and I will understand the functions of different types of attitudes and responses in a way that is species-neutral and unbounded by psychological limitations. Of course, the function of emotional response (e.g. fear, indignation) perhaps cannot be adequately understood in abstraction from such empirical features. So when it comes to emotions, we will have to settle for a less idealized conception of function. However, in the cases of belief and action, their core functions are derived from the two most fundamental ways for us to interact with the world: One is to represent the world with the content of our mental state; the other is to change the world in accordance with such content. As I indicated in the last chapter, such core functions of belief and action are naturally understood in abstraction from species-specific features and their psychological limitations. And it is this basic, fundamental sense of function that I will rely on in my analysis concerning the normative force of reasons for belief and for action. So, for instance, we should say that the function of belief is to correctly represent what is the case, instead of saying that its function is to represent objects that typically contribute to our fitness in surviving and reproducing in our environment, or that its function is to correctly represent what is the case only when the psychological cost for correct representation is not too high.

What is it for a consideration to properly "engage" the function of a type of response then? This has to depend on the specific function of the response type in question. But the rough idea is that each type of attitude or response is constitutively sensitive to some considerations that help it carry out its function, and it is these considerations that are able to properly "engage" its function (I will

explain this in more detail when we come to the functions of different types of response in the following paragraph).[4]

I think the functionalist account plausibly captures the response-focused kind of normativity for the following reasons:

(1) The Extension of Normative Force

On the functionalist account, only considerations capable of properly engaging the function of a type of response can have the normative force in calling for a token response of that type: Belief functions to represent what is the case. It is hence constitutively sensitive to evidence suggesting what holds true in the world. That is why only evidence is able to properly engage the representational function of belief and to exert the normative force on us to believe something. Similarly, intention functions to lead us to realize a prospective state of affairs S that we take to be appealing, and it is therefore constitutively sensitive to the appealing features of S and the feasibility of S (as I argued in Chapter 3). That is why only these considerations are able to properly engage the executive function of intention and to exert the normative force on us to intend to (or not to) bring something about (See Kolodny 2008a: 369). Finally, the emotion of fear functions to help us avoid potential dangers in the environment, and it is constitutively sensitive to features that pose a danger or threat to us. So only such danger- or threat-related considerations can properly engage this function and have the corresponding normative force on us to form a fear response toward something. (Of course, there can be wrong kinds of

[4] Recall that by "response" I mean something that we have voluntary control over and that we can form in reaction to considerations with normative significance. That is why the functionalist account does not entail that there is a reason with the normative force in calling for, say, the constriction of the pupil.

reasons for these responses—I will deal with these later. As I will explain, wrong kinds of reasons are actually considerations with normative significance but lacking the normative force in calling for the responses that they prescribe.)

Thus, on the functionalist account of normative force, the function of a type of response R (say, belief, intention, or fear) determines the range of the considerations that can exert normative force on it: Only the considerations capable of engaging R's function can have the normative force in calling for a token response belonging to R. And this, as we have just seen, comports quite well with our pre-theoretical intuitions about what kinds of considerations can have the normative force in calling for different types of responses. This then provides an initial case for the functionalist account.

(2) The Intension of Normative Force

The functionalist account not only gets right the extensions of the considerations with normative force; it also adequately captures the intension of normative force, that is, it captures how normative force has both a justificatory and a motivating dimension. Roughly speaking, the justificatory dimension of reason consists in how a consideration that constitutes a normative reason can justify a certain response on our part, while the motivating dimension of reason consists in how it can move us and exert some pressure on us in forming the response that it justifies, when we become aware of it.[5]

[5] Of course, since there are various types of response to which the idea of normative reason has applicability (e.g. belief, fear), here I am understanding the idea of motivation in a broader sense, as the psychological force or pressure that a normative reason exerts on us to form a certain response. But this broader notion of motivation applies to normative reason for action as well. This is because we can also say that the motivating dimension of normative reason for action consists in how it can

Now the functionalist account accommodates the justificatory dimension of normative force quite well: Belief, for example, functions to correctly represent what is the case, and that is why truth is the internal standard of correctness for belief. As such, any truth-related considerations capable of engaging the function of belief and suggesting how a particular belief is likely to be true will *ipso facto* be capable of showing how this belief is likely to be correct in light of its functionally-given internal standard. That is why any truth-related considerations are also able to play a justificatory role for belief and thereby have normative force on it. Similarly, fear functions to lead us to avoid potential dangers in the environment, and that is why danger or threat provides the internal standard of correctness for fear, and why fear toward things that do not pose actual dangers or threats are irrational or defective. As such, any danger- or threat-related considerations capable of engaging the function of fear and suggesting how a particular fear response is likely to be directed at a dangerous or threatening object will *ipso facto* be capable of showing how this fear response is likely to be correct in light of its functionally-given internal standard. That is why any danger-related considerations are also able to play a justificatory role for fear and thereby have normative force on it.

More generally, on the functionalist account of normative force, a normative reason's normative force in calling for a certain response consists in its capacity to engage its function. And this kind of force has a justificatory dimension just because the function of a response type R sets an internal standard of correctness for its response tokens (cf. Wedgwood 2017: ch.9), such that any considerations capable of engaging R's function will also be able to show how a certain response token belonging to R is likely to be correct in light of its functionally-given internal

---

exert some psychological force or pressure on us to perform an action.

standard.

The functionalist account also explains why normative force has a motivating dimension as well: Belief, for instance, is constitutively sensitive to truth-related considerations given its function. That is why the considerations suggesting p's being the case will also be able to move us and make us feel the pressure in believing that p. Similarly, fear is constitutively sensitive to danger- or threat-related considerations given its function, and that is why the considerations suggesting an object's being dangerous, or a situation's being threatening, will also be able to move us and make us feel the pressure in forming a fear response toward it. Again, on the functionalist account, a normative reason's normative force can be said to have a motivating dimension just because any considerations capable of engaging the function of a response type R will *ipso facto* be able to exert some pressure on us and make us feel inclined to form a certain response token belonging to R—After all, this is what we would naturally do if we have a proper-functioning system for forming that type of response.

(3) The Phenomenology of Normative Force

What is more, the functionalist account explains very well the phenomenology of the normative force involved in different types of normative reason. So normative reasons for belief make us feel *convinced*; normative reasons for intention make us feel *attracted*; and normative reasons for fear make us feel *scared*. These normative reasons have different phenomenological feelings just because their normative force consists in their capacity to engage the functions of these different response types, and the engagement of them will for sure result in different kinds of

phenomenology.

(4) The Right and the Wrong Kind of Normative Reason

Finally, the functionalist account of normative force provides a simple and elegant distinction between the right and the wrong kind of normative reason. Pragmatic, non-truth-related considerations are the wrong kind of reasons for believing something, just because while they carry certain normative significance, they are nonetheless incapable of properly engaging the representational function of belief. Similarly, the consideration that a rich person will reward you with one million dollars if you form a fear response to a fluffy teddy bear is the wrong kind of reason for fear. This is because this consideration, though it has certain normative significance for you, is still not able to properly engage the function of fear in helping us avoid potential dangers or threats. More generally, on the functionalist account, we can say that the wrong kind of reason is a kind of consideration that (1) has certain normative significance itself but that (2) fails to have full-fledged normative force for the very response that it calls for, due to its inability to properly engage its function. Feature (1) explains why the wrong kind of reason can still be said to carry some normativity and be able to play a counting-in-favor role despite its being of the wrong kind, while feature (2) accommodates our intuition about how the wrong kind of reason is, in a sense, not really normative for the response that it calls for. In a way the wrong kind of reason is wrong just because it picks out a consideration with normative significance but then calls for a response *in a way that is contrary to its functional nature*, given that the very consideration it picks out is simply unable

to engage the function of this response.[6]

In fact, the functionalist account provides a good explanation for the basic features of the wrong kind of reason. As Mark Schroeder has noted (2012: 458-61), the wrong kind of reason for a type of attitude A is often more difficult for us to follow (see also Raz 2011: 36-41). Furthermore, following it compromises our rationality in holding A and violates the internal standard of correctness for A (see also Broome 2013a: 94-5). Now the wrong kind of reason is more difficult to follow just because it is at bottom a consideration that calls for a particular response in a way that goes against its functional nature (just consider pragmatic reasons for belief). Moreover, following the wrong kind of reason compromises our rationality in forming a particular response just because following it requires us to disrespect its functional nature, and to do so is to defy the internal standards of correctness given by its function.

