PRACTICAL REASON AND
THE MYTH OF THE GIVEN

by

Emer Mary O'Hagan

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy,
Graduate Department of Philosophy,
University of Toronto

© by Emer Mary O'Hagan 2001
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-63747-6
Abstract

Practical Reason and the Myth of the Given
Doctor of Philosophy, 2001
Emer Mary O’Hagan
Department of Philosophy
University of Toronto

My thesis argues that the debate about the nature of practical reasoning is hindered by an erroneous distinction between theoretical and practical reasoning which, when redrawn, allows for a better account of the normative force of practical reasons.

Typically, reasoning about action is taken to bear explanatory burdens which reasoning about belief is not. Rational constraints are thought to be categorically binding in theoretical reasoning, but only hypothetically binding in practical reasoning. Instrumentalists, who hold that reason’s role in action is limited to the generation and evaluation of means to ends, ends set by something other than reason itself, cannot appeal directly to reason’s authority in the explanation of how reasons rationally compel action. Instead, they focus on the question of what motivates action. This focus obscures the more fundamental question, ‘how are judgements, practical and theoretical, normatively binding?’ It prejudices an analysis of practical reason by supposing it to involve normative commitments which theoretical reasoning doesn’t.

I outline Wilfrid Sellars’s analysis of the normative commitments implicit in theoretical
reasoning and apply its lessons to an analysis of practical reasoning. In Sellars's terminology, attempts to justify by appeals to non-normative facts employ "mongrel concepts," which illegitimately conflate the causal and normative orders of explanation. Using this concept as a gauge for evaluating accounts of theoretical and practical reasoning, I show how views that characterize reasons simultaneously as mental particulars, merely descriptively true of agents, and as general standards authorizing conduct, fail. I argue that both instrumentalist and Kantian accounts of practical reason invoke mongrel concepts.

To focus on the deeper question of how reasons are binding, I redraw the distinction between practical and theoretical reason, showing that puzzles purportedly specific to moral reasoning are instead, puzzles for rationality itself. I argue that the normative authority of reasons is explained by their public character. The practice of giving and asking for reasons, and the inferences which reasons authorize within that practice, must be understood in intersubjective terms on pain of unintelligibility. According to intersubjectivism, what we have objective reason to do is a function of what good practical discourse would vindicate.
Acknowledgements

My time with the Department of Philosophy at the University of Toronto has been philosophically rewarding. Because the Department is active and diverse, it offered me numerous opportunities to learn, and enlarge the range of my competencies.

I feel very fortunate to have worked with such an excellent thesis committee. I would like to thank Cheryl Misak for the speed and thoroughness with which she responded to drafts of my work, and for the demandingness of her comments which forced me to more carefully develop the epistemological strands in my account of practical reason. I would like to thank Joe Heath for his enthusiasm about my project, and his readiness to engage in discussion about the topics of my thesis as well as offering me helpful advice about writing and presenting my work. I am particularly grateful to my supervisor, Arthur Ripstein, who not only guided my thesis work, but helped me to see that doing good work in philosophy is more than taking up a stand within the existing literature. By encouraging me to see that I had a contribution of my own to make, he helped me to develop my own voice.

My most sincere thanks go to my parents, Mae and Dan O'Hagan, who came from circumstances which did not permit the pursuit of an advanced formal education, but who
nonetheless value intellectual clarity and knowledge for its own sake. Of course, real respect for the value of learning meant that their education simply came from elsewhere, and that their children were given the gift of respect for learning. I want especially to thank my Mom for her faith in my abilities as a philosopher, and for her regular supply of delicious cookies.

My deepest thanks go to my unfailingly patient partner, Eric, who faithfully supports me in all of my endeavours, big and small. Without his kind encouragement it would have been much less fun, and certainly more difficult. Without his willingness to listen to, and engage with, my ideas as they developed, the thesis would have been much less coherent.
# Table of Contents

Chapter One: Introduction ................................................................. 1

Chapter Two: The Myth of the Given ................................................. 13
  The Given .................................................................................. 17
  Sense Data - A Mongrel Concept .................................................. 20
  Looks/Is ..................................................................................... 25
  An Objection ............................................................................... 38
  A Sellarsian Response ................................................................. 41
  Three Lessons for Practical Reason ............................................... 45

Chapter Three: Instrumental Reason and the Myth of the Given ............. 48
  The Promise of Instrumentalism .................................................... 54
  Crude Instrumentalism ................................................................ 57
  Sophisticated Instrumentalism ...................................................... 62
  Smith's Fully Rational Advisor ..................................................... 63
  Bratman's No-Regret Condition on Rational Agency ....................... 79

Chapter Four: Reflective Endorsement and the Myth of the Given .......... 93
  The Normative Question ................................................................ 101
  Korsgaard's Answer to the Normative Question ............................. 104
  The Persuasion Model .................................................................. 116
  The Demonstration Model ............................................................ 123
  Reflective Endorsement and the Myth of the Given ....................... 129
<table>
<thead>
<tr>
<th>Chapter Five: The Constitution of Agency and the Public Authority of Reason</th>
<th>137</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posing the Normative Question of Theoretical Reason</td>
<td>139</td>
</tr>
<tr>
<td>Constitutive Arguments and Agency</td>
<td>149</td>
</tr>
<tr>
<td>Constitutive Arguments and the Myth of the Given</td>
<td>163</td>
</tr>
<tr>
<td>The Public Authority of Reason</td>
<td>173</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter Six: Intersubjectivism</th>
<th>179</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary Of The Argument Thus Far</td>
<td>179</td>
</tr>
<tr>
<td>Making Room For Objectivity</td>
<td>183</td>
</tr>
<tr>
<td>Intersubjectivism</td>
<td>188</td>
</tr>
<tr>
<td>Agent-Relative and Agent-Neutral Reasons</td>
<td>203</td>
</tr>
</tbody>
</table>

| Chapter Seven: Conclusion                                    | 211 |

| Bibliography                                                  | 214 |
Chapter One:
Introduction

Philosophical inquiry into the nature of reasons for action usually proceeds from a distinction between theoretical and practical reasoning that is taken to be uncontroversial. Because it serves as a springboard for analysis, the way in which this distinction is drawn is critical. Typically, reasoning about action is taken to bear explanatory burdens which reasoning about belief is not. While it is commonplace to discriminate between theoretical reasoning and practical reasoning by describing the former as reasoning about belief, and the latter as reasoning about action, it is largely uninformative to draw the distinction this broadly. On the face of it, the thought 'it is already noon' might lead me to believe that I am late, or lead me to rush out the door. Philosophical orthodoxy has it that the normative authority of theoretical reasoning (its power to legitimately command) is straightforward; what we ought to believe will be constrained by theoretical reasoning. At the same time, the normative authority of practical reason over action is held to be puzzling. How could deliberation itself constrain action? It seems to make sense to say that agents are rationally constrained to believe one thing rather than another. But the claim that agents are rationally constrained to do one thing rather than another seems to many to be at odds with the role of the will in practical deliberation. Our desires and ambitions are a
source of motives for action. It is easy to conclude that rational constraints are categorically binding in theoretical reasoning, but only hypothetically or indirectly binding in practical reasoning. Kant's characterization of this distinction is useful here:

Now, all imperatives command either hypothetically or categorically. The former represent the practical necessity of a possible action as a means to achieving something else that one wills (or that it is at least possible for one to will). The categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end...if the action would be good merely as a means to something else the imperative is hypothetical; if the action is represented as in itself good, hence as necessary in a will in itself conforming to reason, as its principle, then it is categorical.1

The question of whether there are categorical reasons for action, or whether all reasons for action are hypothetical is central to a discussion of practical reason. Instrumentalists, for example, hold the view that reason's role in action is limited to the generation and evaluation of means to ends, ends set by something other than reason itself. The influence of reason upon action is, at best, indirect, its imperatives only hypothetical. Instrumentalists are thus inclined to portray reasons for action as more or less sophisticated belief-desire complexes. Because he is committed to the idea that practical reason only prescribes action indirectly, there being no categorical imperatives, the instrumentalist cannot appeal directly to reason's authority to explain how having a reason to act rationally compels action. From this vantage point, the question most urgently needing an answer is 'how can practical reasons be judgements about the world and still motivate action?' But the emphasis on the question of what motivates action obscures a

more fundamental question about reason itself: How are judgements, practical and theoretical, normatively binding?

This more fundamental question is obscured by the dubious but common assumption that theoretical reasoning provides a non-normative starting point from which to undertake a study of practical reasoning. The supposition that theoretical reasoning provides a non-normative starting point misleadingly suggests that the rational commitments involved in belief can themselves be understood in completely non-normative terms. This supposition has the consequence that it makes the rational authority of theoretical reasons unproblematically given in experience while making the authority of practical reasons mysterious. A faulty distinction between theoretical and practical reasoning is at work here, colouring from the very start our thinking about what practical reasons must be like, and leading us into philosophical error. It obscures the need to consider how rational commitments are themselves normative. As a result, it prejudices an analysis of practical reason by supposing it to involve normative commitments which theoretical reasoning doesn’t. My thesis is, in part, an attempt to show how this questionable distinction between theoretical and practical reasoning informs the debate about the nature of practical reasoning which it generates. It is also an attempt to show that by conceiving of both theoretical and practical judgements as involving irreducibly normative commitments, we are better able to account for the normative force of practical reasons.

In both epistemology and ethics, philosophers are confronted with the question of how agents are bound by rational norms, and in both disciplines philosophers are tempted to seek an answer in the form of privileged mental states whose occurrences are self-justifying. But the normative question, as it applies to both theoretical and practical reasoning, can not be answered
merely by an appeal to such mental states or psychological facts about the agent. It is a mistake to suppose that reasons can be justified by appeal to non-normative facts. In order to make this mistake visible, I outline Wilfrid Sellars's analysis of the normative commitments implicit in theoretical reasoning and apply its lessons to an analysis of practical reasoning. In Sellars's terminology, the attempt to justify by appeal to a non-normative fact, constitutes the employment of a "mongrel concept," one which illegitimately conflates causal and normative orders of explanation. The role of mongrel concepts in such explanations provides me with a gauge for evaluating the success of accounts of theoretical and practical reasoning.

With this framework in place, I am able to show how views of reason that characterize reasons simultaneously as mental particulars or facts which are merely descriptively true of the agent, and general standards or norms authorizing or legitimating conduct, employ mongrel concepts. For example, we see that instrumentalist appeals to the hypothetical desires of fully rational agents, or intention formation under idealized conditions, fail to provide a foundation for practical reason. For, if employed as descriptions of reason-making conditions, they cannot serve to explain reason's normative hold on agents. But, if employed as rational standards, they presuppose the normative authority which they must explain. In this way, the instrumentalist account of the normative authority of reasons conflates causal and normative orders in an attempt to ground reason's authority in a mental state. However, philosophers working within the Kantian tradition may also be guilty of employing mongrel concepts in their non-instrumentalist conception of practical reason, as I show in a discussion of Christine Korsgaard's account of reflective endorsement.

I contend that in order to focus on the more fundamental question of how reasons,
practical and theoretical, are binding, we need to draw the distinction between practical and theoretical reason in a different way. When we do this, we discover that puzzles purportedly specific to moral reasoning are instead, puzzles for rationality itself. The answer to the normative question, ‘why do I have reason to do?’ is fundamentally related to the answer to the question ‘why believe what I have reason to believe?’ The distinction between theoretical and practical reasoning should be marked according to their respective domains: theoretical reasoning is reasoning which is authoritative over belief and practical reasoning is reasoning which is authoritative over action.

I argue that a public conception of reasons is needed to explain the normative authority of reasons. Just as there can be no private language, there can be no private reasons, since private reasons would invoke thoughts articulated using private concepts. The practice of giving and asking for reasons, and the inferences which reasons authorize within that practice, must be understood in intersubjective terms on pain of unintelligibility; all agents engaged in reasoning are authorized in making the same moves in virtue of the same considerations. I argue that what is at issue is the rational criticism of judgements, and that theoretical and practical reason do not differ in this respect. It is in the character of their respective domains that practical and theoretical reason differ, not the nature of their normative authority. The intersubjectivist view of reasons which I develop in the final chapters of the thesis has it that what we have objective reason to do is a function of what good practical discourse would vindicate. This view avoids the myth of the given and makes ‘what we have reason to do’ a function of the social practice of reasoning. The structure of the argument is as follows.

In Chapter Two, I outline Wilfrid Sellars’s critique of the philosophical idea of givenness
in order to illuminate the normative structure of epistemic commitments. The framework of
givenness, which he takes to be common to both rationalism and empiricism, applies to any view
that conflates the justificatory and descriptive orders in an attempt to explain how cognitive states
of agents are justified. I begin with a discussion of Sellars’s account of the given in its role as the
mongrel concept of sense data. Sense data are those bits of experience that purport to be both
mental states caused in us by our interactions with the world, and also a basic form of knowing
which don’t presuppose the acquisition of any concepts. The attraction of sense data as
theoretical postulates lies in their character as a form of knowing that is justified merely in virtue
of being had, thereby constituting a foundation for, and explaining the origin of, knowledge. For
Sellars, sense datum theory was just the most visible form of a very general problem resulting
from the attempt to analyse normative concepts into descriptive concepts without remainder. Any
such attempt will ultimately commit itself to a class of concepts which purport to offer a
justification of judgements while claiming to involve nothing more than descriptions of states of
agents. These mongrel concepts obscure the fact that epistemic responses are irreducibly
normative. The commitments involved in these responses require the mastery of concepts
acquired within a public practice of giving and asking for reasons. What Sellars’s account of
epistemic commitments makes clear is that judgement itself is conceptually and irreducibly public,
constitutive of the very concept of agency. The constitutive character of Sellars’s account applies
no less to practical normative concepts than it does to theoretical ones. Applying the lessons of
Sellars’s account, I exploit his notion of the mongrel concept in an analysis of the normative
authority of practical reasons. I argue that the rational commitments of practical reason are
constitutively public and irreducibly practical and not, in their nature, essentially different than the
In Chapter Three, I discuss instrumental theories of practical reason. For my purposes, instrumental theories of practical reason can be defined as those which hold that the role of reason in action is indirect. It is limited to generating means to ends, ends which are set by something other than reason itself. According to instrumentalism, reason is neither normatively nor motivationally practical. Instrumentalists see it as a great advantage of their position over non-instrumentalist positions that it appeals to only minimal rational constraints and facts about the psychology of agents in the explanation of how reasons are justified. Mysterious acts of willing and questionably normative substantive moral principles are ruled out by the instrumentalist's conception of reasoning. Instrumentalists assume that theoretical reasoning is authoritative and that the bindingness of the hypothetical imperative is obvious and uncontroversial. However, as was shown in Chapter Two, theoretical judgements themselves depend upon irreducibly normative commitments. The instrumentalist supposes the authority of theoretical reasoning to be a brute fact of experience and consequently that non-normative descriptions of reasoning are capable of explaining how reasons rationally bind agents. This is the instrumentalist version of the mongrel concept.

I first discuss a crude form of instrumentalism which conflates merely having a disposition to act with having a reason to act. In this conflation, crude instrumentalism mistakes for an identity relation what is merely a necessary condition. In order to explain the normative authority of reasons for action any theory of practical reason will need to offer an account of the commitment, or normative attitude, involved in having a reason for action. In the remainder of Chapter Three I discuss two attempts to offer such accounts. Both Michael Smith and Michael
Bratman attempt to explain the normative force of reasons while remaining true to instrumentalist precepts. Smith attempts to work normative commitments into theoretical reasoning by way of what we desire. Bratman tries to work them into an account of rational intention formation. I argue that both attempts fail because they are unable to explain how reasons bind reasoners without presupposing the authority of the reasons which they set out to explain.

Smith identifies normative reasons with judgements about what one would want if one were fully rational. On his view, the authority of reasons for action is just the authority of well-formed beliefs. It is irrational for me not to desire to act in the ways that I would want to act were I fully rational. By making this identification, Smith supposes that he has shown that the failure to be moved by what one takes oneself to have reason to do is irrational. But Smith owes an account of the rational authority which the desires of the fully rational self have for the actual agent. My arguments push him into a choice between two options. If he claims that the authority of reasons is merely a tendency to act in a certain way he is left with a description, but not a justification, of our rational commitments. On the other hand, if he claims that the authority of reasons is a norm or rational imperative, he cannot ground it in our actual tendencies which are, after all, less than fully rational. I argue that in order to explain the authority of reasons he must grant that at least some reasons directly constrain action and thus must abandon instrumentalism. Our dispositions to desire what we would want if we were fully rational serve in Smith’s account as mongrel concepts, doing double-duty as brute facts and rational norms.

Bratman’s account of planning agency aims to uncover what is rational in the way of intention. I focus on his no-regret condition for rational agency, a condition intended to inform an analysis of how reasons oblige over time. The no-regret condition is motivated by the
thought that agents are concerned with their reasons for action not merely in the present but also after they have acted or failed to act on them. By focussing on regret and anticipated regret Bratman attempts to rationally constrain intention formation. The quality of an agent’s reasoning depends upon her dispositions to care about future regret. The problem with Bratman’s account is that our dispositions to care about future regret, while described as simple tendencies which we have, are called upon to serve as rational imperatives. This is a clear appeal to a Sellarsian mongrel concept. Dispositions to regret can function well or badly, they may serve or fail to serve as *indications* of the presence of reasons for action. The mere existence of such dispositions, however, does not guarantee their rationality or their normative significance. I argue that while both Smith and Bratman seem to recognize the need to turn to the practice of reasoning in order to explain the authority of reasons, their instrumentalist commitments prevent them from adequately developing this insight. Their accounts either slide back into crude instrumentalism or covertly presuppose a conception of rationality too rich to be derived from instrumentalist premises.

In Chapter Four, I turn to a discussion of Christine Korsgaard’s non-instrumentalist, Kantian account of practical reasons. Korsgaard develops the idea that the normative authority of reasons is essentially tied to the practice of reasoning and so does not fall prey to instrumentalist versions of the myth of the given. Korsgaard sees that the dictates of practical reason must be made practical in order to explain their bindingness. She provides an answer to the question of how reasons authorize action by way of her conception of reflective endorsement. Reasons for action, according to Korsgaard, come into being when an agent reflectively endorses (commits to) a disposition to act. The normative force of reasons is thus constitutive of the
practice of reasoning. In reflectively endorsing a disposition to act, the agent gives herself a law which binds her because she has given it to herself. Korsgaard's account is attractive in large measure because it carefully attends to the phenomenology of practical reasoning in the first person. However, this very attention to reasoning in the first person obscures the lessons of Sellars's critique of the myth of given. Reasons must be understood in terms of irreducibly public commitments from the start.

I argue that there is an ambiguity in Korsgaard's account of the normative authority of reasons. I distinguish two ways of interpreting her position and show that neither adequately explains how rational commitments are public. What I call the persuasion model makes rational commitments merely psychological facts about us. It invokes the myth of the given by making such psychological facts play a justificatory role, while failing to acknowledge the inferential connections which give them normative significance. What I call the demonstration model does not conflate the descriptive and justificatory in this way (in the form of a psychological event) but instead fails to explain the authority of reasons by making them essentially private. Korsgaard arrives at an account of the public character of reasons to which she aspires by way of an appeal to the transcendental value of humanity. Her transcendental argument amounts to an attempt to give the skeptic about practical reasons a reason to have reasons. But this presupposes what it attempts to establish. Reasons aren't intersubjective because we are morally bound to recognize the humanity of other agents. Our reasons are more fundamentally intersubjective than Korsgaard grants.

In Chapter Five, I offer a constitutive argument for the authority of reasons according to which reasons are irreducibly public. I argue that reasons' authority is bound up in the practice
of reasoning because reasons function within a public network of inferentially articulated commitments and entitlements. My account avoids the myth of the given by conceiving of the structure of rational norms in non-foundationalist, intersubjectivist terms. I first argue that the authority of normative commitments is as much a problem for theoretical reason as it is for practical reason, through a consideration of Lewis Carroll’s “What the Tortoise Said to Achilles.” This dialogue shows that the explanation of the normative authority of theoretical reasoning faces challenges like those raised in the explanation of the normative authority of practical reasoning. These challenges can only be met without invoking the myth of the given by making reason’s authority constitutive of agency.

After offering a constitutive account of reasoning, I consider Peter Railton’s view that constitutive arguments cannot explain the authority of reasons. Railton argues that constitutive arguments are limited because they fail to provide a non-hypothetical reason for the authority of belief. I argue that what Railton suggests is a limitation of constitutive arguments is rather a limitation on foundationalist conceptions of justification. Railton falls prey to the myth of the given because he links the explanation of the authority of reasons to some feature or state of an agent which is rationally undeniable in the first person. Constitutive arguments will invoke the myth of the given if they presuppose foundationalism about justification in their attempts to find a feature of reason or agency which itself justifies reason. The foundationalist mistake is to suppose that some feature of belief or agency, once identified, will serve as an explanation of the bindingness of rational norms. But this is to take reason’s authority out of the practice of reasoning and lodge it in the agent in the form of a mongrel state. To understand theoretical or practical judgements as propositionally contentful we must take them to have inferential
consequences and antecedents which are conceptually public. The normative authority of reasons cannot be lodged in a private mental state.

In Chapter Six I argue that a public conception of the authority of reasons provides the means to explain the objectivity of reasons without appeal to a foundation outside of the practice of reasoning. Intersubjectivism about reasons is the view that what we have reason to do is a function of what good practical discourse would vindicate. I contrast intersubjectivism with both conventionalism and realism about reasons. I show how conventionalism fails to distinguish between agreements which legitimate social practices and ones which fail because they are biased. Conventionalism cannot make sense of error, nor can it explain how what we have reason to do outruns what we take ourselves to have reason to do. Realism, because it makes reasons independent of the practice of reasoning, cannot explain how they are binding. I conclude the chapter with a brief look at Korsgaard’s discussion of the distinction between agent-neutral and agent-relative reasons. I argue that her treatment of this distinction makes the intersubjectivity of reasons more like a moral principle which we have reason to adopt than an account of the ways in which reasons are in principle shareable. I conclude that the intersubjectivity of reasons is a conceptual, not a moral feature of reasons.
Chapter Two:  
The Myth of the Given

In his justly famous "Empiricism and the Philosophy of Mind," Wilfrid Sellars engages in a sustained critique of the philosophical idea of givenness, under the slogan 'the myth of the given.' When he first gave the lectures later published as "Empiricism and the Philosophy of Mind," sense datum theories and phenomenalism were going concerns and so it is natural that he directed his attention primarily toward them. But the framework of givenness, which is his genuine target, is common to any view that confounds justificatory and descriptive orders in the explanation of normative concepts. The myth of the given is the idea that there can be a state of the agent, or form of awareness, which simultaneously has two properties. First, it is a state which is itself knowledge, or in virtue of which one has knowledge. Second, it is a state which doesn't presuppose the acquisition of any concepts, so that it constitutes or entails knowledge independently of the mastery of any concepts.¹

The myth of the given has many forms.² While Sellars's most sustained critique attacks

¹ Robert Brandom explains the myth of the given this way in his "Study Guide" for Empiricism and the Philosophy of Mind (Cambridge: Harvard University Press, 1997) 122.

² Sellars is clear that the myth of the given takes a great number of forms, of which sense datum theory is only one kind. See his Empiricism and the Philosophy of Mind (Cambridge: Harvard University Press, 1997) 13-118. 32 ff., 68ff.
the idea that some instances of sense experience themselves provide a peculiar sort of self-justified, immediate knowledge which furnish a foundation for empirical knowledge, its lessons apply directly to practical reason. In the next chapter I will argue that many of the problems for the myth of the given which Sellars identifies have parallels in instrumental reason. In this chapter I will limit myself largely to an exposition of Sellars’s argument against the myth of the given as it appears in *Empiricism and the Philosophy of Mind*. I will undertake this task with a view to emphasizing those features of Sellars’s critique that bear lessons for an analysis of practical reason.