Furthermore, the functionalist account also marks the proper boundary between the right and the wrong kind of reason in more peripheral cases. According to the object-/state-given approach (Parfit 2001, 2011, app. A.; Piller 2006), the wrong kind of reasons for a particular attitude A are state-given reasons related to the consequences of having A, while the right kind of reasons for A are object-given reasons related to A's object. However, as Schroeder has persuasively pointed out, some state-given reasons can actually be the right kind of reasons for an attitude

[6] The functionalist account is actually a variant of the constitutivist approach to the right/wrong kinds of reason distinction, according to which the wrong kind of reason for a particular attitude A is a kind of consideration that speaks in favor of A yet is unrelated to the constitutive concerns of A. For this kind of constitutivist proposal, see Hieronymi (2005, 2013) and Schroeder (2010, 2012). I personally prefer the functionalist account because the idea of function gives our constitutive talk a more naturalistically- respectable foundation. Also, as we shall see later, the functionalist account also helps to adjudicate the debate among the constitutivists themselves, for some disagreements among themselves actually arise only because they have different views on the central function of a particular type of attitude.

(2012). For instance, if I can decide now whether to visit my friend next month and form a corresponding intention to do so, then I will be able to let him know now whether I will really visit him later, thereby helping him to make further arrangements in advance. This consideration can be said to be the right kind of reason for me to form an intention now to visit my friend, even though it is actually a state-given reason derived from the effect of my intention in facilitating our interpersonal coordination. Now, on the functionalist account, we can perfectly say that this consideration is the right kind of reason for intention, for it is still able to properly engage another central function of intention, which is to help us to settle on a course of action in order to facilitate interpersonal and intrapersonal coordination (see Bratman 1987: ch. 2).

Of course, we can also follow Pamela Hieronymi and object that the coordination consideration that we have just considered, given that it doesn't really concern the substantive merits or demerits of visiting my friend, still cannot be the right kind of reason for me to intend to visit my friend. It can at most be the right kind of reason for me to get the decision made now (Hieronymi 2013: 122-5). But this kind of verdict can also be perfectly accommodated by the functionalist account. After all, if you have a stricter view about the function of intention and think that its core function is to lead us to realize a prospective state of affairs that we take to be appealing, then you can still say that coordination considerations still cannot properly engage this core function and hence cannot be the right kind of reasons for intending to do something. So the functionalist account actually explains quite well why people's intuitions about the right and the wrong kind of reason would differ in these kinds of peripheral cases.

In sum, the functionalist account adequately accommodates our pre-theoretical

intuitions about what kinds of considerations can claim the normative force in calling for different types of responses. It also explains how the normative force of reason has both a justificatory and a motivating dimension. What is more, it suitably captures the phenomenology of the normative force of different kinds of normative reasons. Finally, it correctly marks out the distinction between the right/wrong kinds of reason. That is why I think it presents a plausible analysis of normative force.

Two things should be noted here. First, although the functionalist account explains normativity in terms of the functions of different types of attitudes or emotional responses, it is not a variant of the fitting-attitude analysis of normativity. This is because the latter is typically offered as an analysis of non-response-focused normativity in terms of response-focused one (say, what is of value is just what would be appropriate for us to desire or what we have a normative reason to desire (Rabinowicz and Rønnow-Rasmussen 2004; Rønnow-Rasmussen 2011; Scanlon 1998: 96-7; Rowland 2016)). The functionalist account, in contrast, is an analysis of response-focused-normativity in terms of the idea of function.

Second, as a meta-normative view, the functionalist account is actually a non-subjectivist account of normative force. This is because on this view, the normative force of reason consists in its function-engaging power. But whether a certain consideration has such a power is not of the agent's own making; rather, it is determined by the functional nature of the very response that is called for. It is just that some types of response, given their functions, are constitutively sensitive to our subjective states. For instance, if you think that one of desire's functions is to fulfill what we already desire as an end or to track what can bring about pleasure within us, then only considerations related to end-fulfillment and pleasure-acquisition

would be able to properly engage its function. These would then be the considerations with the normative force in calling for a desire on the functionalist account. But we would instead end up with a non-subjectivist picture about normative reasons for desire if we think that the function of desire is to track what is objectively good. In other words, whether the functionalist account would entail a subjectivist or an objectivist view about first-order normative reasons actually depends on the function of the response in question. As I will argue later, action is in fact a type of response whose function is typically engaged by intentions. Thus, the functionalist account of normative force will in fact lend support to a version of internalism about normative reasons for action. I will develop this new line of argument for reasons internalism and explain how intentions can be normative for our action later in Section 5. But here I shall first address the question about the normativity of practical rationality.[7]

4. The Normativity of Practical Rationality

As we saw in the last chapter, the normativity of rationality has been called into doubt in recent literature. An agent can always ask, "Why do I have a normative reason to follow rational principles and to maintain the psychological patterns they prescribe?" (Kolodny 2005). It has also been pointed out in the literature that an agent can ask the same question about constitutive principles for agency: "Why do I

[7] As you can see here, the functionalist view actually allows for a hybrid view about first-order normativity (cf. Chang 2013a, 2013b). Perhaps given the function of action in fulfilling an end of ours, only internalist considerations can properly engage its function and thereby have the normative force in calling for an action. But it might turn out that our desire functions to track what is objectively good in the world, such that externalist considerations might still be able to engage the function of desire and claim the normative force in calling for a desire. In other words, we could have both internal and external normative reasons on the functionalist picture. It's just that they will be reasons for responses with different kinds of functions.

have a normative reason to follow the constitutive principles for agency or those for a certain type of attitude? Why can't I be content with being a schm-agent or having a schm-attitude?" (Enoch 2006). These two challenges actually merge into one on the functionalist picture of practical rationality that I defended in the last chapter. This is because on this picture, the requirements of practical rationality are just the constitutive principles for agency, i.e. principles prescribing the psychological steps and patterns constitutively involved the proper function of our agency. So, as you might similarly wonder, even if there are some rational constitutive principles for having proper-functioning agency or a proper-functioning attitude, why do we have a normative reason to follow them? Why are they normative (Scanlon 2007: 86)?[8]

This scepticism about rational constitutive principles, however, seems ill-motivated in many cases. For instance, if an agent refrains from following the rational constitutive principle for the proper functioning of her belief system (i.e. the norm of truth), then no normative reasons for belief can probably have a hold on her: Even if we present evidence to show to her why p is the case, she can simply respond, "Yes, I can see why p is the case, but I simply don't care about believing what is true." Similarly, if the agent rejects the rational constitutive principle for the proper functioning of her executive system (i.e. the instrumental principle), then few normative reasons for action can sway her: Even if we show her how some of her ends might be promoted by an act φ-ing, she can similarly respond, "Yes, I can see how φ-ing promotes my ends, but I simply don't care about acting to promote my

[8] Notice that this is not how Enoch and Kolodny originally conceive of their challenges. Kolodny's main argument is that following rational principles don't seem to help us better conform to what we have reason to do, while Enoch's major complaint is that we cannot ground normativity in constitutive features of agency.

end."[9]

Of course, on the functionalist account, the normative force of a normative reason consists in its function-engaging power, and I have just pointed out that this kind of force holds regardless of the agent's acceptance or denial. For instance, even if the agent doesn't care about the proper functioning of her belief system, truth-related considerations still possess the normative force in calling for a belief on her part. However, what we can see here is that, for an agent who refrains from following the rational constitutive principles for belief and agency, few normative reasons can really get under her skin—and the functionalist account correctly predicts this. This is because, on this view, normative reasons for belief or for action are supposed to exert their normative force on us *by engaging the functions of our belief or agency* and thereby calling for a token belief or action on our part. So if an agent simply fails to follow the constitutive principles for the proper functioning of these two systems, then she probably also won't be susceptible to the influence of the normative force from reasons for belief or for action. This is what the functionalist account can allow us to say.

With this in mind, we can then make the following move. We can first argue that if one wants to make a meaningful request for some normative reasons to follow rational constitutive principles, then one has to remain open to the influence of potential normative reasons for belief or for action that might potentially justify this—otherwise one wouldn't be different from a sceptic who is disposed to deny the force of all kinds of normative reasons that we might offer. But we can then argue that if one is to be receptive to the force of potential normative reasons for

[9] A similar observation is also made by James Dreier (1997), who argues that those who don't accept the instrumental principle cannot be subject to the influence of instrumental reasons at all.

belief or for action, then one at least has to follow the norm of truth and the instrumental principle. In other words, either we have to follow rational constitutive principles for belief or for agency at some level in order to be susceptible to the influence of any normative reasons, or we have to risk becoming complete sceptics who are beyond the influence of normative force.