The term ‘sense data’ is just the term for those privileged bits of awareness purported to grant an immediate and authoritative relation between subject and world. As such they lodged the basic explanation of knowledge in private, incorrigible mental episodes. While practically no one is a sense datum theorist these days, in contemporary philosophy of mind proponents of *qualia* might be thought of as clinging to something like sense data. Still, forms of naturalized epistemology such as reliabilism are very common. Attempts to naturalize the normative by analysing it in causal and descriptive terms may invoke the myth of the given in its more general form, not its foundational form. For my purposes, the relevance of Sellars’s analysis of the myth of the given is, more broadly, its critique of uncritical forms of naturalism as conflating the descriptive and the normative orders of explanation. It isn’t clear that there is a counterpart in theories of practical reason to the sense datum form of the myth of the given. Ethical intuitionism, now much maligned, probably comes closest. Nonetheless, conceptions of practical reason which conflate the descriptive and normative are not, as we shall see, uncommon.

Sellars’s critique of the myth of the given is anti-foundationalist, but also more broadly,
an argument against attempts to explain normative concepts in completely non-normative terms. It is often supposed that the real problem with the myth of the given is its foundational role in knowledge. But Sellars is clear that foundationalism is only one form of the more general problem. He acknowledges that even philosophers who do not reject the idea of inner episodes "find the Myth of the Given to consist in the idea that knowledge of these episodes furnishes premises on which empirical knowledge rests as on a foundation. But while this idea has, indeed, been the most widespread for of the Myth, it is far from constituting its essence." The more general form of the problem resides in the form of the 'mongrel concept.' Mongrel concepts are thus dubbed because they run together two modes of explanation: the causal, or descriptive and the normative or justificatory. Mongrel concepts lodge justificatory powers in states of the agent or world. They are philosophically pernicious when they are supposed to provide a judgement's justification without needing to make reference to other judgements. This offers them fictitious foundational powers. However, when we characterize a state of a person normatively (for example, as an instance of knowing) we are not merely giving an empirical description of it: "in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being

---

3 Sellars, *Empiricism and the Philosophy of Mind*, 33. Sellars goes on to clarify that one can give up on the myth of the given in the form of inner episodes, but still be subject givenness in a variety of forms. His discussion of the "thermometer view," which can be thought of as a precursor to reliabilism, shows that the more general problem lies in the supposition that an explanation of authority of judgements (perceptual and otherwise) can be had without making reference to other propositions or concepts. Reliabilists are not foundationalists in any straightforward sense, but as we shall see, are guilty of employing this more general form of givenness.
able to justify what one says.”

Three broadly Sellarsian themes influence my project. First, Sellars rejects foundationalism of the sort that assumes a presuppositionless mode of discourse. The idea that some special acts of awareness, or knowings, are immediate and justified without reference to other propositions or concepts is untenable. Rather, an explanation of knowledge claims requires reference to a normative relation between the agent and the world in the form of a claim or a commitment about the way the world is. This commitment includes not just the thought that the judgement is true, but also the thought that the grounds for the judgement are normatively adequate and so implicitly commit the agent to many judgements. Sellars’s treatment of justification as an essentially practical attitude of commitment has clear application to my discussion of the inadequacy of purely instrumental conceptions of practical reason.

Second, Sellars recognizes that judgements have an irreducibly public character. Their intersubjective character is fundamental to an account of their justification. In the epistemic case, Sellars makes this point by showing that talk about the way things look presupposes talk about the way things are. Justified uses of concepts require the conceptual mastery of those concepts as they are used by agents in a linguistic community. Judgements, practical and theoretical, cannot be grounded in private mental states. Rather, they occur within the space of reasons which is where the inferential moves constitutive of knowledge claims are made. While the appeal to ‘private appearings’ is no longer common in epistemology, their counterparts in action theory and ethics are common indeed. Given Sellars’s insight we should balk at the idea that practical reasons could have a purely private justification.

---

4 Sellars, *Empiricism and the Philosophy of Mind*, 76.
And finally, Sellars holds that justification is in an important sense *sui generis* - a normative rather than a causal relation, based upon inferential relations. It is the public practice of exchanging reasons which furnishes the possibility of their justification. Justification is essentially and conceptually practical, not reducible to natural facts. In making assertions, agents take responsibility for the inferential role of their assertions. If justification were just a matter of fact, then normative concepts would be explicable in entirely non-normative terms. As the treatment of reliabilism later in this chapter will show, this isn’t so. Instead justification is a matter of taking responsibility and making commitments, in short, it is essentially practical.

The Given

The classical empiricist and classical rationalist both attempt to explain knowledge by grounding it in inferences from other things known and privileged states of awareness which provide a foundation for knowledge. Whether we think of the empiricist’s sense impressions, or the rationalist’s clear and distinct ideas (or *a priori* truths) these states are non-inferentially justified (they are self-justifying) and count as knowledge independently of any other fact about the person. Hence they are purported foundations for knowledge. The given provides an attractive first step in answering the question ‘what can we know?’

Let’s begin by identifying the sort of thought which motivated Sellars’s critique of the myth of the given:
Now the idea that epistemic facts can be analyzed without remainder - even "in principle" - into non-epistemic facts, whether phenomenal or behavioral, public or private, with no matter how lavish a sprinkling of subjunctives and hypotheticals is, I believe, a radical mistake - a mistake of a piece with the so-called "naturalistic fallacy" in ethics.⁵

The above expressed conviction, that epistemic discourse is irreducibly normative, features importantly in Sellars's authoritative and thoroughgoing critique. Where criticism of the naturalistic fallacy takes aim at the identification of an ethical concept with a natural concept, Sellars's criticism of the myth of the given takes aim at the identification of an epistemic (inferential) fact with a natural or descriptive fact. While Sellars uses sense datum theories as his primary target in *Empiricism and the Philosophy of Mind*, his argument is aimed more broadly at positions which attempt to reduce normative concepts to descriptive ones. An appeal to what is given in experience is one loaded with commitments, and Sellars's aim is to demonstrate how, once these commitments are made explicit, the given crumbles under the weight of its own authority. The foundation for empirical knowledge which the given is purported to provide is illusory, a confusion arising out of the conflation of causal and justificatory aspects of belief formation.

A large number of philosophical views seek to ground problematic philosophical concepts which involve norms or rules (such as epistemic and moral terms) in a descriptive account of the nature of things. The given is part of an initially attractive philosophical account of knowledge, referring to the purported incorrigible and immediate apprehension of the contents of sensory experience expressed in present tense, first person reports. It seems appealing because it seems

---

⁵ Sellars, *Empiricism and the Philosophy of Mind*, 19.
to offer a means of escaping an infinite regress of justificatory claims by postulating a special class of basic statements of a conceptually distinct sort which form a foundation for knowledge. While the justification of most (or many) statements is had by reference to other statements, the justification of some class of statements is, so the story goes, given in experience itself. Indeed, it might be argued that nothing but a privileged class of beliefs could do the work of justifying knowledge because mere coherence, consistency among beliefs, won't ensure the truth of the claims involved as the propositions may be consistent but false. This view was common to classical rationalists and empiricists alike. A strategy of this sort is well exemplified by Descartes's search for the indubitable in the *Meditations*, and his subsequent determination that that which can be known clearly and distinctly is true: "I am sure that I am a thinking being; but do I not then know what is required to make me sure of something? Certainly, in this first conclusion, there is nothing else which assures me of its truth but the clear and distinct perception of what I affirm." The common philosophical perception of Locke, Berkeley and Hume as philosophers who ground ideas and concepts in sensory impressions provide us with an empiricist version of the approach.

Proponents of the given claim that it provides a basic and authoritative relation between subject and world (such as particular sorts of perceptual experiences) thereby explaining how empirical knowledge can be founded on a non-inferential matter of fact. Awareness of what is given in experience is immediate in both a causal sense, because it is directly present to us (it is causally immediate) and an epistemic sense, because reports of what is given are justified

---

independently of all other beliefs and evidence. Experience provides the agent with (or forces upon the agent) a self-justified awareness of some feature of her world, occasioned without the exercise of a conceptual apparatus. *Qualia*, and ‘clear and distinct ideas’ are examples of such phenomena. Part of the attraction of the given is that it promises to explain normative, justificatory concepts in purely descriptive terms (such as being in a particular state). I will argue that instrumental conceptions of practical reason fall prey to the myth of the given when they characterize dispositions to act as reasons for action. When the agent’s dispositions are understood both as merely a set of facts about her and as reason-providing, they constitute a classic instance of the conflation of causal and justificatory orders. Retracing Sellars’s argument will help to disentangle some of crucial features of the myth of the given to which I shall attend in my discussion of instrumental reason.

**Sense Data - A Mongrel Concept**

The concept of the sense datum itself results from a cross-breeding of two ideas: the thought that there are particular sorts of inner episodes available to humans without the mediation of learned concepts is conflated with the thought that these inner episodes are instances of non-inferential knowings, providing the foundation on which empirical knowledge is based. Causal and justificatory orders get confused through a failure to appreciate that it is particulars (not assertions) which are picked out by their causal roles, while assertions or claimings (not

---

7 Alvin Goldman characterizes the given as incorporating these two forms of immediacy in “the given,” in *A Companion To Epistemology*, eds. Jonathan Dancy and Ernest Sosa (Oxford: Basil Blackwell Publishers, 1993) 159.
particulars) are picked out by their inferential or justificatory role. There is no reason to suppose that the particulars sensed in acts of perceptual awareness constitute epistemic facts. Epistemic facts concern epistemic, not causal, relations. We can distinguish between causal and epistemic relations roughly by recognizing that causes determine or influence causal relations while inferences or reasons determine epistemic relations. This is not to say, of course, that an epistemic relation exists outside of the causal order, but that epistemic relations are justificatory, or normative, in a manner in which causal relations are not.

While Sellars does challenge the concept of givenness, he does not deny the existence of inner states, but only denies that they possess the epistemic powers that the foundationalist supposes them to have. An appeal to an unmistakable and foundational piece of one’s experience as a privileged class of self-justifying facts requires an explanation of how it is that a sensation can constitute a non-inferential matter of fact. It is one thing to say of a perceiver that she entertains visually a round red patch and another to say that she sees an apple. Propositional contents, not particular sense data are known, and so the sense datum theorist owes an explanation of the (non-inferential) move between sensing a sense datum and the occurrence of knowledge.

By focussing on the move from the sensing of sense data to non-inferential knowings, Sellars points to the source of the difficulty involved in appeals to the given: an awareness of sense data cannot provide the privileged foundation for empirical knowledge which appeals to the given are intended to provide. The defender of the given must explain how it is that awareness of what is given in experience constitutes a form of knowledge. As Sellars sees it, the sense datum theorist has two options: either sensing (or awareness) is unanalyzable or it is analyzable. To say that sensing is unanalyzable is to hold that particular sense data are sensed,
and because sensing is not knowing, the existence of a sense datum doesn’t itself imply the existence of knowledge. In other words, if sensing is unanalyzable, or basic and immediate, the relevant fact involved in an act of sensing is the non-epistemic fact about the sense content (‘that a red, round patch has been sensed’). To say that it is analyzable is to hold that sensing is a form of knowing. In order for it to be a form of knowing it must be the case that facts rather than particulars are sensed. In other words, if sensing is analyzable it implies knowledge of a fact because it is of itself knowledge of a fact (‘there is an apple’). The sensing of sense contents cannot imply the having of non-inferential knowledge unless sensing itself is an immediate form of knowing. If sensing is non-inferential knowledge then, as awareness of what is given in experience, it must be awareness of facts and not particulars.

The problem to which Sellars here points is not exclusive to sense datum theory. His point is that a simple act of awareness cannot both be an awareness of a particular tokening and a fact about that tokening. The task, for the defender of this sort of view, will be to demonstrate how knowledge or belief can be the result of a simple act of awareness without appealing to the features of knowing or believing in order to explain them. Another way to think of the problem is to see it as the problem of explaining the move from non-epistemic, natural facts, to epistemic or inferential facts. First, we may grant that it is because there is a red apple in front of me that I sense a round, red sense content. But second, the sense datum theorist claims that I have the non-inferential belief that ‘there is a red apple’ because of my sensing of a round, red sense content, and third, that because I have this non-inferential belief, along with the belief that apples
are four for a dollar, that I have the inferential belief that this apple costs twenty five cents. 

Sellars challenges the view that the sense datum theorist is entitled to the use of because in the second and third cases. Although the first because does explain a causal relation, the third one surely does not. It is true as a description of me that I have a belief of the apple-expense variety, but not true that the content of that belief can be understood or justified as the causal product of a set of beliefs. My 'apples cost twenty five cents' claims will need to be evaluated through an investigation of the reasons supporting this conclusion, rather than the causes which may have led to it. But what of the second because? Is my 'there is a red apple' belief a strictly causal consequence of the sensing of sense contents, or is it the product of an epistemic relation? This is where the real difficulty for any epistemological position lies and where one is most likely to be tempted by the myth of the given.

Sellars argues that proponents of the given are committed to an inconsistent triad of propositions. The proponent of the given is faced with the following problem. The claim that we just see red, round patches as apples and thus that we immediately know of the existence of apples in our visual field, raises the question of how concepts are formed. The classical empiricist, denying that our concepts themselves are furnished independently of experience, is committed to the following inconsistent triad:

A. X senses red sense content s entails x non-inferentially knows that s is red.
B. The ability to sense sense contents is unacquired.

---

8 This way of laying out the problem (in terms of "because" claims) is suggested by Robert Brandom in his "Study Guide," Empiricism and the Philosophy of Mind, 127.
C. The ability to know facts of the form $x \phi$ is acquired.\footnote{Sellars, \textit{Empiricism and The Philosophy of Mind}, 21.}

If $A$ and $B$ are true, then $C$ must be false. If $B$ and $C$ are true, then $A$ must be false. And, if $A$ and $C$ are true then $B$ must be false. The difficulty facing the proponent of givenness concerns the plausibility of the explanation of a perceptual event as itself constituting a self-justified epistemic belief. What he needs to reconcile is the idea that an act of sensing produces an act of knowing with the idea that sensing sense contents is not an acquired ability, given that the ability to know facts is an acquired ability. While Sellars's discussion of the inconsistent triad is directed at the sense datum theorist, it is easy to see that classical rationalists (those who are proponents of the \textit{a priori}) will be guilty of the inconsistent triad. \textit{A priori} concepts would have to be non-inferentially known and in the appropriate way, unacquired.

The proponent of givenness might opt to give up $C$, the claim that the ability to know facts of the form $x \phi$ is acquired. To do this would be to hold that acts of consciousness with conceptual commitments (thoughts to the effect that something is such and so) were possible without the agent having acquired those concepts, for example, by having learned a language. But we understand the conceptual commitments involved in knowledge claims by understanding their consequences for other beliefs. So the proponent of givenness could not give up $C$ without absurdity. For example, he would be in a position which allowed that one could know something to be red without knowing it to be coloured, or for that matter, anything else. Judgements, even of the simplest perceptual sort, are conceptual.

It might be objected at this point that the argument has become too narrow, sense datum
theorists aren’t plentiful these days, so the argument has really lost its audience because it has lost its target. But the concern expressed by the inconsistent triad would be too quickly dismissed by such an objection. The inconsistent triad reminds us that any explanation of non-inferential knowledge must be squared with a causal explanation of belief formation which nonetheless doesn’t preclude an epistemic, or inferential, explanation of belief, for without the latter justification and knowledge cannot be explained. For Sellars, sense datum theory is simply the most visible form of a very general problem – the attempt to analyse normative concepts into descriptive concepts without remainder.

**Looks/Ls**

Since the myth of the given arises out of the mongrel concept of which the sense datum is a classic form, the resolution or dissolution of the myth requires careful analysis of the inner episodes involved in non-inferential claims. While Sellars grants the existence of inner episodes, he denies that the awareness of inner episodes is a form of empirical knowledge which is immediate (non-inferential), presupposes no other knowledge of matters of fact, and can serve as the foundation for all other factual claims. His strategy is to accept that ‘x is red’ is necessarily equivalent to ‘x would look red to standard observers in standard conditions’ but deny that ‘x is red’ is defined by ‘x looks red under standard conditions.’ The sense datum theorist wants to define knowing that ‘x is red’ in terms of its looking red under standard conditions because this would offer a definition of knowledge in non-epistemic terms. However, as Sellars argues, claims about what looks to be the case presuppose and rely upon an already established conception of
what is the case. There are two parts to his argument: first, a demonstration that ‘looks’ doesn’t express a relation, and second, a demonstration that ‘looks’ talk is not conceptually prior to ‘is’ talk.

What does it mean to say that ‘looks’ doesn’t express a relation? Initially it seems quite plausible to believe that in ‘x looks y to z,’ looks expresses a relation between, say, an apple, its redness, and an observer, or even more simply between an apple and redness (‘x looks y’). If ‘looks’ claims do express relations between objects or properties and observers then it seems plausible that knowledge could be founded upon sense contents. However, Sellars shows us that looks claims don’t express relations because they cannot be taken to be non-inferentially elicited claims about objects. In making this point, Sellars develops the example of John who works in a necktie shop. The example is useful because it draws our attention to some relevant features of perceptual claims. Specifically, by drawing our attention to a form of systematic error resulting from a change in observation conditions, we come to see that claims about what ‘looks’ to be the case are conceptually dependent upon claims about what is the case. ‘Looks’ claims cannot, therefore, provide the relevant foundations for perceptual knowledge.

We are to imagine that John works in a necktie shop and ably identifies ties by their colour. John judges colours correctly under standard conditions. His skill at classifying colours meets with the affirmation of his friends and customers until electric light is installed in the store, thereby altering the conditions under which he makes colour judgements. John quickly learns that his post-electric lighting assertions about the colours of the ties are now often met with dissension. When he says that a tie is green he is contradicted by a customer who takes him out to the natural light to demonstrate that the tie is actually blue. John learns that his in-store colour
judgements are systematically in error. We can imagine John trying to adapt to the new situation by thinking twice about making any assertions about green ties: to his thought ‘this is green’ he might react ‘but wait, if it looks green in the shop it is probably actually blue’. If asked what colour the tie is, he might admit that while it looks green to him, it is actually blue. In order to see the importance of Sellars’s example we must first note the difference between two types of claims which John has made. Some of his claims are evoked inferentially, while others are non-inferentially elicited. When immediately preceded by the thought ‘this tie looks green, therefore it must be blue,’ the claim ‘this tie is blue’ is an inferentially elicited one. Once John becomes accustomed to the new lighting system it is likely that his claims about blue ties will no longer be inferentially evoked, but instead arise as non-inferentially elicited responses to his environment.

The example points to the need for a more careful analysis of the nature of inner episodes. Two distinct features of the non-inferentially elicited report need to be distinguished. The first is the ‘reliable differential responsive disposition’, the disposition possessed by the agent who has been trained to respond to particular features of her environment in particular ways. Pre-electricity John had a reliable differential responsive disposition to respond to green ties with the thought, or the assertion, ‘there is a green tie’. Post-electricity John had an in-store reliable differential responsive disposition to respond to blue ties with the thought or assertion ‘there is a green tie,’ at least during the immediate period after the switch. It is the possibility of systematic error that helps us to see that a reliable differential responsive disposition is a necessary

---

10 Here I follow Robert Brandom in avoiding Sellars’s misleading terminology. What Sellars calls reports, I have referred to as non-inferentially elicited claims. What Sellars calls fact-stating claims, I have referred to as inferentially elicited claims. In a footnote, Brandom notes that there is no reason to deny that non-inferentially elicited claims function in a fact-stating manner. “Study Guide,” Empiricism and the Philosophy of Mind, 138.
but not sufficient condition for non-inferential knowledge. A consistent response to one's environment doesn't demonstrate knowledge of some feature of that environment unless (or without) the agent's endorsement of that responsive disposition. This cannot be done without the application of concepts. In other words, non-inferential claims like 'this tie is green' cannot be understood as simple descriptions of experience; the claim that something (a tie) possesses a certain property (has greenness) is not something that a differential responsive disposition can account for. We may, or may not, endorse the product of our differential responsive dispositions, as we saw when John declined to endorse his 'this is green' response to (what he had collateral reasons to believe was) a blue tie. Endorsement is the second distinct feature of non-inferentially elicited reports which Sellars identifies.

A response is endorsed when it is taken up into the *space of reasons*, the game of giving and asking for reasons: "to say that a certain experience is a *seeing that* something is the case, is to do more than describe the experience. It is to characterize it as, so to speak, making an assertion or claim, and - which is the point I wish to stress - to *endorse* that claim." Endorsement, unlike the differential responsive disposition, requires the capacity to engage in inference. Thus we can explain the difference between a parrot which has been trained to respond to blue ties by uttering 'this tie is blue' and John's utterance 'this tie is blue.' The parrot can reliably respond to blue ties but hasn't mastered the inferential role played by 'this tie is blue' reports. According to Sellars, one has grasped a concept when one has mastered the use of a word. A word expresses a concept if it has an inferential role, that is, if it can be employed in the premises and conclusions of inferences which may be judged correct or incorrect.

---

This makes reasoning intersubjective. When we make assertions, we confer authority upon propositions and in doing so, make ourselves responsible for them. We leave ourselves open to challenge in this way. For example, John’s claim ‘this necktie is green’ authorizes others to believe that it is. During the period when John systematically misidentifies ties, others are likely to hold him responsible for his assertion by directly challenging it — ‘it’s blue, not green!’ But the inferences for which he makes himself responsible will be manifold. A person out to buy a green tie may take John’s assertion that it is green to be a reason to buy it. If he does buy it, he is likely to feel misled. The commitments undertaken in the exchange of reasons are themselves public. Endorsing a claim is taking up a position in the game of giving and asking for reasons. Reasons demand concept use because reason use is based upon inferential relations:

the inferential role is a matter of what conclusions one is entitled to draw from such a statement when it is overheard, what would count as a reason for it, what is incompatible with it and so a reason against it. This is a matter of the inferentially articulated content of the assertional commitment undertaken by the reporter in virtue of the performance that is the reporting: what the reporter is responsible for.\textsuperscript{12}

Thus, we see that our assertions (even non-inferentially elicited ones) contain an irreducibly normative component and cannot be explained in purely descriptive terms.

According to Sellars, saying that ‘x looks green to S at t’ amounts to saying that x has an experience that, if one were prepared to endorse the claim, would describe S’s seeing that x is

\textsuperscript{12} Brandom, "Study Guide," \textit{Empiricism and the Philosophy of Mind}, 140.
One of the advantages of Sellars's account of endorsement is that it is able to explain the varieties of assertions possible with 'looks' talk in a way that appeals to the given cannot. 'Looks' claims may be either qualitative, as in 'the apple looks green,' or existential, as in 'there looks to be a green apple.' In the first case, the agent commits herself to the existence of the apple but not its colour, whereas in the second case, she commits herself to neither the existence of the apple, nor its colour. Generic and determinable observations can also be explained by Sellars's view although they cannot be explained by a theory which posits a given. Looks are often qualitatively generic, such as 'that building looks tall,' as opposed to determinate, 'that building looks to be 100 stories high.' We may see that something looks red without noting that it looks to be a particular shade of red. If scarlet, pink, and so on, were given to us in experience then it would be difficult to understand how one could be appeared to redly. Sellars's position avoids this problem nicely by attending to the variety of ways in which claims are endorsed. It doesn't follow, of course, that redness, or tallness, pink, or scarlet, cannot be understood in part as inner episodes. The point is just that 'looks' talk simply signals limitations on the observer's commitment to endorse the claim. The building may look tall but because I am standing in a deep hole in front of it and so I do not endorse the proposition that it is tall.

On Sellars's view, 'looks' talk fails to make claims about the world and thus cannot express the speaker's commitment to the world being a particular way. Instead, we should think of looks claims as an expression of the agent's differential responsive disposition to say that '{x looks red,' while nonetheless withholding his endorsement of that claim (perhaps, like John, he has become aware that he is systematically in error). Hence, the incorrigibility of looks talk

---

13 Sellars, Empiricism and the Philosophy of Mind, 41-42.
should not be explained by the infallible awareness of sense contents (as Descartes thought) but instead as arising from the failure to make a commitment. One cannot be wrong if one hasn’t committed oneself to anything. We should not, therefore, appeal to sense-contents to ground knowledge.

From this it follows that ‘looks’ talk does not express a relation. Sellars’s contention that ‘looks’ doesn’t express a relation is part of a broader argument intended to show that the awareness of inner episodes, understood as non-epistemic or natural facts, can neither explain non-inferential knowledge, nor provide a foundation for empirical knowledge. The endorsement thesis suggests that claims about how something looks are not reports about how something is but rather expressions of ‘is’ claims to which one is not willing to commit. ‘Looks’ talk thus presupposes ‘is’ talk. This leads us to the second part of the argument: ‘looks’ is not conceptually prior to ‘is’.