Thus, the functionalist account accommodates the normativity of the principles of practical rationality in the following way. We have no normative reason to maintain the proper function of our agency by following these principles just for proper function's own sake. The sceptics are right about this. However, given that principles of practical rationality are constitutive principles for the proper functioning of our agency, those who don't accept and follow them also won't be receptive to the normative force of a special kind of reason, which exerts its normative force by engaging the function of our agency, i.e. normative reason for action. Indeed, those who don't even follow the principles of instrumental rationality, feasibility, and non-contradiction also won't be susceptible to the normative force from instrumental reasons, reasons against an intention based on its infeasibility, and reasons against an intention based on its undermining or frustration effects on other intentions. This then provides a distinctive functionalist answer to the question about the normativity of practical rationality: We cannot find an independent normative reason to follow principles of practical rationality, but our acceptance of them is still the precondition for our receptivity to various types of normative reasons for action. The functionalist account hence allows us to say that practical rationality and constitutive principles of agency are normative for us in this conditional sense.

Of course, I am not suggesting that this is *the* answer to the normativity of

practical rationality, or that this can leave the sceptics satisfied. But the functionalist view of practical rationality and normativity at least shows how the principles of practical rationality are already implicated in our susceptibility to practical normative force, such that we cannot wholly distance ourselves from them and ask the normative question. So the functionalist account, I believe, at least points to an alternative approach to the normativity of practical rationality that is worth exploring.

5. The Normative Force of Normative Reasons for Action

I have just alluded to the idea that the normative force of reason for action consists in its capacity to engage the function of our action and agency. This then naturally leads us to the question about the normativity of intention. So here I shall explain how the functionalist account lends support to both internalism about normative reasons for action and the normativity of intention. In order to show this, I will consider three questions in turn: Does human action have a central function, like the various types of attitudes or emotional response that we have considered? (The answer: Yes. It is to fulfill our ends). Can there be any wrong kind of reason for action, a kind of consideration with normative significance but incapable of properly engaging the function of action? (The answer: Yes. It is an external reason for action wholly unrelated to the fulfillment of our ends). What is the right kind of reason for action? What is its normative force like? (The answer: It is a consideration related to the fulfillment of our intentions or volitional states, and its normative force consists in its capacity in settling our first-person practical question concerning what we are to do).

Let us start with the first question:

(1) What is the function of human action?

Roughly speaking, an action is minimally a bodily movement that is supposed to serve an end by changing the world in one way or another. An end imbues *purposiveness* to a bodily movement and elevates it to the status of an action by rendering it as something that is done for the sake of something else.

But an action does not merely function to promote an end. After all, brute instinct can lead me to recoil away from the hot handle of a pot on the stove simply to avoid possible danger. A hypnotist can put me under hypnosis and prompt me to pick up a glass of water just to impress the audience with his skill. None of these seem to be actions of mine because the ends that they serve are simply given by nature, or by other intentional beings. In order for one's bodily movement to count as an action (or more precisely, an intentional action), the end it serves needs to be provided by an intentional state that one possesses and that has a world-to-mind direction of fit. After all, our bodily movement typically has causal effects on the world and changes the world in one way or another. And an intentional state with world-to-mind direction of fit can naturally imbue *intentionality* to a bodily movement by rendering it as something that is done in order to fulfill an intentional state that calls for a corresponding change in the world. In other words, an end not only has to make a bodily movement purposive; it also has to make a bodily movement intentional by revealing how it is done because one wants the world to be in a certain way.

This characterization of action here is modelled on the teleological conception

of action, which boasts a venerable history dating back to Aristotle, and which has been accepted by virtually all contemporary philosophers of action in one form or another. Since this teleological conception defines action in terms of the idea of end, it trivially follows from this conception that action has a defining constitutive function, namely, to fulfill an end that we have. I take this to be the most minimal and the least controversial view about the core function of human action. With this conception of the function of action in mind, we can turn to our second question:

(2) Can there be any wrong kind of normative reason for action?

The answer is yes. As we saw earlier, wrong kinds of reasons are in fact considerations with certain normative significance yet incapable of properly engaging the function of the response that they call for. In line with this view, the wrong kind of reason for action should presumably be a consideration that is normatively important in itself but unable to properly engage the function of action in fulfilling an end for ours. So presumably it should be a consideration with certain normative significance yet bearing no connection whatsoever to the fulfillment of our current ends and that of the ends that we might possibly adopt after we deliberate on the matter.[10]

This kind of normative reason for action is actually what has been labelled "external reason" in the literature. Indeed, I think external reason for action has precisely the same structure as other wrong kinds of reasons: An external reason for

[10] Another type of wrong kind of reason for action is a consideration that has certain normative significance but asks us to act without having its promotion or the promotion of any other ends explicitly in mind (see Raz 2011: 50). A normative reason would call for an action in this way when our thought about the promotion of an end will undercut an action's promotion effects.

action also calls for an action in a way that is in tension with its functional nature in fulfilling an end of ours, just as a non-truth-related reason for belief calls for a belief in a way that is in tension with its functional nature in representing what is in fact the case. Furthermore, external reason for action also shares the basic features of the wrong kind of reason identified by Schroeder: An agent would have difficulty in performing an action for an external reason that is completely unrelated to her current ends and the ends that she might possibly adopt,[11] for she simply cannot see the point of doing so and cannot count the resulting action as something that is distinctively hers. Furthermore, following an external reason also requires the agent to violate the internal constitutive rational standard for action, which has to do with how an action can effectively fulfil an end that she has. In fact, this point has been precisely highlighted by Bernard Williams's remark that "the only rationality of action is the rationality of internal reasons (1982: 111)".

It is important for us to distinguish between the wrong kind of reason and other "ersatz" normative reasons, that is, normative reasons that pick out the very same consideration but call for other types of response in a way that is in conformity with their functional nature. For instance, the consideration that a vicious rich man will inflict extreme pain on an innocent person unless you believe that the earth is flat is the wrong kind of reason for *believing* so, but the right kind of reason for *wanting* it to be the case that you believe so. Similarly, we should also distinguish between external reasons for action and other ersatz normative reasons for other kinds of response. If the fact that someone is in pain, for instance, is simply unrelated your current ends and the ends that you might possibly adopt, then it

[11] Here I shall leave it open whether we can possibly adopt as an end something that bears no connection whatsoever to our current ends.

would be the wrong kind of normative reason for *action* for you, though it may still be the right kind of normative reason for *adopting an end* or for *desiring*. After all, action functions to serve an end that we have, but this might not be the case for adopting an end, unless you embrace a Humean picture about how we come to adopt an end (Williams 1982). In fact, it might even be the case that our desire functions to track what is objectively good (Railton 1986a, 1986b, 2012: 40-1; Stampe 1987), or what we believe to be good (Davidson 1969; Mele 1992: ch.12; cf. Scanlon 1998: 18-22, 2007: 92-6), or at least something quite close to that (say, our belief about what our idealized selves would desire (Smith 1994: ch.5, 2015)). The topic of the function of desire is beyond the scope of this chapter, and I would be happy to adopt a non-Humean picture about this. My point, however, is that it can be perfectly all right to say that those who don't care for morality still have a normative reason to *adopt ends* or to *desire* in a way that takes moral considerations into account. What is problematic, rather, is the claim that those who don't care for morality have a normative reason to *act* for moral considerations, even when these considerations bear no connections to their current ends or the ends that they can possibly adopt—for this kind of externalist claim actually calls for an action in a way that is in tension with its functional nature in serving an end that one has. I think this is why Williams thinks that externalists are still justified in claiming that those who do not comply with external reasons are "inconsiderate, or cruel, or selfish, or imprudent; or that things, and he, would be a lot nicer if we were so motivated (1982: 110; cf. Williams 1995: 40)." All these claims concern evaluations with different constitutive standards. But when it comes to the internal constitutive standard of action provided by its very own function, it would indeed be problematic to ask people to act for considerations completely out of touch with

their ends.