Instead of holding that what looks to be the case ultimately provides the grounds for what is the case, Sellars argues that an explanation of what is the case is conceptually prior to (or is presupposed by) what looks to be the case. He puts it like this:

The point I wish to stress at this time, however, is that the concept of looking green, the ability to recognize that something looks green, presupposes the concept of being green, and that the latter concept involves the ability to tell what colors objects have by looking at them - which, in turn, involves knowing in what circumstances to place an object if one wishes to ascertain its color by looking at it.\textsuperscript{14}

\textsuperscript{14} Sellars, \textit{Empiricism and the Philosophy of Mind}, 43.
The reference to circumstances here is important as it provides the means for a shift in our understanding of non-inferential knowings away from 'looks' to 'is'. Again, having the differential responsive disposition isn't enough, one must know under which conditions it has application. Mastery of a concept is not reducible to a purely descriptive order because conceptual knowledge requires some capacity to determine the scope of the differential responsive disposition. Hence, an epistemic attitude remains part of the story. 'Looks' is not conceptually prior to 'is' because the epistemic attitude involved in non-inferentially elicited assertions functions to adjudicate the propriety of endorsements of the differential responsive disposition in particular cases and thus expresses an 'is', not a 'looks', claim about a perceptual taking. So, the sentence:

\[ x \text{ is red } \equiv \quad x \text{ would look red to standard observers under standard conditions} \]

is, as Sellars claimed, true not because the right hand side constitutes a definition of what is red, but rather because it defines standards conditions. Standard conditions are those in which things look like what they are. Thus,

\[ x \text{ is red } \equiv \quad x \text{ would look red to standard observers under conditions in which things look like what they are} \]

Perhaps this claim is too strong. One need not know all of the conditions under which an agent's differential responsive disposition has application in order to have mastered a concept, for the set of these conditions may be indefinite. All that needs to be preserved is the thought that there is a normative dimension involved in the employment of the differential responsive disposition.
To put the point even more explicitly,

\[ x \text{ is red } = \text{ } x \text{ would look red to standard observers under conditions in which red things appear to be red.} \]

Standard conditions are those in which an agent's differential responsive disposition is dependable and ought to be endorsed.

Sellars has, in effect, reversed the direction of dependence traditionally taken to hold between appearance and reality in epistemology. The myth of the given arises from the thought that an act of awareness \textit{per se} could have epistemic significance independently of the acquisition or deployment of concepts. As his argument and re-interpretation of the definition of red have shown us, the idea that there could be beliefs which presuppose no other beliefs implicitly appeals to the given. Understanding of the world would be impossible without sensory input but sensory input alone is insufficient for understanding (a position reminiscent of Kant's motto of the \textit{Critique of Pure Reason}, "intuitions without concepts are blind"). Even non-inferentially elicited claims require or presuppose concept acquisition and thus the question of how concepts function within a language is raised.\textsuperscript{16}

According to Sellars, we don't learn the meanings of words by associating words with things of which we are already aware. For example, it is not the case that we become proficient redness detectors and then learn the meaning of red once able to associate the word 'red'

\textsuperscript{16} This amounts to a denial of the second proposition of the inconsistent triad which we earlier examined.
correctly with our detections of redness. We don’t have concepts because we have repeatedly noticed a certain kind of thing, rather, to have the ability to notice a particular kind of thing requires having the concept and cannot account for it. A perceptual taking of a red apple is a conceptual response to it. A non-inferentially elicited judgement about a red apple cannot be had without mastery of the concept red. This is not to say that infants and other non-conceptual animals cannot respond differentially to red things, but is just to say that they cannot be aware of red things as red. Red apples will typically causally elicit ‘there is a red apple’ reports and this does not imply that the concept must already be available to the agent in order for it to be used in non-inferentially elicited reports. The concept ‘red’ is applied to the thing which elicits the report (for Sellars, the apple itself) and so concept formation requires the existence of a differential responsive disposition but does not presuppose a pre-conceptual awareness of conceptual entities such as red.

This has become very complicated. Why not simply allow that an ostensive definition (pointing to a red apple, for example) defines red? Because, the temptation to appeal to ostensive definition is nothing other than the temptation of the myth of the given. Ostensive definition is not a sufficient condition for defining for the same reason that the act of following a rule cannot be understood as simple uniformity of practice (“in which case the lightning, thunder sequence would ‘follow a rule’”) without doing an injustice to the normative attitude the actor brings to

\[17\] Like Wittgenstein, Sellars hold that rule following is public practice which cannot be reduced to mere regularity. *Empiricism and the Philosophy of Mind*, 73. In remark #202 Wittgenstein writes: “...‘obeying a rule’ is a practice. And to think one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule ‘privately’: otherwise thinking one was obeying a rule would be the same thing as obeying it.” *Philosophical Investigations*, trans. G.E.M. Anscombe (Oxford: Basil Blackwell, 1989).
the situation. It is because she brings this epistemic attitude to the situation that the question of whether or not her differential responsive disposition is to be endorsed arises. Endorsement enters and further complicates the picture and not just any old schema of endorsement will suffice. Recall that John readily endorsed his 'this tie is green' response to blue ties because of the change in background conditions. What is needed is a means of conferring authority on those assertions (and those doing the asserting) which get things right (whose reports are reliable) while denying authoritative status to those assertions (and those doing the asserting) which fail to get things right.

As we have already seen, inferential relations play a fundamental role in concept acquisition and are thus relevant to the task of determining whose claims will be authoritative. The authority of beliefs is roughly their capacity to justify other claims. In order for the reported endorsement of a disposition to count as knowledge - in order for it to have authority - it must be taken to be reliable. This brings us back to the space of reasons and the social practice involved in knowledge. Propositional content depends upon inferential relations, which in turn require that a thought be placed in the space of reasons so that it may be assigned an inferential role in the game of giving and asking for reasons. The reporter's capacity to defend the commitment she has undertaken in her speech act is her capacity to justify her report. As we saw in the discussion of 'looks' a commitment is necessary for something to count as a report, this claim is now expanded so that in order for something to count as a justified report (as knowledge)  

---

18 Brandom notes that one need not accept Sellars's (strong) claim that knowledge requires not just that the agent take the report to be reliable, but that she knows it to be reliable. Taking the report to be reliable seems adequate to account for the necessary normative attitude and less contentious, so that is the interpretation I shall adopt. "Study Guide," Empiricism and the Philosophy of Mind, 157.
it must have potential as a defensible commitment - the reporter must have good reasons for holding it. This really amounts to the reporter having the capacity to place the report within a bit of inferential reasoning, for example, 'I was disposed to say red', and 'I am a reliable reporter of redness.' In other words, even though the report was non-inferentially elicited, the reporter must be able to employ the report within an inferential context as the conclusion of a piece of reasoning.

By conceiving of judgements in this way, Sellars is committed to the view that an agent's access to her mental states (even when they are spontaneously elicited) is conceptually public. The agent has no resources to individuate mental states except by the use of concepts and words which are entirely learned in intersubjective contexts. Sellars explicitly challenges the classical view of thoughts as inner episodes to which verbal behaviour owes its meaningfulness. Instead he offers his “myth of Jones,” wherein the genius Jones invents a theory of mental states to explain what causes and connects acts of overt speech. Thoughts are postulated by Jones as theoretical entities, introduced to explain both the inferential gaps between verbal episodes of behaviour and the real underlying cause of those episodes. In thinking of inner episodes in this way, Jones identified thoughts on analogy with pieces of spoken behaviour. As the myth of Jones develops, this theory becomes entrenched among Jones's fellows as a successful instrument which they use for predicting what is going on between episodes of speech and thus for interpreting each others’ behaviour. Finally, in the denouement of the myth of Jones, Sellars imagines that the theory begins to be used for self-description. Agents learn to explain their own behaviour by identifying their inner states using the theory, in the first person. What started as a theory has come to have a first person reporting use. Sellars's story helps us to see that “concepts pertaining
to such inner episodes as thoughts are primarily and essentially *intersubjective*...and that the reporting role of these concepts — the fact that each of us has a privileged access to his own thoughts — constitutes a dimension of the use of these concepts which is *built on* and *presupposes* this intersubjective status."¹⁹ This helps to explain the claim that *is talk* is conceptually prior to *looks talk*. When something looks red to a person the perceptual experience 'it looks red' must be understood by way of concepts which the agent has acquired through a public process of learning.

When we bring these consideration to bear on practical reason, we see the danger in thinking of reasons as private mental states such as desires. The view of practical reasons which I advance is able to make sense of the intersubjective character of reasons. We come to have reasons for action not by uncovering previously hidden mental states, but by being rationally committed to a course of action. One can learn to experience these forms of commitment as mental states in a way comparable to our acquisition of perceptual competencies. One can learn to experience one's hunger as a reason to eat, for example. But clearly, hunger is only defeasibly a reason to eat. If one is fasting, for example, hunger doesn't give one a reason to eat. Accordingly, it would be a mistake (indeed, an instance of the myth of the given) to identify the reason for eating which one may have with the experienced character of one's hunger.

Let me briefly summarize what I take to be of primary importance in this section. Even non-inferentially elicited reports are not simply the product of the world acting upon the agent. Both differential responsive dispositions and endorsement feature in perceptual reports. A differential responsive disposition denotes a reliable relation between world and agent, supplying

¹⁹ Sellars, *Empiricism and the Philosophy of Mind*, 107.
a necessary condition for knowledge. But it is not a sufficient condition because it is unable to account for the epistemic attitude included in the perceptual claim. Endorsement supplies the needed epistemic dimension and proves important for at least two reasons. First, it makes room for the possibility of justification by offering a space for the epistemic appraisal of judgements. Secondly, because endorsement is understood conceptually as a form of practice, it leads us to the space of reasons in which epistemic attitudes are evaluated within the framework of taking others to be responsible and holding others responsible for their claims.

An Objection

On the Sellarsian view a reliable differential responsive disposition is not a sufficient condition for knowledge because of the possibility of systematic error, John in the necktie shop helped to demonstrate this. It might be objected that Sellars's account of endorsement within the space of reasons is unnecessary. The possibility of systematic error may be avoided more directly by building truth into the system itself, so that only those dispositions which reliably produce correct judgements count as authoritative. This way we could circumvent Sellars's claim that a reliable reporter has to internalize evaluative standards regarding the propriety of an endorsement.

According to the objector, beliefs need not be justified by an appeal to standards if we simply hold that they are justified when well-aligned with the truth and go on to determine empirically which beliefs have a high probability of being true. In that case, contra Sellars, we could bypass the normative or epistemic dimension posited to allow for justification and avoid both the space of reasons and the conclusion that there is a non-reducible normative dimension to our epistemic
language. The conflation of causal and justificatory roles could be disentangled in the opposite direction, so that justified belief would be understood in causal terms.

The reliabilist is the likely objector, and here is how he might attempt to out-manoeuvre Sellars. Reliabilism holds that a belief is justified when it is the result of a process or method which is reliable, and reliability measures the likelihood that the belief produced is true. Rational agents are those who are dependably caused or disposed to hold true beliefs about the world, thus rational beliefs are those which are dependably produced by a process of the appropriate sort. An agent who develops the capacity reliably to have true beliefs about some feature of her environment develops a reliable responsive disposition with respect to (at least) that feature of the environment. A belief is justified just in case it is produced by cognitive processes (perception, memory, reasoning, and so on) which are generally reliable and which cause or causally sustain it. The determination of which beliefs are and which are not justified is thus factual and not normative. Thus, the objector concludes, 'endorsement' can be bypassed in favour of an appeal to those processes which yield a high ratio of true beliefs.

According to reliabilism, an agent is an authority on \( p \)-beliefs because she is able to make \( p \)-judgements correctly and non-accidentally. Her \( p \)-claims are authoritative because her \( p \)-beliefs are justified. This authoritative status can further authorize \( p \)-beliefs in others on the basis of her testimony. Her authority thus extends beyond her capacity to arrive at her own justified beliefs, for once the reliable responsive disposition is formed, and so long as it remains intact, there is no reason why it cannot play a role in the justification of \( p \)-beliefs in others who have come to take

---

the agent's testimony regarding $p$-claims as authoritative. The reliability of the psychological processes which cause true beliefs is what is at issue. It is a consequence of this view that the person in whom the reliable process exists need not appeal to (or even be aware of) any reasons which could justify the belief. The focus on reliable causal processes provides a picture of justified belief according to which a fact about, or a state of, the believer (that she has a reliable responsive disposition) determines the normative status of her beliefs. Thus, epistemic norms are reducible to facts.

The psychological processes involved in the justification of more abstract beliefs pose a challenge for the reliabilist because it is less obvious how they might provide a reliable link to the truth. How, for example, is the belief that $'2 + 2 = 4'$ justified according to the causal view? There are no perceptual takings of 2's and 4's which produce the belief because the world does not contain 2's and 4's. Because psychological processes are the critical determinants of justification, abstract beliefs like $'2 + 2 = 4'$, indeed the whole of mathematical knowledge, will have to be explained with reference to brain states and other such empirical data. While this poses a real challenge for the reliabilist, it doesn't serve as a knock-down argument. It remains possible that a more detailed account of the psychological processes involved in such abstract beliefs can distinguish them, and meanwhile the apparent simplicity of the position has an appeal. Nonetheless, as we shall see, reliabilism fails to provide a purely non-normative account of

\[21\] It is important to distinguish between a moderate or weak reliabilist view and a strong reliabilist view. On the latter, in order for a $p$-belief to be justified the truth of $p$ must be an essential feature of the causal chain producing the belief and so it is difficult to see how beliefs which are not based on direct sensory experience of the world could be justified. On the former, is not the case that the fact that $p$ must play a causal role in the formation of the $p$-belief, but less stringently that a reliable psychological process result in a true belief. So there need not be a 'real natural object' which initiates the belief-production process.
A Sellarsian Response

As indicated earlier, Sellars’s argument against the myth of the given is not just directed against classical empiricist or phenomenalist forms of the myth, which appeal to sense data, but more broadly against any theory which conflates the descriptive and the normative. While reliabilism avoids the first-person phenomenalist form of the myth, it does not avoid the myth of the given entirely. It is guilty of conflating the normative and descriptive by smuggling in normative attitudes under the guise of facts about the truth-aptness of psychological processes. Sellars’s argument shows that a normative commitment to standard conditions is inescapable. The reliabilist’s attempt to circumvent a commitment to epistemic norms gives rise to what has been called ‘the range problem.’

The range problem asks us to consider the following question: how ought we to determine the scope of application of a particular process in order to determine its reliability? To answer this we need to know whether to take a process’s reliability to be the truth ratio it exhibits in the actual world, or whether we should consider how it would fare in other possible worlds as well. A well-discussed example helps to clarify what is at issue.²² A person may be a reliable detector of barns locally and yet if she unknowingly happens upon a county in which most objects which look like barns are in fact cleverly appointed barn facades then her status as an authority about

---

the presence of barns will be altered. If she claims 'there is a barn' and it happens that what she spies is really a barn, then the processes which produced in her a barn belief produce a true belief which nonetheless fails to be reliable because it was simply good luck and not a discriminatory capacity which produced the true judgement. In barn facade county, the agent's claim that she sees a barn doesn't express knowledge. Hence, it seems that knowledge is relative to the environment.

This is important because it demonstrates that the reliability of a belief-forming process varies with its circumstances. When not in barn facade county our reliable barn detector manages very well and her success depends on the difference in her surroundings, not on a different set of psychological processes. So reliability depends upon the environment in which a disposition is exercised. This seems a reasonable conclusion to draw: if she lives in a world in which barn facades are plentiful then the reliable responsive disposition acquired in the non-barn facade world would prove inadequate. Indeed, there may be no more carefully developed disposition which would constitute a reliable responsive disposition in such a world. It seems that reliability alone cannot determine the scope of the authority of knowledge claims. The question of the source of this authority is raised all over again and we are forced to distinguish between normative and explanatory adequacy.

In his discussion of this problem, Robert Brandom offers a Sellarsian response to the reliabilist challenge by first questioning how it is that the reference class for reliable beliefs is determined.23 The claim that reliability is an objective matter, determinable by reference to the

objective probability that the beliefs it produces will be true in the specified circumstances, leads to the question of how we are to specify the circumstances which will determine the probability of true beliefs. Under what conditions ought the probabilities of specific sorts of judgement to be assessed? The barn facade example is useful because it demonstrates the variety of options available. If the class of judgements relevant in a determination of the probability assessment is a class of one, in which the agent's belief is true, then a high probability results, but if the reference class includes judgements made within barn facade county then the ratio will be very low. Yet, if the reference class includes barn detection judgements from across the country, the probability ratio will be reasonably high. Which objective fact should determine the probability assessment? Any justification of one reference class over another will rely upon a normative practice and thus a purely causal account is impossible: "There is no objective matter of fact as to which of the various possible reference classes to which the case in question might be assimilated is the right one to assess reliability with respect to." Reliabilism fails because it cannot make good its claim to specify justificatory standards in a purely descriptive manner. It fails to take into account the practical (Brandom calls them 'deontic') attitudes of taking or treating performances as correct or incorrect and, in effect, fails to demonstrate that the endorsement of reliable responsive differential dispositions is superfluous.

It should be noted that Brandom takes Sellars's actual position to be somewhat stronger than the position argued for here. According to Brandom, Sellars argues that for the claim of a reliable observer to be justified, the observer must be able to endorse the claim reflectively by way

---

of an inference in which the claim to reliability is an explicit premise. In other words, in order to be authoritative, non-inferentially elicited claims must be able to be justified by the maker of the non-inferential report. Here justification demands that the reporter explicitly identify herself as reliably making correct judgements of this sort. However, it seems reasonable to hold that an agent can have a justified belief without knowing that she is justified in holding it. If Sellars does hold this view, it is too strong. Indeed, Brandom suggests that for Sellarsian reasons we ought to deny that the claims of non-inferential reporters are authoritative only if they can be justified by the reporter. Reliability is important, but not in the first-person manner that internalism supposes it to be. "Reliability may entitle the reporter to the knowledge claim, may qualify it as knowledge, even if the reporter does not even implicitly endorse the inference that is the practical acknowledgement of the authority of reliability." For my purposes, the details of justificatory internalism and externalism need not be sorted through here. The less stringent justificatory option fits well within the broadly Sellarsian account I am advocating.

Fundamentally, Sellars's criticism of reliabilism shows that the very project of giving a purely naturalistic account of norms is misplaced. This is not to say that we cannot distinguish between objective kinds, "those whose instances are what they are regardless of what any particular community takes them to be," and socially determined kinds, whose instances are whatever we take them to be. It is just that this distinction is one made within our practice and

25 Brandom, Making It Explicit, 213-221.

26 Brandom, Making It Explicit, 219.

thus cannot reach more deeply into nature than our own practices allow. Accordingly, for Sellars, the very distinction between the natural and the social is socially constituted. This need not entail that objectivity is illusory, nor that freedom is a mysterious contra-causal metaphysical force. When we describe something as norm-governed (for example, as an action or a piece of knowledge) we are classifying it as something which meets (or fails to meet) standards which we impose in our understanding of it. It is, in effect, the kind of thing which one of us would, could, or should do. To be an action governable by norms is not a matter of what it is in itself as an objective kind, but instead a matter of how the behaviour is treated by some community. Descriptive facts are themselves socially delimited. It should be clear that while reliability is a necessary condition for knowledge, reliabilism cannot account for the ways in which that relation is taken up by the community.

**Three Lessons for Practical Reason**

I have contended that Sellars offers some important insights into the nature of theoretical reasoning and suggested that these insights have application for a study of practical reason. Before turning, in the next chapter, to an analysis of their import for instrumental reason I will briefly summarize what I take to be the fundamental insights. First, the myth of the given arises out of the confusion of causal and justificatory orders. The mongrel concept of the sense datum provides a good example of how causal and justificatory roles are confused because the awareness of sense data is, in effect, made to do the work of both describing and justifying the contents of one’s sensory awareness. Sellars shows that these orders of explanation must remain
distinct. The inconsistent triad clarifies the point at which the confusion is generated. While Sellars introduces his critique by discussing the classical empiricist version of this conflation, where the sense datum is the culprit, the myth is not primarily about the problem of locating knowledge in the awareness of the knower. It is rather that in this form of the myth the awareness of the knower is purported to be both a fact about the agent and epistemically authoritative. Reliabilism shifts the problem from the first to the third person by replacing the empiricist's act of awareness with a reliable belief forming process, as the fact about the agent which has normative import. This fails to avoid the myth of the given. So, lesson number one is that we should beware of the conflation of causal and justificatory explanations.

Second, the discussion of 'looks' and 'is' talk confirms that reliable responsive dispositions are a necessary but not a sufficient condition for knowledge. The fact that an agent has a disposition to respond to her environment in a particular manner does not, on its own, constitute an act of knowing. Differential responsive dispositions may consistently get things wrong and are not always taken by agents to be indicators of the truth. What is needed, and the differential responsive disposition cannot provide, is a means for introducing normative or epistemic relations into the account in the form of doxastic commitments. This cannot be managed within a purely causal framework. So, lesson number two is that we cannot understand or explain epistemic judgements in purely causal or descriptive terms. As we will see in the next chapter, the failure of the instrumentalist account of practical reasons is that it attempts to explain normativity in purely descriptive terms.

Thirdly, and finally, normativity must be built into epistemic practice at a very basic level in order to explain how it is that epistemic (normative) judgements are obtained. By
understanding epistemic judgements as normative commitments we avoid the problem of explaining how values can be derived from facts and do so without conflating the causal and justificatory orders. The introduction of endorsement as a means of explaining normative commitment is the key to the Sellarsian project, and key to avoiding the myth of the given. Endorsement ultimately leads us to the space of reasons and the practice of giving and asking for reasons. So, the third lesson is that in order to understand the justification of judgements we must attend to the commitments undertaken by rational agents in the space of reasons. This turns out to be more difficult than one might think. Christine Korsgaard, whose views I examine in Chapter Four, grants that the authority of reason is irreducibly practical and offers an account of the commitment involved in practical reasons which locates the source of their authority in practice. By conceiving of practice in the first person, however, her position fails to avoid making the authority of practical reasons into facts about the agent, thus pointing to the need for an account of commitment which is irreducibly social. This is the topic of Chapter Five.
Chapter Three:  
Instrumental Reason and the Myth of the Given

Philosophical investigation into practical reasoning has long been motivated by the thought that because successful practical reasoning issues in action, and successful theoretical reasoning does not, practical reasoning is of necessity grounded in some type of motivating force. Theoretical judgements are not per se productive of action. Thus, in order to explain how theoretical judgements can be made practical, a motivating force is required. The Humean thought that, "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them", supports the view that practical reasons are reasons grounded in an agent's desires. Reasons serve to satisfy non-rational ends. Practical rationality is thus thought to be end-neutral in the sense that an account of what it is rational to do does not incorporate a conception of the good. Reasoning about what to do, it is claimed, differs fundamentally from reasoning about what to believe because what one has reason to do will depend crucially upon desires, while what one has reason to believe will be dictated not by desires but by relevant features of reality.

An instrumentalist theory of practical reason holds that reason's role in action is limited

---

to generating means to ends, ends set by something other than reason itself. Reason points the way to desire satisfaction by providing means to ends which an agent already, in some sense, possesses. The proviso, \textit{in some sense}, is important because in order to respect the capacity of deliberation to produce in us states of desire, it has to be possible that an agent could come to have new desires because she has reasoned something through.\footnote{Thomas Nagel describes the difference between these two different types of desire-states as "motivated", those arrived at after deliberation, and "unmotivated desires", those which simply befall us. \textit{The Possibility of Altruism} (Princeton: Princeton University Press, 1970) 29.} On this way of thinking, the imperatives of practical reasoning are only hypothetical. By contrast, a non-instrumental account of practical reason holds that reasoning about action directly constrains practical judgements. According to the non-instrumentalist, practical reason provides (at least some) categorical imperatives. Non-instrumentalists hold that reason is not merely "a tool designed solely either to achieve or to discover the ends of our actions, but as something that, in addition to its instrumental functioning, quite literally, defines ends of action."\footnote{Jean Hampton, \textit{The Authority of Reason} (Cambridge: Cambridge University Press, 1998) 232.}

Generally speaking, instrumental accounts of practical reason adhere to two related ideas. First, practical reasons are grounded in motivational states. Action demands an unmoved mover of sorts, and thus reason can serve only to promote the most effective realization of goals antecedently given. Our preferences are criticised only on instrumental grounds. In other words, when we criticize a person as practically irrational, we do not criticize the content of her chosen end, but her means of attaining it. Second, all reasoning is theoretical – instrumental reasoning is just theoretical reasoning in the service of action. There is no such thing as a special
form of practical rationality: "reason is neither normatively nor motivationally practical: it grounds neither judgements of what we should do nor motivation to do anything." While the first of these suppositions attracts the most discussion, I will argue, that it is the second which poses the greatest problem for the instrumentalist.