Can we say that an external reason is actually a "compound" normative reason that prescribes us to adopt or desire a certain end E and then to act so as to promote E? In a way I think this is the most sensible way to understand the claim made by externalists. However, as far as normative reason for action goes, this kind of compound normative reason still involves an internal reason for action. This is because this kind of compound normative reason still prescribes us to perform an action to promote an end E which, as externalists assume, we would adopt if we were to follow the normative reason to adopt E as an end. Again, I think this is why Williams contends that externalists can sometimes be seen as "entertaining an optimistic internal reason claim (1982: 111; cf. Williams 1995: 41)", a claim about how there would be a normative reason for an agent to act so as to promote the end that externalists prescribe, *once the agent comes to adopt it*. But if externalists' claim is that the agent has a normative reason to perform an act even though it bears no connection to her ends whatsoever, then their claim would indeed be, in Williams's words, a bluff.

Once we make it clear how external reasons for action are distinct from other ersatz normative reasons for other types of responses, as well as from optimistic internal reasons, a more general argument for reasons internalism emerges. According to our functionalist picture, (1) the normative force of a normative reason in calling for a response R consists in its capacity to engage R's function. Any normative reason that is unable to do so and calls for a response in a way that is contrary to its functional nature will have defect in its normative force and be the wrong kind of reason. But according to the teleological conception of action, (2) the function of action is to serve an end that the agent has, the very thing that imbues

purposiveness and intentionality to the agent's bodily movement and makes it an action. What follows from these two ideas is then a form of internalism about normative reasons for action, for only internal reasons are able to properly engage the function of action in fulfilling our ends and to have full normative force in calling for an action on our part.

This argument, I think, is more modest than other traditional arguments for reasons internalism. After all, I have left open the question about the function of desire. Indeed, if its function is not to just extend our intrinsic desires but to track external reason-giving features in the world, then there might still be external reasons for desire. As such, my argument here can allow moral considerations to provide external normative reasons for desire and to indirectly affect what (internal) normative reasons for action we have. More importantly, this argument is not meant to show that external reasons for action are completely non-normative. In fact, I have already conceded that some considerations can carry normative significance in virtue of mind-independent normative facts. So external reasons for action can still have a kind of non-response-focused normativity, as things of value do. What my argument here is supposed to show is instead that external reasons have a defective normative force on the very response that they call for (i.e. action), and they have a similar problematic structure as other wrong kinds of reason. So external reasons are actually defective when it comes to the response-focused normativity involved in them. And the burden of proof is on externalists to explain why external reasons can still have normative authority over us despite their problematic structure and defective normative force, and furthermore why we should accept external reasons for action while rejecting other wrong kinds of reason.

This argument, on the other hand, may nevertheless have a stronger grip on those with an externalist tendency than other traditional arguments for internalism, or at least I hope so. After all, the resulting internalist picture supported by my argument is still compatible with normative realism, for it recognizes how external reasons for action can carry normative significance on their own, and how there might even be external normative reasons for desire. More importantly, traditional internalist arguments are typically premised on certain ideas that more or less express the internalist intuition, say, that normativity needs to have a psychological grip on us (Korsgaard 1996: ch.1), or that normativity has an explanatory dimension (Williams 1982: 102, 1995: 38-9). The typical reaction from externalists to these claims is to simply deny them and the internalist intuitions undergirding them (Parfit 2011a: 65-70, 2011b: 413-25). My argument here, however, relies only on two ideas that can be supported on independent grounds, i.e. the functionalist account of normative force, and the teleological conception of action. And these ideas at least seem to be something that most externalists would not find particularly question-begging, or at least I hope so.

Of course, externalists can still try to deny or restrict either the functionalist account or the teleological conception in order to allow for external reasons for action. Thus, they might claim that it is perfectly all right for a normative reason to call for an action in a way that is in tension with its functional nature in fulfilling an end that we have. After all, it seems a lot easier for us to act for considerations unrelated to our ends than to believe for non-truth-related considerations. However, as I have just noted, we have to distinguish external reasons for action from ersatz normative reasons for other types of responses, as well as from optimistic internal reasons, which prescribe us to perform an action once we adopt the ends that they

call for. An external reason for action is supposed to pick out a consideration completely unrelated to our current ends and any ends that we might adopt after deliberation, and it asks us to perform an action on that basis. Is it really easier for us to follow this kind of reason than to follow a non-truth-related reason for belief?

Externalists might alternatively contend that the teleological conception of action does not entail that an action always has to serve an end that the agent has. Perhaps as long as the agent can see how an action will serve a purpose and can then set out to act on the basis of it, the resulting bodily movement can still count as an action of hers nonetheless. So perhaps in the end external reasons don't really call for an act in a way that compromises its functional nature in serving an end. However, even if the action required by external reasons can retain its purposive and intentional character, I presume it would nonetheless be something that the agent cannot take full ownership of. In fact, we can say that an end needs to imbue not only purposiveness and intentionality, but also *attributability* to a bodily movement (see Korsgaard 2009: ch.5, cf. Frankfurt 1971, 1977). Only if this is so can an end render a bodily movement *the agent's own* action. And a deeper problem with external reason is that the action that it prescribes, in so far as it has to be done for the sake of a consideration wholly unconnected to our ends, is something that will lack full attributability to us and that we will feel alienated from. It therefore cannot be a full-fledged action of ours (see Williams 1973: 116, cf. Railton 1986a). This fuller characterization of the idea of an end then enables us to answer the last question, namely:

(3) What is the right kind of reason for action? What is its normative force like?

I have mentioned that an end needs to bring purposiveness, intentionality, and attributability to an action. These are the features that render a bodily movement a full-fledged action of our own. Now the final feature of attributability here adds a further constraint on what things can provide the sort of end that makes a bodily movement a full-fledged action: This sort of end needs to be provided not only by an intentional state with a world-to-mind direction of fit, but also by an intentional state that can fully represent us. Only if a bodily movement serves an end like this can it be a full-fledged action of our own. This final constraint then suggests that an end is supposed to be provided by our intention or volitional state—after all, as we saw in Chapter 2, intention or volitional state is an integrated practical state that best represents our attitude as a whole with respect to our practical question. So it is precisely the type of intentional state able to imbue attributability to an action. Therefore, I think we should say that an action functions to fulfill not just an end, but more specifically an end given by our telic intention.

Given this understanding of the function of action, we should say that the normative force of a normative reason for action consists in its capacity to engage the function of action in fulfilling an end provided by our intention. Also, in light of our foregoing analysis of the right kind of reason, we ought to say that the right kind of reasons for action are considerations which have certain normative significance, and which are capable of engaging the function of action in fulfilling our intentions. Thus, the right kind of reasons for action ought to be thought of as considerations related to our intention-fulfillment. This view, I believe, comports quite well with Christine Korsgaard's characterization of the normative force possessed by normative reasons for action. Consider the following passage in which she raises the

famous normative question:

> The normative question is a first-person question that arises for the moral agent who must actually do what morality says. When you want to know what a philosopher's theory of normativity is, you must place yourself in the position of an agent on whom morality is making a difficult claim. You then ask the philosopher: must I really do this? Why must I do it? And his answer is his answer to the normative question (1996: 16).

The work that a normative reason for action is supposed to do, according to Korsgaard, is to present the solution to this kind of first-person practical question. As she says, "normative concepts are not […] the names of objects or of facts or of the components of facts that we encounter in the world. They are the names of the solutions of problems (2003: 116)". In fact,

> The normative word 'reason' refers to a kind of reflective success. If 'good' and 'right' are also taken to be intrinsically normative words, […] then they too must refer to reflective success. And they do. Think of what they mean when we use them as exclamations. 'Good!' 'Right!' There they mean: I'm satisfied, I'm happy, I'm committed, you've convinced me, let's go. They mean the work of reflection is done (1996: 94).

In other words, a normative reason to φ, as the solution to our first-person practical question about what we should do in the future, is something that is supposed to successfully settle this practical question, and to leave us satisfied, convinced, and

committed to φ-ing without further questioning, "Must we really φ?". Now the main problem with normative realism, as Korsgaard notes, is that the supposed normative reasons grounded in mind-independent normative facts cannot satisfactorily settle our first-person practical question in this way.

> Suppose you are being asked [by morality] to face death rather than do a certain action [that you like]. You ask the normative question: you want to know whether this terrible claim on you is justified. Is it really true that this is what you must do? The realist's answer to this is simply 'Yes'. That is, all he can say is that it is true (or there is a fact) this is what you ought to do (1996: 38).