An abundance of the recent literature on practical reason has focussed on the nature of the motivating force underlying reasons for action. According to reasons internalism, the fact that an agent has a reason to act implies that the agent is accordingly motivated (although not necessarily sufficiently motivated). Yet, even if it is granted that a motivating force is a necessary condition of having a reason for action, it remains open whether the motivating force must be an occurrent state in order to be, in the relevant sense, accessible to the agent. And of course, the question of just what counts as a motivating force also remains open. If understood as an occurrent state of the agent, the motivational requirement is too strict, and poses problems for prudence and rational planning. Thomas Nagel gives the example of a person for whom prudence dictates learning some Italian because he will be travelling to Rome, but who is never moved by an immediate desire to do so. Thus he never does what it seems he has reason to do. Instead, he contemplates the arrival of the reason with detached curiosity, knowing that at some point the

---


5 In an attempt to broaden the notion of what counts as a motivating force Bernard Williams introduces the idea of the "subjective motivational set", which contains a wide range of possible motivators including principles as well as desires. "Internal and external reasons," in his *Moral Luck* (Cambridge: Cambridge University Press, 1981) 101-113.

desire will present itself, although likely when it is too late. On the other hand, if the motivating force is not understood as an occurrent state of the reasoner, but as a state which is rationally accessible in some other way, then it isn’t clear just how it serves to ground reasons for action in an action-producing feature of the agent.  

The recent focus on the internalism requirement has obscured a more important issue—the question of how judgements, practical and theoretical, rationally bind agents. Instrumentalists generally assume that theoretical reasoning is authoritative and hence that the authority of the hypothetical imperative is obvious. This assumption needs challenging. In the last chapter we saw that theoretical judgements incorporate, or depend upon, irreducibly normative commitments. Yet, the instrumentalist proceeds as though the authority of theoretical reasoning is a brute fact of experience, pointing to psychological tendencies which are supposed to serve as non-normative descriptions of reasoning which simultaneously justify reason’s authority. I will argue that this is the instrumentalist’s version of the mongrel concept.

For the purposes of my project, instrumentalism can be rather broadly defined. I am concerned with the nature of the rational commitments involved in having reasons for action. Hence, my treatment of instrumentalism concentrates on its capacity to account for these rational commitments. Because my concern is with the more fundamental issue of how practical judgements can be normatively binding and still guide action, an issue which (on my view) is not illuminated by examining the details of the instrumentalist’s narrowly circumscribed account of

---


8 As Jean Hampton puts it, “the “ought” in a hypothetical imperative is just as puzzling as the “ought” in a categorical imperative.” *The Authority of Reason*, 127.
rational constraints, I will not here consider the question of which principles best represent the thinly articulated rational constraints which the instrumentalist claims authorize reasons for action.⁹

As we shall see later in this chapter, the standard instrumentalist principle that desires are not rationally criticisable is challenged by Michael Smith, who nonetheless, takes his position to be instrumentalist. Positions like Smith’s, which intend to solve the more fundamental problem, are important to examine because, whether or not they actually live up to their instrumentalist precepts, they are commonly perceived as avoiding difficulties which plague their non-instrumentalist counterparts. I will argue that such positions cannot explain the normativity of reasons while remaining within the instrumentalist framework. By examining them, I will confront a more fundamental question, which instrumentalism obscures, that is, ‘how are reasons, practical and theoretical, normatively binding?’ Instrumentalism obscures this more fundamental question because it presupposes that the authority of theoretical reasoning is simply given in experience. It supposes that the question: ‘how can practical reasons be theoretical judgements and still motivate action?’ is more basic. My discussion of instrumental reason aims to clarify what the instrumentalist obscures, and in doing so to draw attention to the irreducibly normative commitments that are part of practical reasoning.

My arguments are not directed to the question of which substantive principles of practical reason are correct. In many cases the normative demands that instrumentalists defend seem

---

⁹ Even expected utility theory, arguably the best expression of instrumentalism as a minimalist thesis about rationality, is amenable to a non-instrumentalist interpretation. For a discussion of instrumental and non-instrumental interpretations of expected utility theory see chapter seven of Jean Hampton’s, *The Authority of Reason.*
reasonable. Rather, my arguments are intended to show that the question of how we are bound by reasons generally is fundamental to an analysis of practical reasons. My strategy in this chapter is to focus on the normative authority of practical reasons -- their capacity to rationally bind agents. I argue that to the extent that they remain true to the thought that practical reason only indirectly constrains action, instrumental accounts of practical reason fail to explain how practical reasons rationally bind reasoners. Instrumentalism about practical reason cannot explain how reasons authorize action (cannot explain their normative authority) without giving up the idea that all reasoning is theoretical, and thus giving up the idea that practical reason is wholly instrumental. My strategy will be first to show that the crude belief-desire model which some instrumentalists employ lacks the theoretical means to explain the normativity of reasons for action. Crude instrumentalism conflates descriptive and justificatory orders of explanation by grounding reasons in non-normative dispositional states of the agent. Next, I will show how sophisticated versions of instrumentalism, which attempt to circumvent the problem of crude instrumentalism, while explaining the normativity of reasons for action, either slide back into crude instrumentalism or fail to remain true to the tenets of instrumental reason. Sophisticated instrumentalism fails to explain the authority of reasons because it supposes that all reasoning is theoretical and its authority self-evident. I conclude that even the sophisticated instrumentalist must hold that practical judgements are directly constrained by rational norms, not all of the imperatives of reason are hypothetical.
The Promise of Instrumentalism

Part of the appeal of instrumentalism is its apparent simplicity. Out of a set of facts about our psychology (what we want and desire) and relatively few rational constraints, an accounting of what we have reason to do is generated. Instrumentalists take their account to provide a relatively straightforward explanation of how having a reason to act can lead one to act. As the instrumentalist sees it, reason simply uses the data provided by our beliefs and desires to arrive at an efficient means of attaining one's ends. The instrumentalist is committed to the view that reason cannot itself provide ends but only means to ends, and to a characterization of reasons as explainable in natural or properly scientific terms. Instrumentalists hold that they are able to clarify the normative force of practical reasons by appeal to minimal rational constraints and facts about the psychological make-up of agents. Thus, they claim, their position has a simplicity and neutrality which eludes the non-instrumentalist.

Some philosophers believe that instrumentalism circumvents a variety of metaphysical problems which non-instrumentalism cannot. The instrumentalist strives to work within a naturalist framework in which claims about intentions, reasons, moral judgements can be identified with (or explained by) natural features of the world. Instrumental theorists want, says Jean Hampton, "a conception of reason that grants it no occult powers, and that presupposes a foundation that is utterly acceptable from a scientific point of view."\(^{10}\) The alleged need for non-instrumentalists to appeal to "mysterious acts of the will, or kinds of causation foreign to

\(^{10}\) Jean Hampton, *The Authority of Reason*, 126.
science,\textsuperscript{11} or queer sorts of objective facts which themselves possess the motivation to act in accordance with their dictates,\textsuperscript{12} leads many philosophers to think that an instrumental account is the only plausible option. But, of course, by embracing (typically undefined) naturalist precepts, instrumentalists do not show that non-instrumentalist accounts of practical reason are inconsistent with naturalism. Instrumentalists don't have the market cornered on naturalism, nor on adequate conceptions of science. We shouldn't be swayed by blanket instrumentalist appeals to naturalism, particularly in the absence of arguments which show that non-instrumentalists cannot similarly avail themselves of the goods of naturalism. Whether or not the instrumentalist can adequate explain the normativity of reasons using purely non-normative concepts is precisely what is at issue.

Before moving on to a discussion of crude instrumentalism, let me first clarify the sense of the term 'reason for action' in which I am interested. When we speak of an agent acting \textit{for} a reason (acting intentionally) we suppose the agent to possess a set of standards according to which that action is defensible. So to have a reason for action is to be in a state which, in some sense, commends or advises doing the action, a sense of recommendation or advice which must be accounted for by that reason. Any account of practical reason (instrumental or non-instrumental) must be able to explain both reason's capacity to guide and its capacity to authorize

\textsuperscript{11} Donald Davidson tries to provide an account of intending which doesn't rely upon any such mysterious acts of the will in "Intending," \textit{Essays on Actions and Events} (Oxford: Oxford University Press, 1989) 83.

\textsuperscript{12} Mackie's famous 'argument from queerness' is found in his \textit{Ethics: Inventing Right and Wrong} (Toronto: Penguin Books, 1977).
Purely explanatory models of reasons for action fail to do so. Explanatory reasons offer accounts of why an agent acted as she did which are not restricted to reasons available to the agent. For example, I might explain a friend’s congenital lateness by saying that ‘she is late because she is disorganized,’ a description of my friend which may be accurate, yet not an adequate explanation of her reason for action. Her disorganization may be part of a larger causal structure which makes it likely that she will regularly arrive late and which, given suitable information, will enable us to predict her behaviour. But her own reason for arriving when she does is not likely to have been ‘I will arrive late because I am disorganized’. In the sense in which I am interested in reasons for action “neither ignorance nor prejudice can be a person’s reason for acting.”

Reason’s practical application should, at least in principle, have some consequence for the actor, just as reason’s theoretical application should, at least in principle, have some have consequence for the believer. This is not to say, of course, that to have a reason is to act on that reason. Not only would it be false to say this, it would be uninformative. If all action were necessarily rational, then irrational action would be impossible.

To say that reasons for action are action-guiding is minimally to say that they play a privileged role in the regulation of the agent’s conduct. Determining that I have reason to φ makes it more likely that I φ than it would have been had I not determined that I had reason to

Of course, there are many philosophical issues on which a theory of practical reasoning may focus its attention. In “A Theory of Practical Reasoning,” Robert Audi points to six such issues. American Philosophical Quarterly, 19, (1982), 25-39. My point is just that accounts of practical reason which are substantive enough to underpin moral theory must be able to clarify the structure of reasons for action in such a way that actions can be understood to follow from reasons for action.

do so. There is clearly a connection between having a reason to φ and being motivated to φ. Proponents of instrumentalism take themselves to be in a good position to explain how reasons guide in this motivational sense.\(^{15}\)

**Crude Instrumentalism**

In this section I briefly outline instrumental reason in its simplest form (represented by a crude belief-desire psychology) and show how it conflates causal and justificatory forms of explanation in an attempt to ground reasons in psychological states of the agent. I argue that when reasons are simply conceived of as belief-desire complexes, they are Sellarsian mongrel concepts, illegitimately employed to both describe and justify the contents of one's psychological state. In much the same way that a sensing of sense contents fails to constitute an instance of knowledge, the awareness of one's beliefs as relevantly connected with one's desires fails to constitute a reason for action. Undeniably, the belief-desire model has an appeal. If we assume that people act so as to achieve their desires we go some distance to explaining the apparent rationality of and motivation for action. But then again, sense data also had their own appeal. By purportedly both being the cause of beliefs and providing justification for those beliefs, sense data appeared to bridge the gap between causal explanation and rational justification. Similarly, if we think of belief-desire complexes as states of an agent which simultaneously cause action and

explain why those actions are reasonable, we get a mental state which appears to provide reasons for actions just by causing them.

I am calling 'crude instrumentalism' the view that marks the simple but powerful thought that we can understand why a person did what she did by making reference to her desires and her beliefs about how to achieve those desires. Donald Davidson's early work on action theory provides a good example of an approach motivated by the belief-desire model. Davidson argued that in order to make sense of an agent's behaviour, we must assume that she is rational.\textsuperscript{16} We must suppose, moreover, that her (largely) rational behaviour reflects action aimed at maximal preference satisfaction. According to Davidson, primary reasons (reasons which account for particular actions) are explainable by reference to an agent's belief-desire pair:

\begin{quote}
\textit{R} is a primary reason why an agent performed the action \textit{A} under the description \textit{d} only if \textit{R} consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that \textit{A}, under the description \textit{d}, has that property.\textsuperscript{17}
\end{quote}

'Pro-attitudes,' as the name suggests, constitute a broader class of emotive states than simple desires, thus making it more likely that the model will be able to explain a wider range of reasoned behaviours.\textsuperscript{18} Nonetheless, it quickly becomes apparent just how crude and inadequate such an


\textsuperscript{18} Davidson includes in the category of pro-attitudes: desires, wantings, urges, promptings, moral and aesthetic convictions, economic prejudices and social conventions, public
account is.

Crude instrumentalism of this sort seems promising on the surface because it appears to provide an account of how normative reasons might supervene on psychological states of the agent, and show that practical reason is reason in the service of desire satisfaction. However, Davidson's account actually shows that the belief-desire complex is too insubstantial to serve as an account of practical reason. At times, we are disposed to act in a manifold of ways. In deliberating about what to do we often notice ourselves being pulled in opposite directions. The crude account lacks the means to identify an agent's commitment to a course of action with a belief-desire complex. In committing to a course of action the agent gives herself reason to act on one belief-desire complex rather than another. Crude instrumentalism fails to explain how a commitment to a particular belief-desire complex informs her intention.

Of course, the crude instrumentalist might say that the belief-desire complex to which one is committed just is the one containing the strongest desire. It would indeed be marvellous if the reasonable thing to do were always what one came to want to do most. But it is unclear how, on such a view, reasoning could guide an agent. In deliberation one doesn't simply weigh the strength of all of one's belief-desire complexes in order to discover what one has reason to do. If deliberation were like that, the rationality of one's strongest desire would be pre-guaranteed; one's "strongest desire" would simply designate the winning disposition. Reasons, on such a view, would be more like natural laws than rationally defensible commitments. The problem with


the crude account is clear. If, just as an object falls to earth because subject to gravitational force, agents act because subject to belief-desire complexes, there is no room left for having a reason for action. The crude model supposes that a belief-desire complex is a necessary condition of having a reason for action because it grounds or supports that reason. In going on to claim that a reason for action just is a belief-desire complex, it misidentifies this supporting relation as an identity relation. But it is one thing to base a reason on a dispositional state and another to claim that an agent’s reason just is her dispositional state.

A belief-desire complex (a particular disposition to act) is not *per se* a sufficient condition for having a reason for action, although it may be a necessary condition. Dispositions to act one way or the other may provide us with invaluable sources of information upon which to deliberate, but do not themselves constitute reasons for action. Consider the parallel in the case of reasons for belief. I am disposed to see a straight stick as bent once it is immersed in a glass of water. However, my disposition doesn’t give me, a competent judge of stick-stability-over-time, a reason to believe that the stick has become bent. Instead, it gives me a reason to believe that things are as they should be — straight sticks usually look bent when placed in a glass of water. We might say that having a reason for action requires of its possessor a normative attitude in much the same way that we might say that reason for belief requires of its possessor an epistemic attitude.

An explanation of reasons for action requires that we be able to make sense of that reason as involving a rational standard which binds the reasoner, whether or not she lives up to it. The belief-desire complex cannot account for this sense of the normative commitment involved in having a reason for action. Without the endorsement of, or commitment to, a particular belief-desire complex as one which grounds her action, the agent is not yet in a state in which she can
be said to have a reason available to her. The instrumentalist needs to be able to make sense of the normative attitude which serves to ground a reason in a particular disposition. Having a reason for action involves bringing one's behaviour under the description of reason. Consider an example. As I rush to the meeting, I may recognize how my behavioural patterns leave me, once again, arriving late. Indeed, I may become painfully aware of how my disorganization brought it about that I will be late. But that form of awareness, because it fails to regard my disorganization in the relevant way, prevents me from taking my disorganization as a reason for action at that time. Of course, in a moment of perversity, I could adopt a commitment to disorganization as a lifestyle to be promoted. But in doing so, I would not merely observe my behaviour, but assign it a normative status. By taking responsibility for my disorganization, and my attitude about it, I could no longer perceive it simply as a causal tendency explaining my behaviour.

Crude instrumentalism cannot distinguish these cases because it conflates having a disposition to act with having a reason for action. But having a dispositional state is an empirical description of the agent, and having a reason is a normative condition. In order to explain how a dispositional state might ground a reason for action, the instrumentalist needs to show how a normative attitude can function to authorize some dispositions and not others. Our discussion of Sellars has shown us how an agent's dispositions can come to signal that the agent has a reason when those dispositions are appropriately bound up in larger reason-giving practice. John's experience in the necktie shop taught him that his own differential responsive tie-identification disposition was a source of information on which to ground his beliefs ('blue ties look green under electric light') but didn't provide the content of the belief itself in any straightforward way.
Dispositions to act aren’t themselves reasons for action.

**Sophisticated Instrumentalism**

The problem of the normativity of practical reasons features more prominently in what I am calling *sophisticated instrumentalism*, than it does in crude instrumentalism. Sophisticated instrumentalists attempt to explain the capacity of reasons to authorize action without transgressing the tenets of instrumental reason. In order to do so, sophisticated instrumentalists need to account for the normative attitude, or commitment, involved in having a reason for action without claiming that reason directly constraints judgement about action. I take as exemplars of sophisticated instrumentalism Michael Smith’s fully rational advisor model and Michael Bratman’s fully rational planner model. Both see instrumentalism as a way of avoiding the purportedly controversial claim that rational norms of conduct directly determine the wills of rational agents. According to them, rational deliberation produces beliefs which are not *per se* action guiding. Reasons for action cannot be directly given through rational deliberation because the products of reasoning (beliefs) are inert. Our theoretical judgements by themselves do not directly guide action. Norms of action are tied to our appetitive natures in crucial ways.

Smith and Bratman develop different instrumentalist insights in the course of attempting to explain the normative force of reason. Smith focusses on working normative commitments into theoretical reasoning by way of what we desire. Bratman tries to work normative

---

20 Hereafter, I will refer to sophisticated instrumentalists simply as instrumentalists, unless distinguishing them from crude instrumentalists.
commitments into an account of rational intention formation. Although they both see it as a virtue of instrumentalism that it sticks to a "nonmysterious theory of the will."\(^{21}\) I argue that both make the normativity of reasons no less mysterious than the Kantian alternative we will examine in the next chapter. Instrumentalism promises to build an explanation of reasons for action (standards of conduct) out of non-normative raw materials — psychological facts and simple rational constraints. But both Bratman and Smith are unable to explain reasons for action without presupposing normatively binding standards of rational conduct. The attempt to build norms of conduct out of desires and some facts about rational behaviour is undermined by the presupposition that substantial normative commitments are built into rational deliberation. Or so I shall argue.

**Smith’s Fully Rational Advisor**

Smith’s widely discussed position defends the instrumentalist conviction that reasons are practical in their issue because constituted in part by desire. His position is firmly fixed by the instrumentalist thought that there is no specifically practical form of rationality, that practical reasons are practical because grounded in desires. Properly speaking, practical reasons are the product of theoretical reasoning. Smith wants the purported simplicity and neutrality of instrumentalism, but recognizes that an overly simple belief-desire model (such as the crude view) fails to explain the normativity of reasons. His work is important because it explicitly strives to

construct a normatively robust account of reasons as rational standards out of non-normative facts about the psychological properties of agents, and reason itself together with a lavish sprinkling of subjunctives and hypotheticals. My concern is not with the details of the normative standards he defends, but with the suggestion that its normativity is immediately given in experience.

Smith, like many other philosophers, sees the problem of the status of moral judgements as a direct consequence of a belief-desire model. If moral judgements are judgements of fact, then what explains their capacity to motivate action? On the other hand, if they are simply expressions of an emotive state of the agent then what, if anything, explains their capacity to be objective? Adherence to a belief-desire model for practical reason leads Smith to formulate and offer a resolution to what he calls 'the moral problem,' by explaining how the following three *prima facie* inconsistent propositions can be made consistent:

1. Moral judgements of the form 'It is right that I φ' express a subject’s beliefs about an objective matter of fact, a fact about what it is right for her to do.

2. If someone judges that it is right that she φs then, *ceteris paribus*, she is motivated to φ.

3. An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume’s terms, distinct existences.\(^\text{22}\)

Smith’s starting point is the tension between the claim that moral judgements are

---

expressions of fact and the claim that action arises from desire. If action is ultimately grounded in desire then reasons for action must ultimately be grounded in desire. However, if reasons for action are bound by the limits of an agent’s desires, and beliefs and desires are “distinct existences,” then action from moral reasons would be impossible since, according to Smith, moral judgements are judgements of fact, not expressions of desire, and thus have entirely different logical properties. Since facts about what we believe and desire aren’t themselves reasons for action, how could there be moral reasons for action? Smith proceeds with the conviction that reasoning about belief and reasoning about action are fundamentally different activities, and then tries to bridge the gap between judgement and action by introducing a conceptual connection between two different sorts of reasons. It is his starting position that is mistaken.

Smith distinguishes two sorts of related reasons, taking as basic the belief-desire model and what he calls motivating reasons. Motivating reasons are psychological states of the agent:

\[ R_{at\ t} \text{ constitutes a motivating reason of agent } A \text{ to } \phi \text{ iff there is some } \psi \text{ such that } R_{at\ t} \text{ consists of an appropriately related desire of } A \text{ to } \psi \text{ and a belief that were she to } \phi \text{ she would } \psi. \]

Smith’s account of normative reasons is intended to establish a normatively adequate conception of reason, one capable of accommodating objective moral reasons. Normative reasons are not facts about the psychological make-up of an agent, but rather propositions of the form “A’s \( \phi \)-ing is desirable or required.” To say that an agent has a normative reason is to make a truth claim, where that judgement is justified according to the normative system by which it was

\[ ^{23} \text{Smith, The Moral Problem, 92.} \]
sanctioned. Conflations of these two different categories of reason are really category mistakes. Smith provides the needed connection between motivating and normative reasons by conceiving of normative reasons as judgements about what one would want to do in a relevant circumstance if one were fully rational: "facts about what it is desirable for us to do are constituted by the facts about what we would advise ourselves to do if we were perfectly placed to give ourselves advice."24

Motivating reasons emphasize the explanatory dimension of reasons (their capacity to explain our behaviour) while normative reasons emphasize the justificatory dimension of reasons (the sense in which standards of rationality function to authorize action). Motivating reasons are explanatory because in virtue of being in the state of having a motivating reason an agent is, other things being equal, in a state which is explanatory of her action. In motivating reasons, the relevant desire and belief must be combined ('appropriately related') in order to constitute a reason for the agent. The agent must 'put the relevant desire and belief together.' Once connected in the appropriate way, the belief-desire complex gives the agent a goal. This goal-imparting feature of motivating reasons is important to Smith’s position as it provides the means to offer an explanation of the commitment involved in having a reason for action.

First, he argues that having a motivating reason is *inter alia* having a goal. Second, 

---

24 Smith, *The Moral Problem*, 152. It is worth noting that being fully rational and being perfectly placed to give advice are far from being synonyms. Being rational and being in possession of the relevant facts are not the same: if I don’t know that the heroin will addict me if I continue to regularly inject it then it isn’t clear that reason has much to say against my doing so. Realistically speaking, we want agents to be reasonably well informed and we hold them responsible for failing to have beliefs they ‘should’ have, but not for failing to know everything.
having a goal is being in a state with which the world must fit. Third, being in a state with which the world must fit is desiring. It follows that the desire-belief model provides the necessary and sufficient conditions for having a motivating reason. The argument is simply that in having a reason I have a goal, and in having a goal I must have a desire and an appropriately related means-end belief. By introducing the idea of having a goal Smith goes beyond the crude belief-desire model. To have a goal is to commit to a disposition to act, and in doing so, to have overridden others.

The claim that to have a goal is to have a reason is thus doing a lot of work for Smith. Because goals are desire based, Smith looks to have a way of explaining reason’s capacity to guide. Motivating reasons offer guidance precisely because they are goals. Smith’s appeal to goals goes some distance, although I will argue not far enough, toward addressing the need to incorporate commitment into his account of practical reasons. What is needed, in order to introduce a measure of rational constraint, is an account of the way in which commitment rationally authorizes action.

Our rational commitments extend beyond our adopted goals, of course. Reasons are the sorts of things we can fail to live up to. Our reasons are (well or badly) constrained by rational considerations that rule out some actions and recommend others. If I hope to pass my French exam and believe that studying will facilitate the achievement of this goal then it seems that I have reason to study French - I would be well-advised to do so. This advice recommends a course of action and is authoritative in the minimal sense that it is based in a non-arbitrary set of standards for rational action, but not simply non-arbitrary. One could try to live according to the principle

---

25 Smith argues for a dispositional account of desires.
'do whatever the person you least admire would do.' This principle is non-arbitrary, but not sufficiently rational or authoritative. What is needed is a distinction between what we take ourselves to have reason to do and what we have reason to do. In order to provide this distinction Smith develops his account of normative reasons.

Normative reasons are not psychological states, but facts or judgements. According to Smith, if I believe that I would want to \( \phi \) were I fully rational, then I have a normative reason to \( \phi \). Clearly, he needs to explain how the connection between what I do in fact desire, and what I would desire, were I fully rational, supplies a rational commitment. On Smith's account, "If an agent believes that she has a normative reason to \( \phi \), then she rationally should desire to \( \phi \)."\(^{26}\) If she fails to make this move, what sort of mistake does she make? Consider an example. Perhaps I am tired and grumpy and want to take a long, hot bath rather than grade term papers, the latter course of action being the one recommended by my fully rational self. The conflict between my motivating and normative reasons could be resolved in a variety of ways. I could drink several beers as a means of ensuring that my normative reason would be less influential. Or, I could develop a policy of ignoring normative reasons which have the consequence that strongly desired immediate forms of gratification would be missed. In short, I could experience the conflict between my motivating and normative reasons in such a way that the presumed authority of the normative reason is called into question. Smith owes an account of the authoritative force of normative reasons over motivating reasons which does not violate his commitment to the instrumentalist principle that reason constrains practical judgements only indirectly.

We do make rational mistakes, says Smith, when we fail to be appropriately swayed by

our normative reasons. His notion of the fully rational self is to be thought of as an advisor to the present self, not as an ideal with which the present self should conform: "the model is not one in which we are supposed to emulate the behaviour of our fully rational selves, or to treat their behaviour as an example we are to follow, but rather one in which we are supposed to think of our fully rational selves as perched above us, in a superior position to give us advice about what we are to do in our less than fully rational circumstances." Our fictional, fully rational self doesn’t simply provide an example for us to be moved by, or not. Instead, it advises us given the particularities of the circumstances we are in. A failure to live up to one’s normative reasons is thus a failure of rationality on Smith’s view. His position implicitly appeals to the idea that agents are committed to doing what reason demands of them because, to the extent that ideal selves represent what reason demands of us, a failure to live up to their advice will be a failure of rationality. The question is whether Smith’s assumption presupposes what it is supposed to explain, namely reason’s authority in the guidance of action.

Smith, unlike the crude instrumentalist, sees that he must bridge the gap between what we actually want to do and what we ought to do. He attempts to close this gap counterfactually by distinguishing between our actual wants and what we would want were we fully rational, making the latter normative for us. Normative reasons outrun what we want, and what we do, because they incorporate rational standards. The problem for Smith is to explain how normative reasons (judgements of fact) provide rationally authoritative standards, without claiming that practical reason gives us non-contingent ends. Smith’s account directs the agent to act in

accordance with motives which his hypothetical fully rational self has, and which by hypothesis, he doesn’t have. I will argue that counterfactuals about our fully rational selves don’t explain their normative authority over us without presupposing a commitment to doing what reason demands of us. I am not denying that agents must have some sort of commitment to doing what reason demands of them. Nor am I arguing (although one could) that our ideal desires have no authoritative role in reasoning. Rather, I am arguing that Smith’s invocation of a commitment to right reasoning presupposes what it is supposed to explain. In attempting to resolve the problem of the normative authority of reason, I will argue, his position either slides back into crude instrumentalism because it cannot explain why idealized desires express what we ought to do, or goes beyond instrumentalism by presupposing a conception of reason’s authority which is practical rather than merely theoretical.

Smith argues for the authority of normative reasons by arguing that valuing is a function of belief, not desire, so that the authority of reasons for action is just the authority of well-formed beliefs. On his dispositional theory of value, “facts about desirability are facts about the desires of our fully rational selves.”28 Indeed, he takes it as platitudinous that “what it is desirable that we do is what we would desire to do if we were fully rational.”29 Before considering his argument let me express some concern over Smith’s use of a normatively robust sense of ‘rational.’ It isn’t obvious that being prone to desire well and being rational are so intimately related. Even if we were to grant that the fully rational agent may cultivate an appropriately desiring character it doesn’t follow that the person with the vice of envy, for example, would be


29 Smith, The Moral Problem, 150.
irrational. Vices are notoriously attributable to weakness of the will and other such non-rational, human tendencies. Part of the attraction of instrumentalism lies in its purported appeal to minimal and formal rational constraints. Smith is not entitled to a normatively loaded conception of rationality. Yet, on his view desirability and rationality turn out to be so closely related that there is an analytic connection between them:

there is an analytic connection between the desirability of an agent’s acting in a certain way in certain circumstances, and her desiring that she acts in that way in those circumstances if she were fully rational: that is if she had the set of desires all agents would converge upon if, under the impact of increasing information, they came up with a maximally coherent and unified desire set.30

Although Smith uses both the terms ‘causal’ and ‘analytic’ to describe it, the connection between what is desirable and what is rational must be a conceptual, rather than a causal, relation. As a matter of empirical fact we often (indeed, almost always) fail to have the desires that we believe we would have were we fully rational. When we understand the relation as a conceptual one, we can make sense of the fact that our natures allow for the reconciliation of belief and desire through the exercise of reason. Through deliberation we arrive at beliefs which typically cause us to have corresponding desires - those that it is rational to have given our beliefs.

Despite claiming that the relation between desirability and rationality is analytic, Smith describes it in purely causal language:

Suppose that an agent who does not yet desire to φ deliberates and, as a result,

---

comes to believe that she has a normative reason to ϕ. And suppose further that her coming to have this belief then causes her to desire to ϕ. . . . For her having that belief causes her to have a desire that it is rational for her to have, given her belief. The causal transition between this belief and desire is thus on all fours with the causal transition between, say, the beliefs that p and that p → q and the belief that q; or between the belief that all the evidence supports q, or it is (most) credible that q, and the belief that q. 31

Smith takes our tendency to reason well about what to believe, and what to do, to be somehow simultaneously normatively required and simply caused. In effect, he is arguing that the authority of normative reasons is of a piece with the authority of reasoning about belief (theoretical reason). I am inclined to agree, but not, as Smith would have it, because all reasoning is theoretical and because the normative authority of reasoning about belief is causal and self-evident. By shifting the burden of the normative question to theoretical reasoning he has merely shifted the burden, he hasn’t answered the question. Smith attempts to build the authority of practical judgements into the authority of theoretical judgements, and to explain the latter as given in the nature of reasoning itself.

Smith slips back and forth between causal and normative talk so frequently that the suspicion that he is trying to have his cake and eat it too is unavoidable. This is a classic case of the conflation of causal (descriptive) and justificatory orders of explanation – the mongrel concept. He can’t have it both ways. Either he reduces normative reasons to causal dispositions and his position collapses into crude instrumentalism, or he has to allow that normative reasons are practical in their own right, in which case he must renounce instrumentalism. We have already seen that the problem with crude instrumentalism is that the very distinction between what we do

31 Smith, The Moral Problem, 179.
and what we can be held responsible for is lost. Being authorized to act is not the same as being caused to act; one could not fail to live up to the 'demands' of reason were they simply causes.

To be fair to Smith, it is important to his position that it be able to accommodate the normative, not merely explanatory, dimension of practical reason. He does think that agents are committed to responding appropriately to the recognition of norms. So let us try to understand his attempt to build the authority of practical judgements into the authority of theoretical judgements, and to explain the latter as given in the nature of reasoning itself, in terms of the practice of reasoning itself. He explains the parallel between practical and theoretical reason in the following way:

Just as, if we have a reason for φ-ing we can say that φ-ing is desirable, where desirable is fixed by norms of rationality, if we have (most) reason to believe p we can say that p is (most) credible, where credibility too is fixed by norms of rationality. But now note that just as it is a platitude to say that if φ-ing is desirable then φ-ing is what we would desire if we were fully rational, it is also a platitude to say that if p is (most) credible then p is what we would believe if we were fully rational.....Suppose an agent believes that she would believe that p if she were fully rational and yet fails to believe that p. Is she irrational? She certainly is. And by her own lights! For she fails to believe something she believes she has (most) reason to believe. Indeed this must surely be a paradigmatic case of irrationality.  

Smith is right that the above-described person is behaving irrationally because she doesn't do what she believes reason demands of her. But he mistakes this for proof of the reason's authority. The person who doesn't believe what she believes she would believe were she fully rational assigns authority to the latter belief. But its authority is not the product of her assigning

---

authority to it. The example shows only that one can be irrational by failing to believe or do what one takes oneself to have reason to believe or do. In order to identify this failure, the concept of ‘having a reason to believe x’ must already be in place and so cannot explain it. The difficult case is not the one in which one fails to do what takes oneself to have reason to do, but the case in which one fails to do what there is reason to do. It is difficult because it requires that the reason’s authority be explained, not presupposed.

Rational standards have authority over agents because agents are believers, and as such, have capacities to responsibly form and modify their belief sets. According to Smith, rational norms are evident in the practice of reasoning, particularly in the assumptions we make about the capacities of our fellow believers: “the norms that govern belief enter directly into the definition of the concept of belief itself by defining its proper function, and so enter directly into the definition of the concept of a believer as well.”

When we engage in reasoning (intra or interpersonally) we invoke a set of assumptions about the capacities and responsibilities of the believer. We assume of believers that they are capable of recognizing the norms that govern their beliefs, and that they are capable of appropriately responding to the recognition of such norms. According to Smith, we assume they are capable of responding appropriately, not that they are

---


34 This raises the question of what happens when people don’t respond as they ought to. Presumably, one can lose membership in the class of believers by failing to recognize and respond appropriately to fundamental norms of rational exchange. If, for example, we could no longer assume of a person that she would respond appropriately to the law of non-contradiction, her status as a believer in our community would be altered. Smith doesn’t go into these sorts of issues in any detail. In chapter five, I offer an account of the idea that rational commitments are constitutive of the practice of reasoning.
caused to respond appropriately. To assume that someone is capable of responding appropriately is just to hold them responsible in the relevant ways. Again the ambiguity in his position between the appeal to norm or fact is apparent. Are these practical assumptions about the normative capacities of agents merely descriptively true of what we do when we act rationally? If so, we are back to crude instrumentalism. Although Smith writes as though these normative assumptions are simply predictions about how another agent will behave, he must, nonetheless, employ them in his account as normative demands in order to explain their authority as rational standards.

Smith appears to believe that he has squeezed the normativity of reasons into the agent’s dispositions to respond. He claims that the question ‘what if I don’t care about being rational?’ just doesn’t arise if we adopt the advice model of the fully rational self. In trying to determine what one has a rational justification for doing, the agent will just see the superiority of the fully rational self’s advice. By employing the advice-giving model, Smith supposes that he avoids a re-issue of the normative question. The gap between what I now desire and what my fully rational self would desire is closed because the advice of one’s fully rational self represents an authoritative stance. We are likely to value the advice of our fully rational selves over, say, the advice of our thirteen-year-old selves. Judgements made under optimal conditions, free of fear, intense anxiety or suffering have a priority which others lack. So, suggests Smith, by attending to what my future or idealized self would desire me to now do, I am committed to valuing its

---


pronouncements: "I cannot rationally ignore his desires either." But this response won’t do.

It is true that by invoking the advice-giving model I am considering the desires of my fully rational self, however, it doesn’t follow that those desires have rational priority over the desires I currently experience. Or, at the very least, their priority deserves some explanation. Smith has confused the question of how we engage in rational reflection with the question of the normative authority of rational reflection. He presents the assumptions involved in rational interchange as though they were straightforwardly facts about what we do given the sort of creatures we are, but if they are understood in this way, he cannot make sense of the normative stance that is part of rational interchange. Smith’s description of the fully rational self’s dispositions to respond is supposed to show that reason’s authority is built into the practice of reasoning. We, less than fully rational agents, cannot ignore the advice of our idealized counterparts on pain of irrationality. The force of this ‘cannot ignore’ cannot be merely descriptive of our actual tendencies because as a matter of fact we often ignore such advice. So it must be a normative demand which is not contingent on our actual motives.

The problem for Smith is clear, to the extent that he is an instrumentalist he has no room left to manoeuvre. Instrumentalism precludes an appeal to non-contingent rational ends. Smith’s Humean allegiances demand that he cannot claim that reason directly constrains action by way of categorical imperatives. Note that he has been careful not to claim that ‘one should act in


38 Actually, Smith would deny this although it doesn’t save his position. In fact, he argues for a form of categorical imperatives: “our judgements about what we are morally required to do are simply judgements about what the categorical requirements of rationality or reason demand of us.” They are categorical in the sense that Smith takes it to be a conceptual truth that we expect rational agents to do what they are required to do, where what we expect is a fact
manner consistent with what one would desire were one, in the same circumstances, a fully rational agent.’ To do so would be to invoke a categorical claim about what reason demands of practical reasoners. Instead he argued that ‘if one were to act rationally, one would act in a manner consistent with what one would desire were one, in the same circumstances, a fully rational agent’. But then it remains open whether one must be committed to rational action.

I have pushed Smith into a choice between two options. If he claims that the authority of reason is merely a tendency to behave in a certain way he is left with a description, but not a justification, of rational action. In that case, gross irrationality is merely another tendency to behave. The wildly irrational agent is perhaps an anomaly but does not act unjustifiably. To the question ‘why is the idealized fully rational agent’s advice authoritative?’ Smith can only say, ‘because reasoners tend to act in this way.’ This option leaves him true to instrumentalism but unable to provide the normative account he sets out to provide.

about us, and doesn’t describe a standard to which we hold each other accountable. Because his version of categorical imperatives is explicitly non-normative, the normative problem remains in place. The Moral Problem, 91.

39 This, of course, opens another can of worms. Results of psychological studies question the claim that typical human tendencies in reasoning are adequate expressions of right reasoning. It is quite possible that our actual reasoning is less reliable and rigorous than the standards for reasoning to which we aspire. Stephen Stich offers a concise summary of these results in chapter one of his The Fragmentation of Reason (Cambridge: M.I.T. Press, 1989). However, philosophers like J.L. Cohen dispute the claim that the results of empirical studies demonstrate that we fail to reason rightly. Cohen argues that there could be no way to test for correct rules except by using the rules that we have, so our inferential competence couldn’t be wrong. Stich discusses this line of argument in chapter four.

40 Smith takes his position to avoid problems which the earliest expression of Frankfurt’s higher-order desires model does not. Frankfurt’s higher order desires fail to provide the relevant authority because they are not appropriately constrained by reflection: “Reflection can confer a special status on a desire only if the desire so formed is special relative to that reflective process. But the only way in which desires could be special relative to a reflective process is if, on the basis
On the other hand, if he claims that the authority of reason is a norm or rational standard then, while he would still owe some account of the origin of that norm, he would have provided an answer to the normative question posed. Smith comes closest to getting things right when he argues that as reasoners we are bound to pursue action in concert with rational judgements. But the appeal to our status and responsibilities as reasoners, in the absence of a commitment to the normativity of reasoning more generally, cannot get him what he needs. If the fully rational agent’s advice is authoritative it is so because it is justifiable, not because it defines having a reason. This means that reason can be practical and supply agents with non-hypothetical ends. This option comes at the cost of abandoning instrumentalism, for by his own lights Smith would be forced to deny that reason itself can directly constrain judgement about action.

I have taken no stand on whether the first order details of Smith’s account of normative reasons are correct. In deliberation, it may be useful to think of what a hypothetical fully rational advisor would propose. What I have argued is that Smith’s commitment to instrumentalism and its claim that practical reason is wholly instrumental is inconsistent with his own account. Bratman’s instrumentalism takes more seriously than does Smith’s the turn to norms bound up in the practice of reasoning. As we shall see, the move to practice is promising, but at odds with an instrumental conception of reasoning.

of such reflection, agents would all converge on the very same desires.” The dispositions to which Frankfurt appeals in order to authorize action are just more dispositions and as such are not authoritative over the lower-order dispositions. By seeing the need to introduce dispositions constrained by the authority of rational standards, Smith is on his way to holding that the role of reason in rational action goes beyond the hypothetical imperative. Smith, “A Theory of Freedom and Responsibility,” 315.
**Bratman's No-Regret Condition on Rational Agency**

Michael Bratman doesn’t take himself to be providing an answer to 'the moral problem.' Instead, he seeks to explain how our capacities as rational planners illuminate a discussion of practical reason. Indeed, although he identifies himself as an instrumentalist, Bratman gives up on the belief-desire model of intention and intentional action, complaining that it fails adequately to account for future-directed intention and thus cannot explain an important part of intention itself. For this reason, and because he sees planning as a central feature of rational agency, he explicitly rejects the belief-desire model. Specifically, he rejects the idea that intentional action (action for a reason) can be reduced to an agent’s desires, beliefs, and corresponding actions. According to Bratman, the belief-desire model cannot account for the complexities involved in the formation, retention, reconsideration and abandonment of intentions. Bratman develops his planning theory of intention by focussing on our planning practices ("plans are intentions writ large").

Bratman’s commitment to instrumentalism takes a different form than does Smith’s. He is careful to note that his account of rational planning agency may not be complete: “A theory of instrumentally rational planning agency may not exhaust all that is to be said about rational intentions and plans. But it will, if it is successful, characterize important structures of rational planning agency that are, as it is said, neutral with respect to diverging conceptions of the good.” Nonetheless, Bratman describes his position as instrumentalist, explaining that as such

---


he may remain neutral about the question of whether or not practical reasoning by itself mandates specific ends or moral constraints. His expressed intention is to articulate basic structures of rational planning agency which are compatible with divergent conceptions of the good and for this purpose he takes instrumentalism to be adequate.

According to Bratman's admittedly broad definition of instrumental reasoning, instrumental rationality just is what rationality requires of planning agents in the pursuit of their preferences (desires, cares, commitments, and concerns). Practical reasoning cannot tell us what to value, but it can tell us what to do, given what we value. So, Bratman's allegiance to instrumental reason is both straightforwardly declared and carefully mitigated. For our purposes, what matters is that he wants his account of reasons to remain neutral about the good and sees the instrumentalist rebuff of categorical imperatives as a means of remaining neutral.

My concern is not with the suggestion that practical reason should remain neutral about the good, but with the suggestion that instrumentalism alone explains reasons for action without a substantive appeal to the good. It seems wise to avoid reference to normative moral conceptions which might ultimately influence what counts as rational and irrational action. And of course, a neutral position is not one which rules out the possibility that reason could ever have practical bearing on morality - to do so would be to beg the question against any form of

---


44 Although I am not framing my argument against Bratman in terms of whether or not he employs a conception of the good, it could be put in those terms. For example, Jean Hampton argues that any conception of instrumental reason must involve at least some components of a conception of the good. Quarrels about different views about the nature of instrumental reason are really quarrels about the nature of the good. See, The Authority of Reason, especially chapter five.
rationalism. Bratman's commitment to neutrality with respect to the good is methodological; it is a commitment to a starting point and not a commitment to the idea that reasoning cannot help us to live well.

What Bratman finds unacceptable are reasons for action which involve more than hypothetical imperatives. What we have reason to do cannot be determined directly by reason, because what we have reason to do is always a function of how we are contingently disposed (those desires, ambitions, principles and concerns we hold dear). Reasons get their grip on us (command us) only indirectly, through our adopted ends. Bratman denies that we can be bound to act directly by the imperatives of reason. However, in what follows, I argue that his account presupposes that reason directly constrains action. The rational imperatives he promotes as essential to intention formation are themselves categorical in form.

Bratman describes his planning theory as two-tier and pragmatic. As planning agents we are constantly faced with options to evaluate, and the question of how to proceed in the face of these evaluations. Our plans command both our current selves and future selves. Bratman's account is pragmatic because he grants that as temporally and causally located agents it would not be practically feasible always to later reconsider the overall benefit of what I now intend to do. If this morning I form the intention to enjoy a walk in the sunshine this afternoon, then barring some further reason why I ought to reconsider my intention when it comes time to take a walk, it would be impractical for me to reflect on my afternoon preference rankings and reconsider whether or not they will best be maximized by now taking a walk. While in some cases it will be entirely appropriate to reconsider my plan, in most mundane cases it simply would not benefit me and may disadvantage me to reconsider my intentions on the basis of more up to
the moment preference rankings. So, says Bratman, we need some means of explaining the stability of intentions over time, and accounting for this stability is what makes his position a two-tier theory.

Assessing plan stability requires that we consider when it is and when it is not rational for the agent to reconsider a prior intention. Bratman notes that in most cases whether or not we reconsider an intention is not something we deliberate about - we either do it or we don’t. But in some cases our circumstances will be relevantly modified in unanticipated ways, and so reconsideration of the prior intention becomes important. If, for example, an unexpected thunderstorm blows up, I have good reason to reconsider my previously made plan to go for a walk. Bratman’s strategy is two-tier because it requires first that we consider what sorts of habits of reconsideration and non-reconsideration are reasonable, and then assess whether in a particular case reconsideration or non-reconsideration accords with reasonable habits. Hence, the stability of intentions over time can be understood as a sort of indirect strategy for maximizing rational action:

It is rational of the agent, in that particular case, to reconsider (or not reconsider) just in case she thereby manifests reasonable habits of (non)reconsideration. This two-tier principle is not, of course, intended for use by the agent in deliberating about whether to reconsider… This principle is for the external assessment of the agent, not a principle to be used internally, in the agent’s deliberations about what to do. It is a principle of practical rationality whose role is not directly to guide the action of the agent, but rather to be applied from an external perspective in

45 Bratman is particularly concerned with these “nonreflective” cases of reconsideration or non-reconsideration which he attributes to “the upshot of relevant general habits and propensities.” Intentions, Plans, and Practical Reason, 64.
assessing the rationality of the agent.\textsuperscript{46}

What is needed is a principle linking an account of the stability of intentions to the actual practice of reasoning. Agents who reconsider prior intentions to the point of foolishness are less than admirable rational agents, just as are those who fail to reconsider prior intentions when it would be wise for them to do so. Bratman supplies this connection with, what he calls, the linking principle: “rationally to decide now, on the basis of present deliberation, to A later in circumstances in which I expect to have rational control of my action, I must not suppose that when the time and circumstance for A arrive I will, if rational, abandon that intention in favour of an intention to perform some alternative to A.” \textsuperscript{47} The linking principle connects the rationality of a prior intention with the rationality and execution of that intention at a later time. It represents a double commitment to a reason for action: a commitment both to form and to retain it. The linking principle makes the reasonable demand that in forming an intention, in coming to have a reason for action, my commitment to that reason extends appropriately into the future, so that if circumstances remain the same, the authority of the reason remains stable. A reasonable level of stability will fall somewhere between an overly rigid adherence to prior intentions and the tendency to reconsider prior intentions that wouldn’t, in the long run, be beneficial to the agent. A well-discussed example shows how Bratman’s happy medium runs into trouble.

\textsuperscript{46} Bratman, “Toxin, Temptation, and the Stability of Intention,” 68.

\textsuperscript{47} Michael Bratman, “Planning and Temptation,” \textit{Faces of Intention} (Cambridge: Cambridge University Press, 1999) 52.
Gregory Kavka’s toxin puzzle⁴⁸ shows that there may be cases in which the circumstances surrounding an intention are predictable and so, according to the linking principle, the agent should remain committed to her prior intention. However, in special cases, maintaining the commitment to a prior intention seems irrational. The example goes as follows. A rather curious billionaire has the technological means to detect the intentions of others. This billionaire offers to give me a large quantity of money on Tuesday if on Monday I form the intention to drink a non-lethal but terribly foul toxin on Wednesday. The deal is set up so that I need not actually drink the toxin on Wednesday. As long as I have formed the intention on Monday, I will be able to bank the money by Tuesday. This leaves me with no conclusive reason to drink the toxin on Wednesday.⁴⁹ Instrumental reason’s most compelling intuition is that on Wednesday, given my preference set, I will have no reason to drink the toxin. The problem for Bratman is that the linking principle makes it impossible for me to formulate the intention to drink the toxin on Wednesday precisely because I can see that when it comes time to drink the toxin on Wednesday I will no longer have reason to do so. According to Bratman, I can’t rationally form the intention ‘drink toxin on Wednesday’ because forming that intention includes the further commitment to maintaining it under the circumstances of the example. I cannot rationally maintain the intention to drink the toxin because the reason for drinking it has disappeared.