On Korgaard's view, any normative reason for action should have a kind of normative force that is capable of having a grip on our psychology, settling our first-person practical question and leaving us satisfied without further questioning, "Must we really do this?"

But our intention or volitional state just seems to be the kind of intentional state with such a psychological grip. As we saw in Chapter 2, our intention represents an integrated practical attitude of us as a whole, carries our all-out commitment to a particular course of action, and settles our practical question concerning what we are to do. So if you really intend to φ, then it typically signifies that you have already settled on φ-ing and committed to φ-ing in the future. And you probably will not further question, "Must I really φ?". After all, if this question is still open for you, and if you are not really satisfied, then perhaps you have not really settled your practical question and formed an intention to φ in the first place. Of course, I am not saying that all intentions or volitional states can fully settle our

first-person practical question and leave us satisfied. Perhaps only a subclass of them will be able to do so. Perhaps they are those formed after our critical reflection; perhaps they need to be backed by our evaluative attitudes; perhaps they should carry our commitment to some long-term projects. I am not able to go into detail here. But I think we can at least say that a subset of our intentions or volitional states can possess the required force to settle our first-person practical question and leave us satisfied, and that the considerations related to their fulfillment can also have a similar force over us.

This then suggests that a volition-based account of normative reasons for action might present the most promising subjectivist analysis concerning the normative force of reasons for action. After all, the functionalist account of normative force, and the most plausible contemporary conception of action (both of which can be supported on independent grounds), together entail that the normative force of reason for action is supposed to consist in its capacity in engaging the function of our action in fulfilling an intention of ours. Furthermore, volition-based reasons seem to have the kind of normative force that Korsgaard describes. Considerations related to the fulfillment of our intentions (or least a subset of them) are capable of not only engaging the function of action, but also addressing our first-person practical question without leaving us further questioning, "Must we really do this?" Therefore, intention or volitional state does seem to the subjective state that can best capture the normative force of reason for action. What else can move us, have a psychological grip on us, and leave us thinking "I'm satisfied, I'm happy, I'm committed, you've convinced me, let's go", if not our intentions or volitional states?

This then answers the question about the normativity of intention. Of course, I

am not saying that all intentions or volitional states are normative for our action. Perhaps only some of them are. What specific kinds of volitional states can best ground normative reasons for action is an issue that I must leave open here. But what we can see is that the functionalist account at least provides an interesting and novel way for us to understand reasons internalism while at the same time lends support to volition-based accounts of reasons for action.

6. Conclusion

I have explained here how the concept of normative reason involves a kind of response-focused normativity, and why the functionalist account of normative force presents a plausible view about the nature of this kind of normativity. This account adequately accommodates our intuitions about the extension of the normative reasons for different types of responses. It also explains how normative force can have both a justificatory and a motivating dimension, while also correctly capturing the phenomenology of different kinds of normative reason and marking out the distinction between the right/wrong kinds of reason.

The functionalist account then points to a new way to secure the normativity of practical rationality. As I argued in the preceding chapter, principles of practical rationality are grounded in the proper function of our agency. Furthermore, as the functionalist account of normative force suggests, normative reasons for action are supposed to exert their normative force on us by engaging the function of our agency and action. Thus, what follows from these two ideas is that our accepting and following the principles of practical rationality is a precondition for our susceptibility to the normative force of normative reason for action. So although we

do not have an independent normative reason to follow these principles just for the sake of the proper function of our agency, they are still normative for us in a conditional sense.

Finally, the functionalist account of normative force and the teleological conception of action together provide a new line of argument for reasons internalism, and they in turn lend support to volition-based accounts of reasons for action. But does this mean that we have no normative reason to be moral? No. Although the justification for morality is beyond the scope of this thesis, I think there are at least two ways for us to accommodate moral reasons in the overall functionalist account of normativity that I have developed here.

One way to do this is to explain how moral or evaluative considerations can still play a role in influencing our intentions and thereby indirectly determining our normative reasons for action. Perhaps our intentions are typically formed in light of our desires, whose central function is to track the good, either the good construed in a non-naturalist or a naturalist sense. Or perhaps our intentions are governed by the Enkratic principle and are therefore sensitive to our normative beliefs, which likewise function to track the good. So although normative reasons for action are provided by our intentions or volitional states on the functionalist picture, the formation of these states might still be sensitive to evaluative and moral considerations in the end.

Another way for us to accommodate moral reasons on the functionalist picture is to re-conceptualize "the reason to be moral". Most contemporary philosophers seem to have taken this to be a normative reason *for action*, that is, a normative reason for acting in accordance with moral requirements. But this way of conceptualizing the justification of morality may have already led us to set off on the

wrong foot. After all, the distinctive functional character of action already makes the question "what actions should I do?" constitutively sensitive to instrumentalist considerations, in the same way as the question "what beliefs should I form?" is constitutively sensitive to truth-related considerations. So it is already a fairly infeasible task to prove how a non-instrumentalist consideration wholly unconnected to our ends can somehow justify our acting morally. However, unlike the question "what actions should I do?", the question "what kind of person should I become?" is not constitutively responsive to instrumentalist considerations. Indeed, the point of shaping our character in a particular way is not to fulfill some ends we already have; rather, to become a person with a different kind of character is to have a different set of ends to begin with. Thus, morality might actually be able to better engage us if the central question about its justification is formulated not as a question about *action*, but as one about *character*.

The questions we ask determine the kinds of answers we are likely to receive. Some questions will even foreclose certain types of answer in the first instance. The question about the normative reason for *acting* morally, I suspect, might very well be like this: Action has very a different internal logic and constitutive rational standards, so it is not surprising that morality can scarcely have a hold on it. But this doesn't mean that morality cannot be justified. Rather, this means that we would be better off re-conceptualizing the question about the normativity of morality and trying to ask it in a different way.

Conclusion

1. The Rational Principles for Intention

In the first three chapters of this thesis, I showed how a proper examination of the functional nature of intention lends support to a neo-Humean account of the rationality of intention, a view that goes beyond mere instrumental and formal coherence principles. Intention settles our practical question with its all-out volitional character. This settling function in turn grounds the principles of integration and non-contradiction, which respectively require us to form intentions by first engaging in an all-things-considered integration process, and to avoid contradictory intentions that undermine each other's volitional all-out-ness and settling force. Furthermore, intention leads us to act and to realize the prospective state of affairs that we have settled on and committed to. This executive function then gives rise to the principles of instrumental rationality and feasibility, which respectively prescribe us to take the psychological steps indispensable for our general intention-fulfillment, and to avoid intentions targeting at prospective states of affairs that cannot be realized in the first place. Finally, although space does not allow me to examine formal coherence principles codified in contemporary decision theory, such principles can also be seen as contributing to the proper function of our intention. In fact, given that most of these principles are vindicated by instrumentalist or pragmatic justification, we can reasonably claim that they promote the proper use of our general resources for intention-fulfillment and

indirectly facilitate the executive function of our intention. Therefore, a fuller picture of the functional nature of intention reveals that intention is not governed by instrumental and formal coherence principles alone. The orthodox Humean picture of practical rationality is hence inadequate in so far as it omits other rational principles that are similarly grounded in the proper function of our intention.

But this should not be taken to suggest that the principles that I have just mentioned constitute all of the rational principles governing intention. In fact, the first three chapters of this thesis all leave ample space for further development. In Chapter 1, I adopted a rather minimal characterization of the function of an actual human agent's executive system. There, I claimed that to fulfill our telic intention, we human agents have to consider and select the means to it, form instrumental intentions for the suitable means, and then trigger the psycho-physiological process leading to action by these instrumental intentions. But one thing to notice is that this kind of executive system already presupposes a rational hierarchy among our intentions: Our telic intentions are the main intentions to be carried out, and we are rationally supposed to revise our instrumental intentions in accordance with them, not the other way around. Indeed, one problem with the means-end coherence conception of IR is that it equally recommends the revision of our telic and instrumental intentions and hence fails to accommodate the rational hierarchy among them.

But this kind of hierarchy seems extendable to our telic intentions themselves: Among our telic intentions, some also seem to be the main intentions that boast higher priority in being fulfilled. In fact, if there were no rational hierarchy among our telic intentions, then fulfilling one telic intention by sacrificing another would always reflect a success with the use of our rational executive system—for in doing

so, we still successfully realize *one* telic intention in the end. However, our ordinary judgments about executive success do not seem to support this view. We tend to think that satisfying one telic intention (say, the intention to binge-watch TV series) at the cost of another (say, the intention to finish a term paper on time) can sometimes constitute an executive failure. This suggests that apart from the instrumental principle (which tells us how to fulfill our telic intentions), there are also rational principles that establish a rational hierarchy among our telic intentions by prioritizing the fulfillment of some of them over others.

Furthermore, I explained in Chapter 2 that the setting function of intention presupposes a fair degree of first-person integration. What I focused on there was the *synchronic* settling force of intention: If our intention is to carry our volitional all-out-ness and settle our practical question at the very moment of intention-formation, then it has to be based on a fair degree of psychological integration. However, I left open the question concerning how our intention can have *diachronic* settling force, that is, how our intention can extend its settling force to our future selves, make them remain non-ambivalent in their commitments, and eventually get to be carried out by them. So apart from the integration principle (which asks us to reach synchronic integration in our practical stance), there are perhaps also rational principles that prescribe our future selves to hold onto our intentions (unless there is substantive change in objective circumstances or in their desires), to regard their practical question as being settled by these intentions as well, and to thereby reach a kind of diachronic integration in our practical stance.

Of course, Humeans have their stock answers about these further rational principles. For most of them, the rational principle for prioritizing our telic intentions is nothing more than the *strength principle*, according to which a stronger

motivating state claims a higher priority in being fulfilled. Also, most Humeans have conceptualized the continuation and the implementation of our intention in terms of the cooperation between our past and future selves. On their view, it is rational for our future selves to cooperate with us and to carry out our intention only if (1) our intention still remains in line with what they desire or prefer the most in overall sense, or (2) they have a generally advantageous policy or habit to not redeliberate on our intentions (Bratman 1987, 1998, 1999), or to deliberate in terms of a "plan" that takes our past choice history into account (Machina 1989; McClennen 1990, 1997). What underlies this way of conceptualizing the issue is *the principle of the primacy of the present*, according to which it is rational for our future selves to hold onto and to carry out our intentions only when doing so contributes to their current desires or general interests.[1]

These two principles, however, seem to wholly defer the synchronic settling force of our intention, as well as the power of our executive system, to our future selves' current preferences, interests, and deliberative policies. Indeed, on this view, as long as our future selves re-deliberate on the practical question and detect a change in preferences, it will generally be rational for them to abandon our earlier intentions, even when the change in question is only temporary. Thus, we might think that a further dimension of our agency consists in the capacity to have a practical point of view that is not only *integrated* but also *extended*, a practical point of view that represents us as a whole—both synchronically and diachronically. And

[1] Indeed, even philosophers who argue in favor of sticking with our intentions have more or less embraced this principle. Although they think that sometimes our future selves should rationally refrain from deliberating from their strongest current desires and preferences, their justification for this still lies in how doing so can promote our future selves' *current interests*: Perhaps this kind of deliberative policy or habit can either save their resources, which would be otherwise spent in re-deliberating about what to do (Bratman 1987, 1998, 1999) or make them more stable agents who can carry out more advantageous projects over time (McClennen 1990, 1997).

perhaps our intentions should be based on this kind of practical point of view and be distinguished from both strong desires within a part of us and fleeting desires that only persist temporarily. Correspondingly, our rational executive system should perhaps take this kind of integrated and extended practical attitudes as its main inputs and treat partial or fleeting desires as "noises", even if they might happen to be the strongest at the moment. If we can show that the basic units to be fulfilled by our executive agency are these integrated and extended volitional states rather than the strongest current desires we have, then this would cast doubt on both the Humean strength principle and the principle of the primacy of the present. This can in turn support a new model for diachronic agency and rationality that departs away from the temporally-sliced view of the agent, a model that will potentially challenge the orthodox approach adopted by contemporary decision theorists and economists. And our neo-Humean picture about the rationality of intention here actually lays the groundwork for this interesting alternative model.

2. The Functionalist Picture of Practical Rationality and Normativity

In Chapter 4, I explained how the neo-Humean account of the rationality of intention is underwritten by a more general functionalist conception of practical rationality. This is because the settling and executive functions of intention also form the core functions of our agency in making decisions and in fulfilling them. So the rational principles for intention that we just examined can all be said to be grounded in the proper function of our agency: They are principles prescribing the psychological patterns and steps constitutively involved in the proper function of agency. Indeed, the functionalist conception plausibly explains the grounds of the principles of

practical rationality, as well as their independent, constitutive, and normative character. Ultimately, the functionalist conception shows that practical rationality is embedded in the environment: Basic principles of practical rationality prescribe the mental patterns and steps that are constitutively involved in the proper function of agency, when agency is understood in its most fundamental sense. But what these patterns and steps in fact amount to has to depend on empirical features of the psychological mechanisms that realize our agency, which are in turn shaped by the external environment and our evolutionary history. This then reveals the contingent character of practical rationality and undermines the rationalist view about the nature of rationality.

The functionalist conception of practical rationality finally leads us to the corresponding functionalist account of normative force. In Chapter 5, I spelled out how the concept of normative reason involves a kind of response-focused normativity, that is, a kind of normative force that a consideration has in calling for a certain response on our part. I then argued that such normative force consists in a consideration's capacity to properly engage the function of the response that it calls for. This functionalist account of normative force is plausible, I explained, because it captures both the extension and the intension of the normative force of different types of reason, as well as their phenomenology and the right/wrong kinds of reason distinction. This functionalist account ultimately points to a new line of argument for reasons internalism. This is because the function of our action is to fulfill an end of ours, and only internal reasons can call for an action in a way that respects this functional nature.

Although I lack sufficient space to develop this account fully, the overall functionalist picture I have developed here also paves the way for a volition-based

account of normative reasons for action. As we saw in Chapter 3, the central line of thought behind motivation-based subjectivism is that our favoring attitudes toward the realization of prospective states of affairs, and our agential capacity to fulfill them, can together ground normative reasons for action. In other words, the thought is that normative reasons for action are reasons based on our agential capacity for realizing the prospective. And the most plausible sources for such reasons are our intentions or volitional states. After all, these states are constitutively involved in the proper function of our agency: They represent the answers to the practical question that we have settled ourselves on through our decision-making process, and they in turn provide the typical inputs to our executive process. What else mental states, if not our intentions or volitional states, could better ground agential-capacity-based reasons?

More importantly, the functional nature of intention or volitional state imposes certain internal constitutive constraints on it. And these constraints are indeed what we need for a sensible subjectivist picture of normative reasons for action. To begin, given its settling function, intention is supposed to represent a practical attitude of us as a whole and arise from an all-things-considered deliberative process. Moreover, given its executive function, intention is also supposed to be feasible and based on true beliefs about the world. So volition-based accounts of normative reasons for action need no further rationale for invoking idealized attitudes—for intention is already supposed to arise from an all-things-considered deliberative process that incorporates true beliefs about the world. What is more, given that intention is supposed to be a practical attitude of us as a whole, volition-based accounts can also exclude alien desires which, though backed by strong motivating forces, do not

really represent "where we stand" as a whole with respect to our practical question.[2] Finally, intention already incorporates our commitment and settles our practical question about what to do. Thus, a normative reason for action based on our intention can likewise possess the force to successfully satisfy us and leave us without further asking, "Why must I do this?"

Of course, this is only the beginning for a full-fledge volition-based account of normative reasons for action. We perhaps also need to bring in our evaluative attitudes or indeed elements of critical reflection. Nevertheless, the overall functionalist picture presented here at least explains why volitional states should form the basic building blocks of any plausible subjectivist accounts of normative reasons for action. It hence provides the foundation for us to further develop a more sensible subjectivist picture.

The functionalist picture that I explored in this thesis not only lays the groundwork for a non-temporally-sliced model for diachronic agency and rationality, as well as a volition-based account of normative reason for action. It also raises other interesting questions that can be further pursued: (1.) Are the principles of integration and feasibility also applicable to group agents? (2.) Can liking-based subjectivism provide a plausible analysis of non-response-focused normativity (say, value, good, or well-being)? Can our likings also ground normative reasons for action? (3.) Does the functionalist conception of rationality also hold for theoretical rationality? (4.) Can non-response-focused normativity be reduced to response-focused one? Is it the other way around? Or are they in fact irreducible to each other? (5.) Can the functionalist account of normative force also explain what

[2] For the idea that some attitudes can better represent the agent's practical standpoint, see (Frankfurt 1987: 166; Bratman 2018: ch.4 and ch.7).

normative reasons we have for different types of emotion? These are all important questions worth further exploring, and the overall picture that I have developed here also paves the way for these various lines of inquiry.