As Bratman sees it, the linking principle is a necessary feature of a theory of instrumental

⁴⁸ Gregory Kavka, “The Toxin Case,” *Analysis*, 43, (1993), 33-36. In its general form, the toxin case is essentially the problem facing Hobbes’s fool, whether it is rational to keep commitments when they are no longer advantageous.

⁴⁹ And of course, if the billionaire deceives me and doesn’t pay up on Tuesday then I still have no reason to drink the toxin on Wednesday.
rationality because any theory of instrumental rationality owes an account of those conditions in which an agent should reconsider and abandon a prior intention. The linking principle constitutes a constraint on rational deliberation-based intention formation, a constraint which we have now seen is in tension with the thought that instrumental reasoning maximizes preferences. The linking principle and instrumental rationality combine in such a way that an instrumentally rational agent cannot form the intention to drink the toxin on Wednesday, and thus cannot gain the benefit of formulating that intention. So the instrumentally rational agent is at a disadvantage in such a case, for rationality requires intention stability over time:

we seek general habits and strategies of reconsideration that are, in the long run, effective in the pursuit of what we (rationally) desire. In a particular case we reasonably implement such pragmatically grounded general habits and strategies and, depending on the case, reconsider or refrain from reconsideration. This means that a planning agent may sometimes rationally follow through with a prior plan even though she would have rationally abandoned that plan if she had reconsidered it in light of relevant unanticipated information.

Although it appears to be in genuine tension with the demands of instrumental reason we ought not to abandon the linking principle. The constraints imposed by the linking principle on

50 Bratman considers a couple of different methods of resolving this conflict, methods which rely upon the idea that the instrumentally rational agent should invest in generalizable procedures conducive to her life going as well as possible. He rejects them on the grounds that they do not adequately respect the fact that when it is time to execute the plan the agent is not rationally bound by them. It remains true that at the time of action, her preferences will not be maximized. Here he finds too much significance in the overriding force of the current preference set, and fails to heed his own remarks about the need to make sense of the stability of intentions over time.

practical reasoning parallel rational constraints in the case of theoretical reasoning, which are equally compelling. Suppose that after deliberating I rationally conclude that \( p \). My \( p \) belief is defeasible; without evidence to the contrary, or a change in circumstances leading me to reconsider my commitment to the belief, said belief will persist, but not only persist. My commitment to that belief at \( t_1 \) extends into the future. At \( t_1 \), I endorse \( p \), not \( p \) at \( t_1 \). Of course, some beliefs may be temporally fixed: \( I \) believe that I am now typing\( \) contains a temporal indexical and so makes reference to the time at which the claim is made. But my commitment to even indexically expressed beliefs extends into the future. At \( t_2 \) I would be committed to the claim that \( 'at \ t_1 I \) believe that I am now typing'\( \) was true. In other words, I would need some further reason to rationally change my belief, to withdraw my commitment to it.

Bratman is trying to define the extent and form of our commitments as rational agents and is right to hold that something like the linking principle functions as a rational constraint on intention formation. It describes a way in which we may be held rationally accountable for our reasons and actions. However, it would be a mistake to hold that it is a condition of having an intention at all, that it conform to the linking principle. The problem the toxin puzzle poses for Bratman is not a puzzle about human psychology. At issue is not whether it is psychologically possible to form the intention to drink the toxin given the authority of the linking principle. The real issue is whether it is reasonable to do so. An appeal to psychological facts which rationally bind us should set off alarm bells.

Bratman tries to resolve the tensions made apparent by the toxin puzzle by introducing what he calls, the no-regret condition. The no-regret condition is motivated by the thought that as planning agents we will be concerned with our reasons for action not just at the moment at
which we adopt them, but also after we have acted, or failed to act, on them. Cases of temptation provide good examples. Bratman gives the example of Ann who enjoys good beer at dinner and a good read after dinner. Ann knows that more than one beer at dinner diminishes her enjoyment of a good book. Her problem is that having consumed one good beer at dinner she will want another, a preference which competes with her earlier preference for enjoying a good book after dinner. Ann also knows that her desire for a second beer is more temporary than the first. Having just finished her first beer she will have a higher preference for another beer over no more beer, but after dinner she will be glad to have refrained from indulging in the second beer. She will regret having the second beer if she does indulge. According to the linking principle if, prior to dinner, Ann forms the rational intention to have only one beer with dinner, knowing that she will desire a second during dinner, she is committed to the rationality of refraining from a second beer at dinner even though the directive to maximize preferences at that time would instruct her to have the second beer.

Bratman introduces the no-regret condition in order to offer a rational resolution to cases of conflicting dispositions such as this. He puts the no-regret condition like this:

(a) If you stick with your prior intention, you will be glad you did.
(b) If you do not stick with your prior intention, you will wish you had.
So, other things being equal,
(c) Though you now prefer to abandon your prior intention, you should nevertheless stick with it.

...Let us say that when (a) and (b), suitably interpreted, are true, following through with one's prior intention satisfies the no-regret condition.52

---

We aspire to be reasoners who don't regret our actions and to develop habits for the reconsideration of prior intentions which serve us well. Ann cares about her intention to drink just one beer not only when she forms it. Her rational commitments, as a result, extend into the future. We tend to think about how we will react to our intentions and our continued commitment to those intentions when we are making decisions. Ann's deliberations may include an awareness of how likely she is to be assailed by a desire for a second beer and perhaps even, on a bad day, the thought that she is likely to cave in to such a desire. The quality of her reasoning, suggests Bratman, depends upon such forms of awareness. If we were the sorts of creatures whose consciousness did not include a connection to a future sense of self-evaluation we would likely be less well off, rationally speaking, than we are.

The rational constraints which Bratman introduces are instructive. But are they compatible with the theoretical commitments of an instrumental account of practical reason? The linking principle and no-regret condition are substantive normative constraints on rational deliberation. They express rational norms binding rational agents. Bratman takes these rational constraints to be consistent with instrumentalism, perhaps, because although they provide us with non-contingent ends, as ends they serve as standards of rationality not standards of conduct. So he might defend these normative constraints as consistent with instrumentalism by pointing out that they are norms of rationality, not morality, and as such do not constitute a conception of the good. Yet, as his examples make clear, these standards of rationality are crucially standards of conduct after all. They are standards dictating one course of action over another. They promote, for example, thoughtfulness about anticipated future regret over inattention to (or apathy about) future regret. Bratman has cloaked these norms in terms of our dispositions to care about future
regret — his version of the mongrel concept. The bare fact that we are disposed to care about future regret cannot justify its normative significance — it cannot serve as a rational imperative. He needs to explain how, when an agent is tempted, the fact that she may care about anticipated future regret serves as a rational constraint on what she does.

Bratman explains the authority of the no-regret condition by noting that “the agent comes to the temptation with relevant prior intentions, and so there is already an issue about whether she may rationally simply give them up....so long as she has these intentions she is a planning agent in a manner that makes salient relevant, anticipated future regret.”53 This seems an apt description of a familiar situation. However, the relevant question isn’t ‘what does it feel like to be tempted?’ It is ‘why should I take the standards of rationality (such as the no-regret condition) to be authoritative constraints on what I do?’ Bratman hasn’t really answered the challenge.

Bratman supposes that our disposition to care about future regret is itself rationally authoritative. According to Bratman, the force of the no-regret condition “is not grounded simply in the recognition that one is a “temporally extended being.” The force of the no-regret condition is grounded, further, in one’s actual engagement in relevant planning agency, and in the resulting significance to one of how one will see matters specifically at plan’s end.”54 What is relevant is that the agent cares about her future evaluation of her present decision. However, if understood as a simple description, concern about one’s future evaluation of a prior judgement has no normative force. Bratman’s no-regret condition presupposes the whole idea of having a reason and thus cannot explain it. Regret and anticipated regret per se do not give one a reason to act.


Rather, regret and anticipated regret serve (at best) as indicators of reasons which are already present. If regret were just an unpleasant feeling then it would have no normative significance. As Sellars's myth of Jones suggests, it is only by having mastered public skills of taking responsibility and finding lessons for the future in our present experience that we acquire the ability to find in our spontaneous reactions reliable indicators of reasons. It is because we assign significance to regret and anticipated regret that they can have normative force.

Dispositions to regret, like other emotional responses, can function well or badly. In a normal case one's disposition to regret functions as a reliable indicator of reasons relevant to deliberation. For example, one might drink a cup of coffee in order to avoid a caffeine-withdrawal headache. One wants to avoid the unpleasantness of the headache and so has a reason to avoid pain. Avoiding regret is a different matter. One tries to avoid regret not because it is painful, but because it is an indication that one has made a mistake and this makes it a useful epistemic tool. What ought to be avoided is acting in ways which will leave us with reasons to regret, not just the unpleasantness of regret itself. Dispositions to regret, like dispositions to identify blue ties, must be endorsed to warrant an intention. The move from the causal to the justificatory order of explanation is made by assigning normative authority to one's disposition to regret. Just as in the case of John in the necktie shop, the possibility of systematic error demonstrates that the normative force of the no-regret condition is a function of the propriety of the endorsement of the disposition and not the disposition itself. If one's dispositions to regret are functioning badly they will not serve well as indicators of reasons. For example, when deciding whether or not to go out Friday night I may anticipate regret no matter what I do. If I go out I can anticipate that Saturday morning I will regret my less than studious activities of the
night before. If I stay home I can anticipate that Saturday morning I will regret having put study before the good of camaraderie. This shows that in this case my disposition to regret is fickle and uninformative. It cannot serve to recommend one option over another. Of course, Bratman can accept that our tendencies to regret can be misguided and uninformative. He cannot, however, appeal to anticipated regret in order to explain the concept of having a reason without presupposing it.

The point against Bratman could also be made by showing that his account of practical reason depends upon a minimal conception of what constitutes our good after all. The person who systematically fails to care about anticipated future regret mistakes the good. A person who takes a regret-suppressing pill, for example, is not made more rational by virtue of the fact that she never violates the no-regret condition. She is criticisable on more than instrumentalist grounds. She is criticisable because she fails to see the good in attending to reasons which may become apparent to her through reflection about what she is likely to regret. The instrumentalist, wedded to the claim that criticising someone as instrumentally irrational doesn't involve a criticism of what they take to be good, cannot deal with this sort of rational mistake.

Bratman's no regret condition cannot explain how it provides us with a reason to act without appealing to it as a reason for action, and hence presupposing its normative force. Bratman's account, like Smith's, either slides into crude instrumentalism or presupposes a conception of rationality too rich to be derived from instrumentalist premises. On the one hand, if attention to regret and anticipated regret are understood merely as descriptions of how we tend

---

55 I do not wish to foreclose on the possibility that there may be circumstances where one faces genuinely incommensurable goods with the result that one will have something to regret no matter what one does, but only to point out that dispositions to regret can be defective.
to behave, they have no normative force. In this case, agents who don't consider their future regrets may be less well off in the long run, but they make no rational mistakes. If, on the other hand, the no-regret condition (and linking principle) are rational norms, while the question of where they obtain their authority remains, we must understand them as demands which hold authority over practical reasoners. On this view agents who ignore the no-regret condition are making a mistake - a rational mistake. This second view is surely the most plausible and the one which Bratman intends to promote. The problem for his account is not with the ambitious rational standards it incorporates, but with the claim that such rational imperatives are not categorical.

In this chapter I have attempted to show that instrumental accounts of practical reason fail to live up to their promise to explain reasons for action without going beyond the idea that the imperatives of reason are only hypothetical. I have argued that to the extent that they remain true to the instrumentalist principle that practical reason delivers only hypothetical imperatives, such accounts fail to explain how practical reasons rationally bind reasoners. Reasons (both practical and theoretical) do not bind us because they constitute apt empirical descriptions of our psychological make-up or tendencies to reason well. Even sophisticated instrumentalists are guilty of employing mongrel concepts when they suppose that reason's authority can be completely explained in purely causal terms. Practical reasoning is not reducible to theoretical reasoning plus desire; it is normatively practical in its own right. In the next chapter we will examine Christine Korsgaard's explanation of the normative authority of practical reasons. Korsgaard recognizes that the authority of practical reasons must be, in crucial ways, a feature of the practice of reasoning itself.
Chapter Four:
Reflective Endorsement and the Myth of the Given

In the last chapter we saw that even sophisticated instrumental accounts of practical reason conflate justificatory and descriptive orders of explanation when they propose that the normativity of practical reasons is reducible to a set of non-normative facts, and thereby fall foul of the myth of the given. Before moving on to examine Christine Korsgaard's even more sophisticated attempt to avoid the myth of the given, let us briefly review the argument thus far.

According to the instrumentalist, an agent has a reason if she is in a state in which her dispositions to act are consistent with a particular set of proposed rational constraints. But, as we saw, a further proviso is needed in order to distinguish between acting for a reason and behaviour which is explainable by reasons. Agents do not simply pass unaware into and out of 'reason-providing states' which are inexplicably capable of rationally (as opposed to causally) influencing action. It must be in principle possible for the agent to become aware of her reasons for action. There is no interesting sense in which an agent can be said to have a reason to act if it doesn’t make sense to say of that agent that had she been rational she would have acted on her reason. A purely causal account fails to explain practical reasoning.

The parallel in the case of theoretical reasoning demonstrates the significance of this
failure. On a purely descriptive basis, if an agent believes that A and believes that B then she is
in a state which, according to theoretical reasoning, could be explained by her believing that A
& B. But this is not to say that she actually does believe that A & B, nor that she has a reason to
believe that A & B, whether she happens to or not. Because on this view reasons for belief
amount to descriptions of a state of the agent, this view fails to explain how it is that such an
agent fails rationally if she fails to believe that A & B. So, in addition to a description of the
conditions under which an agent has reason to act, the instrumentalist must introduce a condition
which can explain how acting for a reason involves commitments to rational standards. This led
me to introduce the idea of a commitment as a necessary feature of reasons.

Commitment involves awareness of one’s inclinations to act in a particular way, but goes
further than simple awareness. One can be aware that one has a strong desire or propensity to
ϕ without taking oneself to have reason to ϕ, in much the same way that one might have a strong
propensity to identify certain shades of neckties as blue. Just as differential responsive
dispositions (propensities to believe) are not themselves sufficient conditions for having a reason
for belief, similarly complex responsive dispositions (propensities to act) are not sufficient
conditions for having a reason for action. In the case of belief, the agent’s commitment to a truth
claim is based on, but not reducible to, her differential responsive disposition. Her responsive
disposition supplies a reaction to her environment, and thus serves as a source of data for herself
and others. The responsive disposition is thus crucial in belief formation because it signals a
connection to the world and beliefs are, after all, about the world. But an agent’s connection to
the world is not yet a belief, much less an incorrigible belief. Simple awareness of one’s
dispositions to identify green neckties, or to have a second beer with dinner, aren’t as such
reasons for asserting that a particular necktie is green, or for proceeding to indulge in a second beer.

In practical reasoning normative attitudes of taking or treating dispositions as reason-providing are a necessary condition for practical judgements. Commitment provides the connection between the agent and her reason for action by making it in principle possible for her to acknowledge that she is in a state which reason condones. The challenge I posed for the instrumentalist was to clarify this criterion without conflating causal and justificatory orders of explanation, without arguing that the fact that an agent can be described as being in a particular, reason-giving state rationally binds her to act in accordance with that reason. In the last chapter I argued that in order to avoid an appeal to the myth of the given, the instrumentalist’s explanation of the normative authority of reasons for action must explain commitment without presupposing a mysterious realm of functionally normative facts. I concluded that the instrumentalist is in a precarious position. Adherence to the precept that reason is inert leaves the instrumentalist unable to hold that reasoning itself provides the normative force of reasons for action.

In this chapter I turn to an examination of Christine Korsgaard’s account of the normativity of practical reasons. Korsgaard’s account is complex and attentive to many of the Sellarsian points which motivated my criticisms of instrumentalist accounts of practical reason. Her position can be seen as one which attempts to avoid the myth of the given by making explicitly non-instrumental claims about the nature of practical reason. Very roughly, she argues that the normative authority of practical reasons is a function of our engagement in practical reasoning itself, which is in turn grounded in our capacities and nature as reflective, autonomous
agents. Korsgaard’s position seems promising, at least *prima facie*, as one which respects the lessons of Sellars’s critique of the myth of the given precisely because she begins by asking how it is that episodes of reasoning have the power to command or oblige. This seems promising because it is likely to rule out accounts which conflate descriptive and justificatory aspects of practical reasons. The mongrel concept employed by the instrumentalist to both explain and justify the normative force of practical reason is one which Korsgaard rejects. She argues, for example, that reasons must be public and that reasons cannot be understood as non-normative facts about agents. In this chapter I hope to demonstrate that while Korsgaard’s position is an advance on the instrumentalist’s it is unclear that it succeeds in altogether avoiding the myth of the given. The myth of the given has more than one form and I will argue that ambiguities in Korsgaard’s account of commitment show that she does not avoid it. Before turning to the particularities of Korsgaard’s position let us rehearse the basic form of Sellars’s argument against the myth of the given. This should leave us in a good position to evaluate Korsgaard’s position.

Although Sellars’s argument in *Empiricism and the Philosophy of Mind* concerned theoretical reasoning, it has direct parallels in the practical case. The sense data which Sellars dubs *mongrel concepts* are so dubbed because of their functional role in the justification of knowledge. While understood in purely descriptive terms, sense data are employed by the empiricist to do a normative job, to provide a foundation for knowledge claims. Sellars demonstrates that an explanation of knowledge claims requires reference to a normative or epistemic relation between agent and world. Facts are known, sense data are merely experienced. The classical empiricist wants to ground knowledge in a form of causality and is guilty of an appeal to the myth of the given because in doing so he conflates the causal and justificatory
aspects of belief formation.

Sellars's argument makes it clear that justification is a normative, not a causal notion, one which requires reference to the epistemic attitudes involved in knowledge claims. What is sensed is a particular, what is known is a generality. The having of sensations is the sensing of particulars. The mere having of an experiential state accordingly cannot be knowledge. In order to serve as a basis for knowledge, sensed particulars would have to function as generals or rules applying to, and justifying, new cases. The empiricist attempts to explain an essentially epistemic (or inferential) relation with a causal one by proposing that sense data provide a special sort of self-justified awareness. However, only states of consciousness can be individuated by the awareness of sense data, beliefs cannot. Beliefs, or knowledge claims, are not explanatorily exhausted by the sensual input which give rise to them, a fact made clear by the possibility of systematic error, as illustrated by the example of John in the necktie shop. Tendencies or dispositions to respond to one's environment are a necessary but not a sufficient condition for knowledge. They are not a sufficient condition for knowledge or belief because until they are taken to serve as reliable indicators of what the world is like in circumstances productive of reasons for belief, they are just occurrences and make no claims about what the world is like. An epistemic attitude toward the contents of one's sensual experiences, among other things, is necessary to an account of belief formation, for it is only once an attitude toward one's experience is taken that a claim can be made. In short, a propositional attitude is not a purely descriptive state of a person but a claim or commitment about the way the world is.

Sellars's term for the epistemic performance involved in belief formation is endorsement. Endorsement describes the commitment involved in knowledge claims, commitment is the parallel
I am developing in the practical case. Endorsement involves a manifold of commitments. First, it includes a commitment to the thought that the agent to whom the claim can be attributed is one who has the relevant capacity to reliably identify features of his environment. Second, it includes a commitment to the thought that the agent’s reliable responsive disposition has application in the particular circumstances - that it is properly functioning in this case. So it includes a commitment not just to the thought that the assertion is true, but also to the thought that the grounds of the judgement are normatively adequate. To say that it is normatively adequate is to say that the endorsement is non-arbitrary and grounded in a commitment to a set of conditions which qualify as accurate or acceptable judgement-producing conditions. The endorsement involved in knowledge claims is thus grounded in the agent’s commitment to the propriety of her own judgement. So a commitment to standard conditions is inescapable.

The example of John in the necktie shop illustrated the fact that an agent’s differential responsive disposition is not by itself sufficient for knowledge or belief. The claim that a necktie is green, for example, is an endorsement not merely of one’s responsive disposition, but also of the proper functioning of one’s responsive disposition within a particular environment. Other agents with reliable differential responsive dispositions would, in such conditions, produce the equivalent sorts of judgements, and agents not in a position to make this claim are authorized to take John’s judgements as authoritative. John learned that his disposition to identify the colours of neckties was not accurate after electric lighting was installed and so refrained from endorsing his then questionable necktie colour identification disposition. His claim, ‘it looks green,’ was an acknowledgment that he remained uncertain about the appropriate scope of his disposition to identify the colour of neckties, and was thus a restriction on what he could endorse as true.
Through endorsement authority is conferred upon some responsive dispositions and not others, so endorsement explains how knowledge claims involve epistemic, not merely causal, relations.

It is clear, however, that endorsement isn’t simply self-justifying. The authority involved in knowledge claims presupposes a form of endorsement which goes beyond self-authorization. After all, the aim of belief is truth. Sellars’s way of putting this point is to say that ‘looks’ talk presupposes ‘is’ talk, which is to say that ‘is’ talk is conceptually prior to ‘looks’ talk. Sellars’s discussion of looks/is is intended to explode the idea that we come to know the way things are by investigating the ways things look. Assertions about how things look cannot provide a foundation for knowledge because claims about how things look cannot themselves be understood unless claims about how things really are, are already in place. Sellars notes that in ‘x is red = x would look red to standard observers under standard conditions’ the left hand side of the equation has the same truth conditions as the right hand side of the equation but is not defined by it. ‘Is red’ cannot be defined by ‘looks red’ because the equivalence is only true because we have mastered the normative standards involved in using ‘red’ correctly. That is, it requires conceptual mastery of ‘red’ and thus cannot be said to define it. A person has mastered the concept ‘red’ when red things look as they are to her under standard conditions, and is thereby entitled to use the word to report the presence of red things. It is a necessary truth because standard conditions are the conditions under which red things look red to standard observers. Looks claims are grounded in is claims because mastery of a normative concept is first needed in order to be able to make the associated looks claims. Necessary truths like the equivalence just mentioned reflect inferential moves implied by a mastery of concepts.

The space of reasons is the space in which the inferential moves constitutive of knowledge
claims are made. The practice of giving and asking for reasons provides the public context in which an assertion may be endorsed. Importantly, for Sellars, endorsement requires the capacity to use language in public contexts. As we have seen, endorsement is not a self-justifying enterprise. An agent’s commitment to a propositional content per se does not authorize that belief in others. Neither can beliefs be made true merely by being willed to be so. Reasons do not come into being simply because we have intentions. If the endorsement involved in making knowledge claims were merely a private, psychic act then the state in question would carry no commitments about the way the world is and thus fail to be an epistemic attitude. Beliefs could not function interpersonally as reasons because they wouldn’t involve shared commitments. By conferring authority on propositional contents I may authorize beliefs in you. The game of giving and asking for reasons makes it possible for reasons to be exchanged and this requires that the entitlements involved in having reasons be conceptually public. Reasons demand the public mastery of concepts because reasons, and concepts themselves, are based on inferential relations which are in principle shareable. Because these inferential relations reflect correct mastery of concepts and their empirical application they delimit the permissibility of inferences preserving knowledge. Knowledge is thus conceptually public. In making an assertion, the agent takes responsibility for the inferential role of his assertion. Based upon his assertion further judgements may be legitimately drawn, and his endorsement makes him responsible for all of these. Endorsing is thus the practice of taking responsibility for the set of inferences which may be legitimately inferred from the utterance itself. Let us now turn to Korsgaard’s account of commitment in order to see how fully it captures these features of endorsement.
The Normative Question

Why do what I have reason to do? This question expresses a thought had by most of us at one time or another, a thought most likely to arise when the prospect of living up to one's reasons seems onerous or overwhelming. Having judged that I ought to do the selfless act and forgo some pleasure, I may be led to question the legislative force of my own judgements. Of course, I may not be so led. Most often, we are not led to question our own judgements in this way. Having judged that I ought to give my change to the beggar instead of stopping at the bakery for a treat, I am likely to do just that. I don't pause, coins in hand, to consider just how it is that by forming an intention I am being done out of a tasty baked good. But I might. It may be that one is more likely to experience this sort of detachment from one's own reasons when one is overtly aware of how compliance with those reasons leads to a personal loss. I think it is a mistake to focus on personal hardship, to the extent that Korsgaard does, when conceiving of the normative question. The authority of one's reasons may seem puzzling even when no threat of significant personal loss emerges from them. Having decided that 'I ought to go to bed early tonight so that I'll enjoy myself tomorrow,' I may yet continue to wander around my apartment, feeling dubious about the legislative jurisdiction of this reason. The experience of detachment from, or skepticism about, one's own reasons should not be thought of as mere reluctance to do something one finds difficult. The normative question, 'why do what I have reason to do?' challenges us to consider how reasons for action rationally bind agents. So an answer to this question will tell us something important about the nature of reasons for action.

By posing the normative question, Korsgaard confronts one form of mongrel concept
head on. She is careful to distinguish between the explanatory and justificatory features of moral theories. They must be able to describe in general terms the actions that morality demands, a task which is expressible in third-person terms, but also explain why we ought to do what we ought to do. This latter task is distinct from the first, a theory may very well describe moral behaviour without explaining how it is justified. The explanatory features of moral theories may provide the means to pick out moral actions without thereby accounting for the obligation to perform moral actions. I might be independently inclined to do the very things morality demands of me, but until I know why I ought to do what morality demands an explanatory moral theory will only give me a description of what moral conduct looks like and not a reason to act. Reasons for action haven’t been fully accounted for unless their normative authority has also been explained.

The normative authority involved in the claim that ‘you ought to give money to Oxfam’ is not explained by adding that it is your duty to do so, or by declaring that such acts are conducive to the greatest good. Such features do not explain the normative interest we take in morality. Remember that at issue is not which substantive principles are correct, but rather what gives any substantive principle which we accept, normative force. An explanation of the normative authority of practical reasons is an account of their capacity to bind or obligate. Why ought I to do what I have reason to do? What justifies the claims that morality makes on me? The normative question gets its force, Korsgaard tells us, from being asked in the first person. An adequate resolution must be able to provide an answer to the person who, confronted with the demands of morality, questions its authority over him. Here is how Korsgaard puts it:
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.¹

Korsgaard goes on to say that the answer to the normative question must meet three conditions. First, it must actually succeed in addressing someone in the first person position -- an agent who doubts whether he ought to do what he believes morality tells him to do. Second, answers to the normative question must be transparent. It must be possible for morality's justification to be made public without that system of morality thereby faltering. Third, the answer to the normative question must appeal to an agent's identity. Morality is peculiar amongst the normative judgements because it is connected with our identities in a unique way. People have been willing to die to defend moral, but not mathematical, truths and this difference is relevant to its justification. Because the demands of morality may, in the most extreme cases, threaten our very lives, and in less extreme cases our self-conceptions, their justification requires special attention to the intimate ways in which obligation and personal identity converge.

I have concerns about the first and third of these conditions, upon which I will later expand. For the time being I will simply note that the requirement that an answer to the normative question be 'successfully addressed' to the first person is ambiguous between success in persuasion and success in mustering a thoroughgoing defence or demonstration. Furthermore, the requirement that an answer to the normative question must appeal to one's personal identity

is ambiguous between the particular conditions of personal identity, and the more general sense in which identity can be understood as a feature of personhood. Both of these forms of ambiguity will be discussed later in this chapter.

**Korsgaard's Answer to the Normative Question**

Korsgaard’s starting point is promising because it demonstrates that she is keenly aware of the gap between the descriptions of morally right action provided by moral theory and accounts of how it is that right action so described bears the power to legitimately command. The promise lies in her clear separation of the descriptive and justificatory features of the explanation of reasons for action. Careful distinctions may preclude the invocation of a mongrel concept employed to simultaneously describe and justify practical reason. Indeed, Korsgaard’s strategy in advancing her own position is to argue that an answer to the normative question naturally leads to an account of justification as fundamentally practical and thus, to her account of *reflective endorsement*. Korsgaard’s arguments for reflective endorsement depend upon her suggestion that moral theories which do not offer an account of reflective endorsement will conflate descriptive and justificatory forms of explanation - they invoke mongrel concepts.

Essentially, the challenge posed by the normative question as Korsgaard raises it, is that of explaining how a reason is made practical. Moral theories offer up reasons for believing that particular acts are permissible, impermissible, or obligatory. So in one sense the project of answering the normative question is the project of explaining how reasons for belief become
reasons for action. How can the truth of a proposition, a fact, carry with it the demands of an obligation, a norm? Korsgaard's answer is that obligations arise because we are reflective creatures with the power to endorse or refrain from endorsing the various ways in which we are disposed to act. This power provides us with the means to give ourselves, and others, reasons for action and thereby incur obligations. There are norms because we are normative creatures, not because there are normative facts which theoretical reasoning can detect.

Korsgaard's separation of descriptive and justificatory orders of explanation provides the basis for arguing, as she does, that the dictates of any moral theory must be made practical in order to explain how moral reasons command. Moral theories don't themselves supply reasons for action, although they do supply principled accounts of the theoretical requirements and consequences of a particular set of beliefs. Reasons for action are practical - they bind the choices of agents who have them - and so any moral theory owes an account of how theoretical reasoning is made practical, how it comes to have the normative authority to require particular actions. Any moral theory needs to explain how to turn reasons for belief into reasons for action. But reasons for action, Korsgaard argues, come about because agents rationally adopt plans for action as their own, so the normative authority of such reasons must be accounted for by explaining how they are made practical, or reflectively endorsed.

Reflective endorsement provides a solution to the normative question because it makes the bindingness of reasons a feature of reasoners and in doing so appears to avoid the mistake made by those who confuse the descriptive and justificatory roles of explanation. Reasons do not

\footnote{I say \textit{in one sense} because although this way of putting the point is consistent with the commonplace assumption that theoretical and practical reasoning are separable by the gulf that exists between fact and norm, I will be challenging this very assumption.}
bind us by causing us to act in accord with them for, of course, we might not act in accord with them. As in the case of theoretical reasoning, where the endorsement of a particular disposition (to believe or to respond to one's environment) featured in the justification of belief, in the practical case, endorsement of a particular disposition to act features in the justification of reasons for action. Sense data are not themselves beliefs, nor are dispositions to act themselves reasons for action. Furthermore, just as a *sensing* of sense data cannot serve as a justification of belief in the theoretical case, a simple *awareness* of one's disposition to act cannot serve as a justification of reasons for action in the practical case. Inclinations to act become reasons to act when, through reflection, we confer authority upon those inclinations. Reasons for action only arise once a disposition to act has been reflectively endorsed.

Korsgaard adopts a Kantian conception of the agent as one who is moved by desires to act but who is not reducible to those desires; agents have wills as well as motives. Sensation may not be a reason to believe, desire may not be a reason to act. Because for humans there is a gap between inclination and a reflective response to that inclination which other animals do not have, our lives possess a normative dimension which the lives of non-human animals lack. Reasons describe a type of reflective success available to us because our dispositions to act and to believe are not the end of the story, they are as Sellars says, entry moves into the space of reasons. For Korsgaard reasons are practical from the ground up. We have reasons for action because we need to act. Because we act under the idea of freedom we require the means to authorize the adoption of one disposition to act over others. Our reasons are fundamentally constructed, they are not, as the realist proposes, discovered.

What needs explanation at this point is the nature of the authority of these reasons for
action and here Korsgaard makes good use of her Kantian roots. Reasons for action are not authoritative merely in virtue of the fact that an agent has condoned one disposition over another. As an explanation of the normative authority of reasons it would be too arbitrary and leave unexplained how reasons obligate over time. If the justification had through reflective endorsement were simply self-justification then reasons would have authority (if indeed they could have authority) only at the moment they were produced. If I decide now that I will later drink only one beer with dinner, thereby endorsing my disposition towards moderate beer consumption, when I discover half way through dinner that I desire a second beer, my previously endorsed disposition toward moderation neither rationally advises nor obligates me to refrain from indulgence at this moment. Of course, I could again commend my moderate disposition and let it serve as a reason to act in this moment, but that is beside the point since I could equally well do the opposite. Obligation across time isn’t the only problem. Reasons must be rationally authoritative -- in the sense that they claim priority over all competitors -- not authoritative in virtue of an arbitrary act of willing. They command us, we do not command them. If the justification involved in reasons for action were just self-authorization, the sense in which reasons for action are requirements of rationality would be lost. Some account of rationality must feature prominently in reflective endorsement and for Korsgaard the rationality involved in practical reasoning is bound up with the categorical imperative.

Because reasons are created, not discovered on Korsgaard’s view, their authority is traceable to the manner in which they are formed. She notes that “since the will is practical reason, it cannot be conceived as acting and choosing for no reason.”

Reflective endorsement

---

3 Korsgaard, The Sources of Normativity, 97-98.
isn't arbitrary self-authorization because endorsements are law-giving acts of willing. The authority of reasons, Korsgaard explains, comes from the categorical imperative as represented by the formula of universal law. Because the will is a form of causality it requires its own form of law and because the will is free its law cannot be forced upon it from the outside. So the law of the will is the principle which tells it to make its own laws. This is the law of the free will, imploring us to act only on maxims which we could at the same time will to be universal laws. The only constraint is that one's maxim have the form of a law. Korsgaard sums it up this way: "the categorical imperative is the law of a free will. It does not impose any external constraint on the free will's activities, but simply arises from the nature of the will. It describes what a free will must do in order to be what it is. It must choose a maxim it can regard as a law."

But if introducing the categorical imperative into the picture resolves the problem of the arbitrary endorsement of ends, it raises the question of the normative authority of reasons all over again. After all, if I am the kind of creature who has a will I am the kind of creature who reflectively endorses inclinations and in doing so arrives at reasons (if I am determined to reflectively endorse my inclinations) then my reasons are a function of my nature. If my reasons are a function of my nature, my constitution as a reflective agent, how can they bind me? In other words, if the categorical imperative is the source of the normative authority of reasons and is just the human tendency to give ourselves laws, in what sense is it an imperative? How does normativity arise out of a human tendency?

---

4 Here Korsgaard parts strict company with Kant by distinguishing between the categorical imperative and the moral law. Her own answer to the normative question leads her to argue that the moral law is implicit in the categorical imperative.

5 Korsgaard, *The Sources of Normativity*, 98.
We appear to be presented with a dilemma. Are reasons normatively binding because we have, as a matter of fact, committed to them? Or, do our rational commitments stem from the authority conferred by the principle of reasoning itself? Neither horn of the dilemma gives us what we need. If reasons are made normative simply by the actions which give rise to them, the source of their normativity is not practical reasoning. The simple fact that a propositional content has been reflectively endorsed, cannot serve as an explanation of that reason's authority lest the normativity of a reason arbitrarily depend upon the movement of one's will. On the other hand, if reasons are normative for us in virtue of their law-like form, then the source of their normativity seems external to them. Reasons bind us because of a formal property which, of necessity, they possess. Korsgaard comes close to resolving this dilemma with a constitutive argument. In offering a constitutive argument for the authority of reasons she correctly moves towards locating the normative authority of reasons in the public practice of reasoning. However, by making their binding force depend upon psychological operations in us, rather than on the intersubjective web of inferential commitments which gives them authority, she makes private (first person) reasons conceptually prior to public (intersubjective) reasons.

Korsgaard’s solution is to argue that these options present a false dichotomy; our reasons are intrinsically normative because the authority of the categorical imperative is operational in the laws we give ourselves. The law-like quality of reasons which grants them authority is not

---

6 This would leave the view in a position similar to that in which Harry Frankfurt’s account of the rational will finds itself. Endorsing, would be roughly equivalent to the act of ‘identifying with one’s first order desires,’ in that it would perform the same regress-stopping function. See his, “Freedom of the Will and the Concept of a Person,” The Journal of Philosophy, 68, (1971), 5-20. Frankfurt’s position was roundly criticized, most notably by Gary Watson in, “Free Agency,” The Journal of Philosophy, 72, (1975), 205-220.
external to them, but is constitutive of reason-giving itself. She clarifies this by asserting that the principles of reason are constitutive of the act of reasoning and so the gap between fact and norm, where fact is a description of the universalizable manner in which we reason, and norm is the capacity for that universalization to offer guidance, doesn’t arise.

I don’t mean to suggest that these principles are descriptive rather than normative; nor do I mean to suggest either that we always follow them automatically, or that we are always conscious of being guided by them, when we engage in these activities. It is the nature of activities, as opposed to mechanical processes, that one who engages in them is self-guided (in an extensive sense, therefore, autonomous). The rules constitute the activities in the sense that what it means to be engaged in them is to guide yourself in accordance with these rules. For an activity to be self-guided, it must be one that you perform consciously, but you need not be conscious that you are doing it, or that you are guiding yourself by these rules.7

The problem of showing how norms fit into the natural world exists at some level for any non-reductionist, naturalist philosophical position. At some point the stories we tell about what we do and what we are will be called upon to provide the means to account for the normative features of ourselves and our world. So the problem of norms isn’t a problem exclusive to Korsgaard’s project. The question we need to ask is thus whether Korsgaard’s account has the resources to explain the normativity of practical reason without invoking the myth of the given.8

7 Korsgaard, The Sources of Normativity, 236. Korsgaard makes this point in response to a rebuttal advanced by G. A. Cohen (174) who argues that she, unlike Kant, cannot claim that agents must be moral on pain of irrationality. Korsgaard doesn’t have the means to avail herself of this Kantian dictum because she grounds morality in human nature and not in reason as Kant does. I discuss this tension in her position later in the chapter.

8 One way to put this would be to ask how a constitutive account can avoid the myth of the given. One might wonder whether Korsgaard’s search for the source of normativity amounts
This question is made complicated by an ambiguity in her position. Earlier I suggested that Korsgaard’s view of an appropriate answer to the normative question was ambivalently described as needing to be ‘successfully addressed’ to the first person. It is unclear whether in order to successfully address the first person you need to persuade that person that he is indeed obligated in this case, or whether you simply need to demonstrate his obligation by making explicit that to which he is already implicitly committed. Korsgaard’s answer to the normative question has called upon both the agent’s activity in reasoning (reflective endorsement) and the principle of reasoning itself (the categorical imperative). It remains unclear whether the normative authority of reasons issues from the categorical imperative or the practice of reasoning in accordance with the categorical imperative.

It would be a mistake to assume that it is her flirtation with naturalism (her move to what we do in reasoning) that poses the difficulty for Korsgaard’s position. Moreover, it would be a gross mistake to suppose that naturalism is inconsistent with Sellarsian scruples. In its broadest scope, Sellars’s philosophy is an attempt to mediate between the competing claims of the scientific and perennial conceptions of persons in the universe (the scientific and manifest to a general skepticism about reasons, a demand that no theory can satisfy. At least sometimes her answer to the normative question appears to be an attempt to give the skeptic a reason to have reasons.

9 Note that on the persuasion view the role of the philosopher is to use the means at her disposal to lead the dubious agent into a particular state of belief, a state in which he affirms that he is indeed obligated. Failure to bring about that state of affirmation in the agent constitutes a failure to successfully address him in the first person. But according to the view in which successfully addressing the first person requires that the commitments implicit in the agent’s practice are made explicit to him, the agent’s failure to actually be persuaded of his obligation does not denote failure.
images). And, of course, non-naturalist, *a priori* accounts of justification are just as likely to postulate illegitimate mongrel concepts as naturalist attempts. The problem isn’t naturalism. Moreover, Sellars’s own view is clearly a form of naturalism. It is only forms of naturalism that attempt to offer purely descriptive explanatory accounts of norms which invoke the myth of the given. A naturalistic explanation of norms can avoid the myth of the given by conceiving of them in essentially practical and public terms. We will see that Korsgaard’s explanation is appropriately practical but fails to make reasons sufficiently public in their origin.

In order to explain obligation, in the first person terms which she deems necessary, Korsgaard turns to a discussion of the role of practical identity in the formation of reasons. According to Korsgaard, reflective endorsement is the source of reasons and practical identity is the source of obligation. Although reflective endorsement explains how we come to have reasons for action it doesn’t complete the explanation of how we are obligated by those reasons. It establishes that the source of our norms, the source of morality’s bindingness, originates with our capacity to give ourselves laws - to construct reasons for action. “The test of reflective endorsement is the test used by actual moral agents to establish the normativity of all their particular motives and inclinations. So the reflective endorsement test is not merely a way of justifying morality. *It is* *morality itself:*”11 This is just to say that our reflective nature requires

---


11 Korsgaard, *The Sources of Normativity*, 89. Korsgaard’s turn of phrase here mimics a remark Kant makes in the *Critique of Practical Reason*. When describing how the moral law
that we give laws to ourselves, and in turn that autonomy is the source of normativity. Our practical identities and our reasons overlap in important ways.

It is because we can give ourselves laws, because of our reflective capacities, that we have self-conceptions. Consider an example. Sophie is a graduate student who loves philosophy. While committed to education for its own sake, she has a deep practical streak which regularly inclines her to think about a move to a more ‘sensible’ profession. She wonders if it isn’t short-sighted of her to remain in graduate school where she is happy, but frets that a doctoral degree in philosophy doesn’t hold much promise for a career with financially security. Indeed, the wisdom of pursuing a career in philosophy has been more often in her thoughts since her friend, Gorgias, dropped out of the program in order to take a high-paying job with an advertising firm. From time to time she contemplates getting out of the academic rat-race and into the real rat-race, until one day Gorgias tempts her with the promise of a job as an advertising executive in the very firm for which Gorgias works. Gorgias encourages her to put her philosophical skills to good practical use and give up on the idea that the powers of reasoning and speech are best employed in uncovering truths – advertising is where it’s at.12 Sophie is sorely tempted by the offer: no provides its own incentive to action, and explaining the value of action from the sake of duty, he writes: “Thus respect for the law is not the incentive to morality; it is morality itself, regarded subjectively as an incentive, inasmuch as pure practical reason, by rejecting all the rival claims of self-love, give authority and absolute sovereignty to the law.” Critique of Practical Reason, trans., Mary Gregor (Cambridge: Cambridge University Press, 1997) AK V, 76.

12 In this respect Gorgias is much like her namesake. In the famous dialogue, Gorgias says that the art of rhetoric, the craftsman of persuasion, is about “being able to persuade by speeches judges in the law court, councillors in the council, assemblymen in the assembly, and in every other gathering whatsoever, when there is a political gathering. And indeed with this power you will have the doctor as your slave, and the trainer as your slave; and that moneymaker of yours will be plainly revealed to be making money for another and not for himself, but for you who can speak and persuade multitudes.” Plato, Gorgias, trans. James Nichols Jr. (Ithaca:
more financial worries, new possibilities for excitement and challenge. But advertising? Her thesis, after all, is about the Socratic notion of the reflective life. How could she reconcile her belief that the unexamined life is not worth living with the ideals of advertising? Isn't advertising just the promotion of non-reflective forms of living?

As Sophie sees it, the reflective life and the life of the advertising executive are not compatible. She acknowledges that philosophers can work in advertising and still remain philosophers, and advertising executives can pursue philosophy in their leisure time. But because she believes that the goals of advertising conflict with the goals of philosophy, the question of whether or not to take the job is complicated in ways it would not be for other philosophy students. At issue here is not whether advertising and philosophy are compatible. Sophie may be wrong about this. For the purposes of explaining the account of obligation we get from reflective endorsement it doesn't matter whether or not she gets it right. Her doubt concerns the authority of her own reasons to guide her action not whether she should give up her reasons because, as it turns out they are bad ones.¹³ Korsgaard wants us to see that our reasons inform our identities in important ways — ways relevant to an accounting of their normative authority. Sophie's decision will contribute to her self-conception. Her practical identity is informed by what she takes to be reasons for action.

Practical identity, Korsgaard tells us, is "a description under which you value yourself, a

¹³ What is relevant are the reasons why she believes the roles of the philosopher and the advertising executive to be at odds. Self-deceit will always be a possibility. Thoughts of soon to be affordable sunny vacations may induce her to now see that the roles are not at all incompatible, but that is another matter.
description under which you find your life to be worth living and your actions to be worth undertaking." Reasons aren't just rules, they are expressions of personal identity. We have obligations because we have personal identities which rebel against particular courses of action - obligations arise from what our identities forbid. ‘I couldn't live with myself if I did that’ expresses the rejection of a possible course of action as incompatible with commitments which inform one’s self-conception. So when the opportunity for financial security presents itself in a manner incompatible with Sophie’s commitment to a philosophical life her temptation to take the advertising job is blocked by the conflict generated by these two inconsonant alternatives. Because she thinks ‘I couldn’t live with myself if I made a career out of exploiting people,’ she is provoked into a rejection of that alternative, and feels obliged to decline the job.

Her focus on practical identity leaves Korsgaard well-placed to respond to the moral skeptic, and to answer the normative question, in the first person. We can imagine Sophie telling her that she feels that morally speaking it would be wrong to take the high paying advertising job, but going on to question how any theory of morality could really be justified in imposing such a significant demand upon her. Korsgaard has set herself up with the means to address her query quite directly. ‘Aren’t you committed to the philosophical life?’ she might reply. ‘Doesn’t the philosophical life which you claim to value preclude the promotion of non-reflective forms of existence?’ Sophie’s response to these questions provides an answer to the normative question as Korsgaard has framed it. At this point, let’s return to my worry about the ambiguity in Korsgaard’s view of what constitutes an answer to the normative question which successfully addresses the first person.

---

On the *persuasion* view, success in addressing Sophie’s version of the normative question requires that Korsgaard convince her that she ought not to take the job by appealing to features of her personal identity in such a way that she is indeed led to feel the force of the reflective rejection. ‘Yes, Korsgaard is right, I would be ashamed of myself, I’d be giving up on something very important to me’, she must be led to think. In contrast, on the *demonstration* view success in addressing the normative question requires only that Korsgaard clarify how particular features of Sophie’s personal identity comprise commitments to beliefs and attitudes which themselves ground her obligation. She need not be persuaded. The question we must now consider is whether for Korsgaard, the first person justification of reasons requires the creation of the relevant state of psychological conviction, or is instead a demonstration of the existence of that obligation. Let us look a bit more closely at each of these interpretations in turn, keeping in mind that it is not entirely clear which Korsgaard intends. What is clear is that she cannot have both. In any case, the persuasion model of first-person justification fails, as I will now show.

**The Persuasion Model**

It seems that Korsgaard must have the persuasion model in mind when she complains that the moral realist is without the genuine means to provide an answer to the normative question. According to Korsgaard, moral realism is a position which refuses to answer the normative question, or holds that it cannot be done, because it neglects the question of how reasons legitimately command. Sophie, for example, will not be satisfied with a demonstration that the moral principles to which she subscribes do, in fact, reject as impermissible a career in advertising.
Her concern isn’t with what she can derive from her principles but with the authority of these principles to guide her action. The normative question arises when the demands of morality are immense and threatening: “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 question is simply ‘Yes’. That is, all he can say is that it is true that this is what you ought to do.”¹⁵ The realist fails because his own response to an expressed lack of confidence in the authority of moral reasons is just a re-statement of confidence in its authority. Korsgaard complains that moral realism fails to answer the normative question because it has nothing to say to persuade the skeptic that obligations are indeed binding. When the skeptic considers rejecting the authority of her own reasons it will not do to restate the truth of the relevant moral principles, as the realist is limited to doing. The skeptic is looking for practical, not theoretical reasons, which will show her that morality is personally beneficial (“the idea is to show that morality is good for us”).¹⁶ The realist fails because he doesn’t have the means to instill confidence in the skeptic. He cannot instill confidence in the skeptic because he lacks any means with which to persuade her of her obligation.

There is another, even stronger, reason to think that Korsgaard has the persuasion model in mind. It appears to be a consequence of the view that obligation is a form of reflective rejection. Obligation is made possible because we have practical identities, and arises when one’s identity is threatened. Unconditional obligations spring, says Korsgaard, from conceptions of oneself that one couldn’t violate without critically damaging one’s practical identity:


It is the conceptions of ourselves that are most important to us that give rise to the unconditional obligations. For to violate them is to lose your integrity and so your identity, and to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking.  