3. Final Words

I promised in the Introduction that my overall functionalist picture would present a sensible and interesting alternative to the orthodox Humean view about practical rationality and normativity. Now we can see why this is indeed the case. The functionalist picture is sensible because, on the one hand, the proper function of agency provides a plausible ground for the Humean view about practical rationality and normativity. But on the other hand, the very idea of proper function also prompts us to move beyond the overly restrictive Humean view. Indeed, although I have been sticking with a minimal Humean characterization of the proper function of agency in this thesis and hence arrived at a Neo-Humean account of practical rationality and normativity eventually, my functionalist picture is compatible with both Kantian and realist views about normativity—All we have to do is to build on this picture by giving more substantive characterizations of the proper function of agency, intention, and desire.

The functionalist picture is interesting and significant because it undermines three long-held and widely-accepted views about practical rationality and normativity. It challenges the orthodox Humean view about the rationality of intention, for it shows how there are more rational principles for our intention than mere instrumental and formal coherence principles. Furthermore, it casts doubt on the rationalist picture of rationality by revealing how principles of practical

rationality are not necessary requirements that apply across all possible worlds. Finally, it discredits externalism about normative reasons for action, since it shows how external reasons are considerations with normative significance yet lacking the normative force in calling for an action properly. Thus, I hope to have succeeded in carrying out my bold intention to provide a sensible and interesting view about the nature of practical rationality and normativity.

References

Arpaly, Nomy (2002). *Unprincipled Virtue: An Inquiry into Moral Agency* (Oxford: Oxford University Press).

Brandt, Richard (1979). *A Theory of the Good and the Right* (Oxford: Clarendon Press).

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

———. (1998). 'Toxin, Temptation, and the Stability of Intention', in his *Faces of Intention: Selected Essays on Intention and Agency* (Cambridge: Cambridge University Press): 58-92.

———. (1999). 'Planning and the Stability of Intention', *Minds and Machines* 2 (1999): 1-16.

——— (2009a). 'Intention, Belief, and Instrumental Rationality', in D. Sobel and S. Wall, eds., *Reasons for Action* (Cambridge: Cambridge University Press): 13-36.

——— (2009b). 'Intention, Practical rationality, and Self-governance', *Ethics* 119 (3): 411-443.

——— (2018). *Planning, Time, and Self-Governance: Essays in Practical Rationality* (Oxford: Oxford University Press).

Broome, John (1999). 'Normative Requirements', *Ratio* 12 (4): 398–419.

——— (2005). 'Does Rationality Give Us Reasons?', *Philosophical Issues* 15 (1):321-37.

——— (2007). 'Wide or Narrow Scope?', *Mind* 116 (462):359-70.

——— (2013a). *Rationality through Reasoning* (Chichester: John Wiley and Sons).

——— (2013b). 'Practical Reasoning and Inference', in D. Bakhurst, B. Hooker & M. Little, eds., *Thinking About Reasons: Themes From the Philosophy of Jonathan Dancy* (Oxford: Oxford University Press): 286–309.

Brunero, John (2012). 'Instrumental Rationality, Symmetry and Scope', *Philosophical Studies* 157 (1): 125-40.

Chang, Ruth (2002). 'The Possibility of Parity', *Ethics* 112 (4): 659-688.

——— (2009). 'Voluntarist Reasons and the Sources of Normativity', in D. Sobel and S. Wall, eds., *Reasons for Action* (Cambridge: Cambridge University Press): 243-71.

——— (2013a). 'Commitment, Reasons, and the Will', in R. Shafer-Landau ed., *Oxford Studies in Metaethics: Vol. 8* (Oxford: Oxford University Press): 74-113.

——— (2013b). 'Grounding Practical Normativity: Going hybrid', *Philosophical Studies* 164 (1): 163-187.

Dancy, Jonathan (2000). *Practical Reality* (Oxford: Oxford University Press).

Davidson, Donald (1963). 'Actions, Reasons, and Causes', in his *Essays on Actions and Events*, 2nd ed. (Oxford: Oxford University Press): 3-20.

——— (1969). 'How is Weakness of the Will Possible?', in *Essays on Actions and Events*, 2nd ed. (Oxford: Oxford University Press): 21-42.

——— (1978). 'Intending', in *Essays on Actions and Events*, 2nd ed. (Oxford: Oxford

University Press): 83-102.
——— (1982). 'Paradoxes of Irrationality', reprinted in his *Problems of Rationality* (Oxford: Oxford University Press): 169-88.
——— (1984). 'Expressing Evaluations', in his *Problems of Rationality* (Oxford: Oxford University Press): 19-38.
——— (1985a). 'Incoherence and Irrationality', in his *Problems of Rationality* (Oxford: Oxford University Press): 189-98.
——— (1985b). 'Replies to Essays I –IX', in E. LePore and B. McLaughlin eds., *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell).
——— (2001). *Essays on Actions and Events, 2nd ed.* (Oxford: Oxford University Press).
——— (2004). *Problems of Rationality* (Oxford: Oxford University Press).
Dennett, Daniel (1987). *The Intentional Stance* (Cambridge, MA: MIT Press).
Dorsey, Dale (2017). 'Idealization and the Heart of Subjectivism', *Noûs* 51 (1): 196-217.
Dreier, James (1997). 'Humean Doubts About the Practical Justification of Morality', in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Clarendon Press): 81-100.
Enoch, David (2005). 'Why Idealize?', *Ethics* 115 (4): 759-87.
——— (2006). 'Agency, Shmagency: Why Normativity Won't Come From What is Constitutive of Action', *Philosophical Review* 115 (2):169-98.
Frankfurt, Harry (1971). 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* 68 (1): 5-20.
——— (1977). 'Identification and Externality' in his *The Importance of What We Care About* (Cambridge: Cambridge University Press): 58-68.
——— (1987). 'Identification and Wholeheartedness' in his *The Importance of What We Care About* (Cambridge: Cambridge University Press): 159-76.
Ginet, Carl (1990). *On Action* (Cambridge: Cambridge University Press).
Griffin, James (1987). *Well-Being: Its Meaning, Measurement and Moral Importance.* (Oxford: Oxford University Press).
Heathwood, Chris (2005). 'The Problem of Defective Desires', *Australasian Journal of Philosophy* 83 (4): 487-504.
Hieronymi, Pamela (2005). 'The Wrong Kind of Reason', *Journal of Philosophy* 102 (9):437-57.
——— (2013). 'The Use of Reasons in Thought (and the Use of Earmarks in Arguments)', *Ethics* 124 (1): 114-127.
Holton, Richard (1999). 'Intention and Weakness of Will', *Journal of Philosophy* 96 (5): 241-62.
Hublin, Donald (1991). 'Irrational Desires,' *Philosophical Studies* 62 (1):23-44.
Hume, David (1739-40). *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed.

revised by P.H. Nidditch. Oxford: Clarendon Press, 1975.
Kavka, Gregory. (1983). 'The Toxin Puzzle', *Analysis* 43 (1): 33–6.
Kolodny, Niko (2005). 'Why Be Rational?', *Mind* 114 (455):509-63.
——— (2008a). 'The Myth of Practical Consistency', *European Journal of Philosophy* 16 (3): 366–402.
——— (2008b). 'Why Be Disposed to Be Coherent?', *Ethics* 118 (3): 437–63.
Korsgaard, Christine (1996). *The Sources of Normativity* (Cambridge: Cambridge University Press).
——— (1997). 'The Normativity of Instrumental Reason', in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Clarendon Press): 215-254.
——— (2003). 'Realism and Constructivism in Twentieth-Century Moral Philosophy', *Journal of Philosophical Research* 28 (Supplement): 99-122.
——— (2009). *Self-Constitution: Action, Identity, and Integrity* (Oxford: Oxford University Press).
Lillehammer, Hallvard (2000). 'Revisionary Dispositionalism and Practical Reason,' *Journal of Ethics* 4 (3): 173-190.
Lin, Eden (2019). 'Why Subjectivists about Welfare Needn't Idealize', *Pacific Philosophical Quarterly* 100(1): 2-23.
Littlejohn, Clayton (2018). 'Stop Making Sense? On A Puzzle about Rationality', *Philosophy and Phenomenological Research* 96 (2): 257 -272.
Locke, John (1690). *An Essay Concerning Human Understanding*, edited by Peter H. Nidditch, Oxford: Oxford University Press, 1975.
Machina, Mark (1989). 'Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty', *Journal of Economic Literature* 27 (4): 1622–68.
McClennen, Edward. (1990). *Rationality and Dynamic Choice: Foundational Explorations*. (Cambridge: Cambridge University Press).
——— (1997). 'Pragmatic Rationality and Rules', *Philosophy and Public Affairs* 26 (3):210-58.
Mele, Alfred (1992). *Springs of Action: Understanding Intentional Behavior* (Oxford: Oxford University Press).
Nagel, Thomas (1970). *The Possibility of Altruism* (Princeton, NJ: Princeton University Press).
Nozick, Robert (1993). *The Nature of Rationality* (Princeton, NJ: Princeton University Press).
Parfit, Derek (2001). 'Rationality and Reasons', in D. Egonsson, J. Josefsson, and T. Rønnow-Rasmussen, eds., *Exploring Practical Philosophy: From Action to Values* (Aldershot: Ashgate): 17-39.
——— (2011a). *On What Matters*, Vol.1. (Oxford: Oxford University Press).
——— (2011b). *On What Matters*, Vol. 2. (Oxford: Oxford University Press).
Pietroski, Pual (2000). *Causing Actions* (Oxford: Oxford University Press).

Piller, Christian (2006). 'Content-Related and Attitude-Related Reasons for Preferences', *Royal Institute of Philosophy Supplement* 81(59): 155-82.
Rabinowicz, Wlodek, and Rønnow-Rasmussen, Toni (2004). 'The Strike of the Demon: On Fitting Pro-Attitudes and Value', *Ethics* 114 (3): 391–423
Railton, Peter (1986a). 'Facts and Values', *Philosophical Topics* 14 (2): 5-31.
——— (1986b). 'Moral Realism', *Philosophical Review* 95 (2):163-207.
——— (2012). 'That Obscure Object, Desire', *Proceedings and Addresses of the American Philosophical Association* 86 (2):22-46.
Raz, Joseph (1998). 'Incommensurability and Agency,' in his *Engaging Reason: On the Theory of Value and Action* (Oxford: Oxford University Press): 46-66.
——— (2005). 'The Myth of Instrumental Rationality', *Journal of Ethics and Social Philosophy* 1 (1):1-28.
——— (2011). *From Normativity to Responsibility* (Oxford: Oxford University Press).
Ripstein, Arthur (2001). 'Preferences', in R. Frey and C. Morris, eds., *Value, Welfare, and Morality* (Cambridge: Cambridge University Press): 37-55.
Robinson, Terry and Kent Berridge (2003). 'Addition', *Annual Review of Psychology* 54: 25-53.
Rønnow-Rasmussen, Toni (2011). *Personal Value* (Oxford: Oxford University Press)
Rosati, Connie (1995). 'Persons, Perspectives, and Full Information Accounts of the Good', *Ethics* 105 (2): 303.
Rowland, Richard (2016). 'Reasons as the Unity Among the Varieties of Goodness', *Pacific Philosophical Quarterly* 97 (2):200-227.
Scanlon, Thomas (1998). *What We Owe to Each Other* (Cambridge, MA.: Belknap Press).
——— (2007). 'Structural Irrationality', in G. Brennan, R. Goodin, F. Jackson, and M. Smith, eds., *Common Minds: Themes from the Philosophy of Philip Pettit* (Oxford: Clarendon Press).
Schroeder, Mark (2004). 'The Scope of Instrumental Reason', *Philosophical Perspectives* 18 (1): 337-64.
——— (2007). *Slaves of the Passions* (Oxford: Oxford University Press).
——— (2009). 'Means-End Coherence, Stringency, and Subjective Reasons', *Philosophical Studies* 143 (2): 223–48.
——— (2010). 'Value and the Right Kind of Reason', in in R. Shafer-Landau ed., *Oxford Studies in Metaethics: Vol. 5* (Oxford: Oxford University Press): 25–55.
——— (2012). 'The Ubiquity of State-Given Reasons', *Ethics* 122 (3): 457–88.
Setiya, Kieran (2007). 'Cognitivism about Instrumental Reason', *Ethics* 117 (4):649-673.
Shah, Nishi (2003). 'How Truth Governs Belief', *Philosophical Review* 112 (4):447-482.
Sinhababu, Neil (2011). 'The Humean Theory of Practical Irrationality', *Journal of Ethics and Social Philosophy* 6 (W): 1-14.
Smith, Michael (1987). 'The Humean Theory of Motivation', *Mind* 96 (381): 36-61.
——— (1994). *The Moral Problem* (Oxford: Oxford University Press).

——— (2015). 'The Magic of Constitutivism', *American Philosophical Quarterly* 52 (2): 187-200.
Sobel, David (1994). 'Full Information Accounts of Well-Being', *Ethics* 104 (4): 784-810.
———(2001) 'Subjective Accounts of Reasons for Action', *Ethics* 111 (3): 461-92.
——— (2009). 'Subjectivism and Idealization', *Ethics* 119 (2): 336-52.
——— (2012). 'Parfit's Case against Subjectivism', in R. Shafer-Landau, ed., *Oxford Studies in Metaethics*, Vol. 6 (Oxford: Oxford University Press): 52-78.
Stampe, Dennis (1987). 'The Authority of Desire', *Philosophical Review* 96 (3): 335-81.
Street, Sharon (2012). 'Coming to Terms with Contingency: Humean Constructivism about Practical Reason', in J. Lenman and Y. Shemmer, eds., *Constructivism in Practical Philosophy* (Oxford: Oxford University Press): 40-59.
Sheeran, Paschal and Thomas Webb (2016). 'The Intention-Behavior Gap', *Social and Personality Psychology Compass* 10 (9): 503-18.
Titelbaum, Michael (2015). 'Rationality's Fixed Point (Or: In Defense of Right Reason)', in T. Gendler and J. Hawthorne eds., *Oxford Studies in Epistemology*, Vol. 5 (Oxford: Oxford University Press): 253-94.
Velleman, David (2000). 'On the Aim of Belief', in his *The Possibility of Practical Reason* (Oxford: Oxford University Press): 44-281.
Wallace, Jay (2001). 'Normativity, Commitment, and Instrumental Reason,' *Philosophers' Imprint* 1 (1):1-26.
——— (2004). 'Constructing Normativity', *Philosophical Topics* 32 (1/2): 451-76.
——— (2012). 'Constructivism About Normativity: Some Pitfalls', in J. Lenman and Y. Shemmer eds., *Constructivism in Practical Philosophy* (Oxford: Oxford University Press): 18-38.
Way, Jonathan (2010a). 'Defending the Wide-Scope Approach to Instrumental Reason', *Philosophical Studies* 147 (2):213-33.
——— (2010b). 'The Normativity of Rationality', *Philosophy Compass* 5 (12): 1057-1068.
——— (2011). 'The Symmetry of Rational Requirements', *Philosophical Studies* 115 (2): 227-39.
——— (2012). 'Transmission and the Wrong Kind of Reason,' *Ethics* 122 (3):489-515.
Wedgwood, Ralph (2002). 'The Aim of Belief', *Philosophical Perspectives* 36 (s16): 267-97.
——— (2011). 'Instrumental Rationality', in R. Shafer-Landau ed., *Oxford Studies in Metaethics: Vol. 6* (Oxford: Oxford University Press): 280-309.
——— (2017). *The Value of Rationality* (Oxford: Oxford University Press).
Wieber, Frank, Lukas Thürmer, and Peter M. Gollwitzer (2015). 'Promoting the Translation of Intentions into Action by Implementation Intentions: Behavioral Effects and Physiological Correlates', *Frontiers in Human Neuroscience* 9: 395.
Williams, Bernard (1973). 'A Critique of Utilitarianism,' in J.J.C Smart and B. Williams, eds., *Utilitarianism: For and Against* (Cambridge: Cambridge University Press).

——— (1982). ‘Internal and External Reason’ in his *Moral Luck* (Cambridge: Cambridge University Press): 101-13.
——— (1995). ‘Internal Reasons and the Obscurity of Blame’ in his *Making Sense of Humanity* (Cambridge: Cambridge University Press): 35-45.
Velleman, David (2000). 'On the aim of belief', in his *The Possibility of Practical Reason* (Oxford: Oxford University Press): 244-81.