Without self-conceptions we couldn’t have obligations. If the normative question arises because an agent doubts that obligations have normative authority over her, then in order to explain the form of obligation in the first person one must appeal to the ways in which the failure to do what one takes oneself to have reason to do constitutes a threat to one’s personal identity. If someone is doubting that she is really obligated to do what she takes herself to have reason to do, and being obligated is experiencing a threat to one’s personal identity, then establishing the reality of the obligation just amounts to establishing the threat to her personal identity of which she, at the moment, fails to feel the force. Clarifying the threat to her personal identity sounds a lot like convincing her that there is an incompatibility across the different aspects of her identity. To choose one is to forgo the other. But it wouldn’t be enough to just point out the incompatibility, of which she may be perfectly aware, for her problem is that at the moment she doesn’t experience it as a threat. So justifying her obligation to her is really helping her to feel appropriately threatened.

While this sounds a bit strange, it makes sense of Korsgaard’s account of the role of the moral emotions in the authority of reasons. Moral emotions can be understood as a feature of reason’s power to command. We can experience ourselves as fulfilling both legislative and active roles precisely because we give ourselves reasons for action. When we fail to live up to the

---

reasons which we set ourselves we tend to experience regret, shame, or guilt. Indeed, anticipation of such unpleasant emotions can sometimes motivate us to do what we ought to do. Emotions like regret, shame and guilt function in Korsgaard's account of reasons by contributing to the authority of the legislative part of the self over the active part of the self: "Although we are not supposed to do our duty out of the fear of punishment or the hope of reward, no one who cannot impose sanctions on us is in a position to require anything of us....a person's own mind does indeed impose sanctions on her: that when we don't do what we should, we punish ourselves, by guilt and regret and repentance and remorse." Korsgaard is clearly influenced here by Kant's discussion of the moral law's power to humiliate a less than perfect will. In the Critique of Practical Reason, Kant argues that our inclinations to self-love are humiliated by the supreme authority of the moral law, producing a pathological feeling of pain. However, humiliation under the moral law also produces a positive, non-pathological, feeling which is respect for the moral law. This is important for Kant because it allows him to explain how it is that the moral law itself serves as an incentive to action. Reference to the moral emotions is important for Korsgaard because it allows her to explain how it is that obligation represents a

---

18 Korsgaard, The Sources of Normativity, 151. The appeal to emotions such as regret isn’t unique to Korsgaard’s position. Michael Bratman’s “no regret condition” on planning also makes use of the information which anticipated regret offers. Appeals to moral and epistemic emotions, however, must be made with care because an emotion as felt response to a thought or experience itself offers no normative guidance. As features of normative operating systems (such as rational agents) emotions serve as indicators, as evidence that reasons might be at hand.

19 Kant, Critique of Practical Reason, A K V, 74. In her replies to her critics Korsgaard clarifies her Kantian affinities, noting that she thinks that Kant is mistaken in holding that our grounds for adopting maxims are restricted to either self-love or duty. Korsgaard sees the need for a less restricted set of options and so introduces practical identities as a substitute for Kant’s view of self-love. In, The Sources of Normative, 243.
threat to personal identity.

It is important to Korsgaard's position that the threat of sanction gives reason an authority it would not have were these sanctions not available. This leads to the question of whether the reason's authority lies in something intrinsic to it, or in its capacity to punish, a question which Korsgaard herself raises. However, her answer is less than clear. She says that a reason's authority doesn't "depend upon the experience of the negative moral emotions, but it absolutely implies it. A mind that could not perceive its reasons, after all, could not function as a mind at all."20 She seems to be saying that the power to sanction arises because reason has authority. But she has just said that without the power to sanction reason would lack authority, so we have come in a circle.

Emotions are not, for Korsgaard, simple physical or psychic sensations, but cognitive phenomena. They are perceptions of reasons. "Pain is the perception of a reason: that applies when we look back as well as at the present and when we look ahead. Someone who recalls failing to do what she was obligated to do will experience pain, and that is what remorse and regret are."21 So her own answer to the question of whether reason's authority lies in something intrinsic to it, or in its capacity to punish should be that the authority of reason doesn't depend on its capacity to punish because what is important about the negative emotions isn't their unpleasantness per se. It is because they represent reasons which the agent already has that they can serve as guides for action. For example, I might now regret having stayed up to watch the late movie because I will be overly tired, work ineffectively as a result, and because I knew that

20 Korsgaard, The Sources of Normativity, 151.

21 Korsgaard, The Sources of Normativity, 151.
I needed an early start and am responsible for the bad situation. My regret hinges on the fact that the reason was available to me and I declined to act on it.\textsuperscript{22} In contrast, if last night I had no reason to go to bed early then I cannot now (in the relevant way) regret having stayed up late to watch it. That applies even if I did, as a matter of fact, rise earlier than I would have liked (perhaps I was awakened by a phone call) and so must spend the rest of the day exhausted. In that case I might wish that I had gone to bed earlier or wish that I had not received the phone call, but I won’t regret my decision to stay up late. I won’t regret having stayed up late because to do so would be to imply that the indiscretion was mine, that I should have known better, and this was not the case.

Although a number of Korsgaard’s remarks suggests that it is so, she cannot genuinely hold that the authority of reasons derives from their power to sanction. To hold this is to make their authority contingent upon human tendencies to be moved by guilt and shame - tendencies which are notoriously unreliable. The problem with the persuasion model’s answer to the normative question is that it makes the justification of moral reasons depend upon an agent’s tendency to be properly educated into, and motivated by, these emotionally persuasive tendencies and this presupposes the normativity it is supposed to explain. That is to say, if a successful explanation of why a doubtful agent is obliged depends upon both the fact that she is disposed to care about emotions she is likely to experience, and that she is disposed to experience those emotions as having a particular normative force, then its very success depends upon the norms

\textsuperscript{22} Of course, regret is also possible (perhaps even more likely) in those circumstances where one adopts a course of action but fails to live up to it. But the object of regret is slightly different in that case. In the case of weakness of the will, I don’t regret the foolishness of my decision, but instead regret the folly of my behaviour; I rue my weakness of character.
in question already being in place. So it presupposes what it sets out to explain. Of course, one could pare down the persuasion model by removing some of these normative suppositions. But this would only move the problem one step back. We could tell Sophie that she is likely to feel like a phony if she takes the advertising job, and others are likely to accuse her of hypocrisy. But she may retort that with her massive salary she will be able to afford a steady supply of Valium, so that when such feelings arise she’ll be able to deal with them. The point is that unless the emotion is normatively charged it cannot aid in an answer to the normative question, but if it is normatively charged then it isn’t really doing the work. So however you look at it the persuasion model fails. Either it fails because it employs a non-normative concept of emotion in which case it doesn’t explain the normative authority of the obligation. Or, it fails because it employs a normative concept of emotion in which case it answers the normative question by appeal to the myth of the given. For this reason, the persuasion model employs a normative concept of emotion which instantiates a Sellarsian mongrel concept.

Earlier I suggested that Korsgaard’s position was ambiguous in its account of what constitutes an answer to the first person. I also pointed to an ambiguity in the suggestion that a first person justification of authority must appeal to personal identity, noting that practical identity could refer to particular features of an agent’s identity or to features of practical identity generally speaking. The relatedness of these two forms of ambiguity should now be apparent. We have just seen that although Korsgaard sometimes writes as though the job of the first person answer is to be persuasive, the model itself is fundamentally flawed. Next we will turn to an assessment of Korsgaard’s position in which her justification of practical reason is a demonstration of the norms implicit in practical reasoning, and the appeal to personal identity is understood in general
The Demonstration Model

Just as there is textual evidence supporting the persuasion model, there is also textual evidence suggesting that Korsgaard sees it as improbable. Elsewhere, while arguing that instrumental accounts of practical reason presuppose the normativity of generalizable principles, she notes the folly of attempts to persuade the skeptic that certain principles of reason are intrinsically normative. It is a hopeless task to try to persuade someone of basic principles of practical reason (for example, that you have reason to take the means to your ends) in the same way that it is hopeless to persuade someone that they have reason to believe in basic principles of theoretical reason (for example, that you have reason to accept the principle of non-contradiction):

The possibility of self-government essentially involves the possibility of its failure; and the principles of reason are therefore ineluctably normative. It is worth pointing out that an exactly parallel argument could be made about believing. We are neither inevitably inclined nor logically necessitated to believe the logical implications of our beliefs....And trying to persuade someone who actually doubted the instrumental principle that she should act on it would be like trying to persuade someone who actually doubted the principle of non-contradiction that he should believe it. It would be exactly like that.23

The persuasive task is hopelessly wrong-headed because the norms which are being justified are

themselves constitutive of (internal to) the practice of reasoning.

The demonstration model is the second interpretation of Korsgaard's response to the normative question we will examine. Because she holds that the practice of giving oneself reasons for action is ruled by the categorical imperative (reasoning about what to do just is employing the categorical imperative) Korsgaard's explication of their bindingness will, quite plausibly, be limited to a demonstration of their role in particular episodes of reasoning.²⁴ In other words, the rational necessity involved in doing what you take yourself to have reason to do is internal to the act of taking yourself to have a reason for action. According to Korsgaard, giving yourself a reason for action (willing an end) just is committing yourself to realizing that end.

So willing an end is equivalent to committing yourself, first-personally, to taking the means to that end. In willing an end, just as Kant says, your causality – the use of means – is already thought. What is constitutive of willing the end is not the outward act of actually taking the means but rather the inward, volitional act of prescribing the end along with the means it requires to yourself.²⁵

The practice of giving ourselves reasons for action is rationally normative action, governed by the principle which says that any rules we give ourselves should be universalizable. Korsgaard grants that it is neither logically nor causally necessary that an agent recognize obligation or the

²⁴ It isn't entirely correct to say that Korsgaard holds that practical reasoning is governed by the categorical imperative. In "The Normativity of Instrumental Reason," 244, she argues that practical reasoning is governed (or constituted) by the hypothetical imperative: "the act of making a maxim – the basic act of will – conforms to the instrumental principle by its very nature. To will an end just is to will to cause or realize the end, hence to will to take the means to the end." Nonetheless, it turns out that on her view the hypothetical imperative in turn depends upon something like a categorical imperative for its justification.

bindingness of reason (that is what makes it possible to raise the normative question). We are, nonetheless, rationally bound by the commitments we make in giving ourselves reasons for action. So our question becomes what, for Korsgaard, do those commitments consist of?

Let us return to the example of Sophie to see how Korsgaard might explain her obligation in these terms. The threat to Sophie’s personal identity arises, if and when it does, because what she values and what she wants collide. Her obligation to refuse the job offer springs from her practical identity, which forbids the adoption of plans obviously incompatible with the reflective life. Her reason to pursue a reflective life also gives her a reason not to pursue a life with which it is incompatible. The threat to her identity is something she has created. An account of her obligation would outline the reasons she takes herself to have, the values implicit in them, and the prohibitions they imply. Her reasons for action constitute rational commitments and so an explanation of their authority demands an accounting of the nature of those commitments.

Korsgaard’s position has the means to demonstrate at least two features of the rational commitments implicit in reasons. First, as we have already seen, by committing oneself to an end, one thereby commits to the means to that end. Giving oneself a reason is a way of understanding oneself as a form of causality: “willing is regarding yourself as the cause of the end in question - as the one who will bring it about. This distinguishes willing from mere wanting or wishing or desiring.”26 According to Korsgaard, when one takes oneself to have a reason to act one adopts a purpose and implicitly a means to attaining that purpose. A maxim which, if universalized,

would thwart the purpose for which it was adopted, invokes a contradiction. Minimally, for Korsgaard, when one takes oneself to have a reason to \( \phi \), one commits oneself to the means necessary for \( \phi \)-ing. So, it wouldn't do for Sophie to claim that she values the reflective life above all else, but lacks a reason to pursue the means to it. In order to demonstrate the form of her obligation to her we need only show her that she must understand herself as a form of causality in this way.

Second, by their nature, reasons involve a universalizability test which incorporates a commitment to the propriety of that judgement over time. The endorsement at issue is, after all, reflective and so includes a commitment to standard conditions. What is endorsed is not a psychological state, but a response to a set of conditions which one takes to be productive of a reason for action. The authority one lends to a reason by endorsing it is something for which one makes oneself responsible. By now judging that I have reason to \( \phi \) I am committing myself to the propriety of \( \phi \)-ing under the relevant circumstances in the future. The parallel in the theoretical case is not unlike the complex sorts of commitment accrued by the agent who comes to have a belief through an endorsement of his differential responsive disposition. Not only does he endorse his disposition, but also his commitment to the proper functioning of that disposition in these circumstances. Again to return to Sophie, her commitments would not disappear were she to claim that although she believes herself obliged to refuse the advertising job offer today, tomorrow, when she is likely to feel depressed, this reason need not apply. Her reason to avoid exploiting others, unlike her reason to choose chocolate over vanilla ice cream, does not vary

---

27 This is the view of contradictions in the formulation of maxims which Korsgaard develops in "Kant's Formula of Universal Law."
across circumstances in this way. Of course, it would be legitimate for her to change her mind and accept the job, but that is another matter. We can demonstrate the form of her obligation to Sophie by noting that having a reason for action commits her to claims about the conditions (standard conditions) which ground that reason; unless these conditions have changed in the relevant ways, the reason stands.

Roughly then, Korsgaard's short answer to the person who asks 'why must I act on my reason for action?' will have to be 'because in coming to have that reason you bound yourself to it - you made it normative for you!' By the very same act of will which made it your reason, you obligated yourself. But Korsgaard must also have a longer answer, for she insists that the answer to the normative question not only be first-personal, but also appeal to the doubter's practical identity. While the persuasion model seems more compatible with an appeal to particular features of an agent's practical identity, the demonstration model seems more compatible with an appeal to generic features of practical identity. On the demonstration model the application of the categorical imperative, for example, seems to count as a generic feature of practical identity. Thus an assessment of Korsgaard's answer to the normative question raises the issue of whether something like the categorical imperative is a generic feature of personal identity. This prompts another question. Is practical identity in the particular, or general sense, doing the normative work in reflective endorsement?

Because, for Korsgaard, reasons arise from reflective endorsement and obligation arises from reflective rejection, there is a conceptual difference between an explanation of how we come to have a reason for action and how we come to be obligated, a difference which points to a tension in her position. Normativity is traced both to a normative capacity (autonomy) and to a
psychological disposition (a threatened self-conception). Korsgaard errs when she attempts to explain justificatory notions like obligation, by means of descriptions of the psychological makeup of rational moral agents. A threat to a person’s identity is only normative for her (something that could possibly obligate her) if she takes a normative attitude toward it, and so the mere description of a state in which an agent’s personal identity is threatened cannot explain the nature of obligation. I may think of myself as both generous and compassionate and yet walk hurriedly past the beggar on the way to the bakery. My obligation to the beggar is not explained by noting that my guilt confirms a tension between my self-conception and my reasons for action in this case.

The short version of Korsgaard’s answer appeals to the reasons to which you are committed, and the long version appeals to threats to your practical identity. In an important way her answer to the normative question is double-edged. On her view, your reason for action binds you because you conferred authority upon it when you endorsed it. Moreover, you are bound by your reason because your self-conception leads you, even compels you, to act in accordance with your reasons in order to maintain a unified and non-discordant sense of self. Korsgaard tries to show that these two features of reasoning are really the same thing, and it is this move that deserves closer inspection. Her claim is that the rational commitment involved in reflective endorsement and the emotional commitment involved in reflective rejection are two-sides of the same coin. The question is whether this two-sided coin is genuine currency or that counterfeit, the mongrel conception of the myth of the given.
Reflective Endorsement and the Myth of the Given

Korsgaard’s naturalized Kantian view has the great virtue of being able to illuminate features of our moral psychology which Kant’s own position does not. But this illumination comes as a result of attempting to straddle incompatible positions. Korsgaard’s attempt to join these positions with a constitutive argument needs to be addressed. On the demonstration model, her answer to the normative question is an answer which addresses the doubter by outlining the form of her obligation. It doesn’t need to persuade the doubter, just clarify to the agent how she is bound by her own will. According to Korsgaard, she is bound by her own will because she has reflectively rejected a particular plan for action. Her obligation arises out of her reflective rejection and so an explanation of her obligation is an explanation of her reflective rejection. An explanation of one’s reflective rejection is an account of how it is that a person’s self-conception is in a discordant state. The ambiguity discovered when we try to unpack the idea that “obligation always takes the form of a reaction against a threat of a loss of identity” can be put roughly by posing the following question: is the obligation a result of the actual threat to identity

---

28 I think it uncontroversial to claim that Korsgaard’s account of practical identity and reflective rejection more subtly identifies the phenomenology of moral agency than does Kant’s account of the incentives of pure practical reason. Consider his brief discussion of our tendency to act on inclinations likely to benefit ourselves (out of self-love) and the dominance of the moral law within us, capable of humiliating us while keeping these tendencies in check. “For, all inclination and every sensible impulse is based on feeling, and the negative effect on feeling (by infringement upon the inclinations that takes place) is itself a feeling. Hence we can see a priori that the moral law, as the determining ground of the will, must by thwarting all our inclinations produce a feeling that can be called pain; and here we have the first and perhaps the only case in which we can determine a priori from concepts the relation of a cognition (here the cognition of pure practical reason) to the feeling of pleasure or displeasure.” Critique of Practical Reason, AK V, 73.

29 Korsgaard, The Sources of Normativity, 102.
(in the form of conflict in the agent’s self-chosen ends) or is it a result of the account of rational agency which is presupposed by the view? If it is threatened identity which obliges then obligation is the result of a causal process. If it is the consistency requirements constitutive of rational agency which oblige then we need an account of how obligation is formed. But as we have seen, she cannot really be arguing that it is my threatened identity which obligates me. That would be just to say that a state in which I happen to be obligates me, and that would just be a fact about me of the sort which Korsgaard has elsewhere ruled out. So the weight of her argument must fall on the constitutive view of the nature of rational agency.

Korsgaard’s view of having a reason can be put in the following way: A has a reason to \( \phi \) if \( A \) reflectively endorses a reason, or takes herself to have a reason to \( \phi \). Endorsement is a rule-governed, law-giving activity. Necessarily, when we give ourselves reasons we construct universalizable claims. This is what leads Korsgaard to assert that practical reasoning just is employing the categorical imperative, according to the unorthodox view of the categorical imperative which she adopts. Universalizable claims are those made with reference to certain circumstances which crucially circumscribe and define them. It is part of the nature of agency itself that willing is making oneself a causality and so committing oneself to the thought that in these conditions I should be or am necessitated to act in this way. In other words, universalizability is a kind of appeal to standard conditions – conditions which are suitable and thus reason-giving. More precisely then, we can interpret Korsgaard’s formula for reasons in the following way: A has a reason to \( \phi \) if \( A \) takes herself to have a reason to \( \phi \) under standard
conditions. It is Korsgaard’s account of the standard conditions which gets her into trouble and leaves her open to the charge that she is guilty of the myth of the given.

Actually, there are two issues here. One is the question of what makes conditions standard for Korsgaard. We have seen from Sellars that the recognition of conditions as standard requires an acknowledgement of the social character of those standards. That certain conditions may be correctly identified as standard conditions is a consequence of the authority those standards have in our practice. The other is the question of the direction of explanation. If Korsgaard understands the equivalence as a definition of what a reason is, she is guilty of the form of the myth of the given which, in epistemology, is called phenomenalism. She would have to, in effect, build intersubjective (and objective) reasons out of subjective reasons.

Korsgaard is clear that the authority of reasons originates in the practical act of taking oneself to have a reason for action. This she states in response to the charge that she has only identified the phenomenology of obligation and not obligation itself. She replies: “it is the endorsement, not the explanations and arguments that provide the material for the endorsement, that does the normative work.” In other words, it is because we take ourselves to have reasons that we actually have them. This sounds a lot like the claim that we come to know what something is by seeing what it looks like. Here Korsgaard is putting too much normative stock in what we take ourselves to have reason to do. Taking ourselves to have reasons to act presupposes our capacity to accurately distinguish between those circumstances which are, and

---

30 I have consciously phrased this to parallel Sellars’s formulation of the necessary equivalence between looks red and is red. The force of the claim that this equivalence is constitutive is captured by the fact that it is, when properly understood, a necessary truth.

31 Korsgaard, *The Sources of Normativity*, 257.
those which are not, reason-giving. It presupposes a conception of having a reason for action which is more robust than her account of reflective endorsement can provide. Reflective endorsement cannot establish a more substantial normative conception of reasons because it makes the psychological act of assenting authoritative, and is thus too thoroughgoingly a first-person notion. This becomes apparent in her explanation of interpersonal obligation and reason-giving.

According to Korsgaard, reflective endorsement will only yield an account of agent-relative reasons, and thus end in relativism, unless the first person commitments implicit in reflective endorsement impose commitments across reasoners. Her tack is to try to show that having a practical identity commits one to valuing humanity, which commits one to valuing other person’s reasons for action in the same way that one values one’s own. Her argument goes roughly as follows. We are the sort of creatures who act for reasons. Having reasons requires having values and practical identities. Because we have practical identities we are committed to valuing our humanity, for it is our humanity which makes it possible for us to have practical identities. Thus, my humanity and the humanity of others is valuable. Her transcendental argument is supposed to show that if you value anything at all, if you acknowledge the existence of practical reasons, then you must value your humanity as an end in itself. Valuing humanity carries with it obligations to others, which is tantamount to being open to their reasons for action. This transcendental argument is supposed to supply a foundation for her account of the intersubjectivity of reasons. Korsgaard asserts that a public, as opposed to private, conception of reasons is needed. With this I agree. However, because her answer to the normative question grounds the authority of reason in first person psychological acts of endorsement, her
transcendental argument is insufficient to ground the public character of reasons. Phenomenalism about reasons is no more successful at constructing norms out of 'private reasons' than phenomenalism about physical objects is at constructing things out of appearances.

The problem with trying to get public reasons from private reasons is that the move to intersubjectivity from reflective endorsement requires an account of why your reasons should matter to me. If we start with an essentially private conception of reasons and try to build intersubjective reasons out of it, we can never bridge the gap to shareable reasons. For example, it turns out that I have an obligation to myself to treat you in ways which accord with the value I place on you, but I don't owe anything to you as such. Korsgaard herself makes these very criticisms of private conceptions of reasons. She is committed to the view that reasons, like language, cannot be strictly private and still communicable and thus is committed to the view that reasons are public in the same way that language must be. We give each other reasons, she says, in much the same way that we give ourselves reasons. When you ask me to close the door, you do not thereby force me to close the door but, according to Korsgaard, you have obligated me by giving me a reason. The reasons of others have much the same role as our own desires. We don't need a reason to take them into account, we need a reason not to take them into account. To the egoist challenge that we need not treat the demands of others as reasons (unless we have some further personal reason to do so) Korsgaard replies that reason-giving interactions are instances of confronting or reminding the other of the value of our humanity. 'How would you like it if I did that to you?' causes the other to think 'I would dislike it,' which forces him to acknowledge the value of one's humanity.

There are several things to be said in response. First, we can see that Korsgaard appears
to be influenced by what I earlier identified as the persuasion model of justification. The transcendental argument looks like an attempt to persuade the egoist that other people's reasons matter, yet there may be nothing we can do to persuade the egoist. The bindingness of reasons does not depend on our success in persuading the skeptic that he is bound by the reasons of others. Second, while it is true that reasons must be in principle shareable, and in this sense public, Korsgaard's account of what obligates (namely, reflection in the first person) makes the bindingness of reasons independent of their shareability. Korsgaard, in effect, tries to build the space of reasons out of an aggregate of the private reflective acts of endorsers who are tied together by the transcendental fact of their humanity. But this makes the space of reasons the product of the essentially private activity of autonomous agents. The space of reasons cannot be constructed out of individual acts of reflection and transcendental glue because insofar as the individual acts of reflection constitute reasons for action they already are stands in the space of reasons.

It should now be clear how her transcendental argument fails. It presupposes what it sets out to explain by having the fact of the other's humanity serve as a norm. The skeptic might accept that other agents instantiate the humanity Korsgaard talks about, but deny that this fact gives him a reason to respond one way or another. It need not give him a reason because on the reflective endorsement model, a thought or judgement serves as a reason only once reflectively endorsed. Korsgaard's appeal to the transcendental value of humanity amounts to an attempt to give the skeptic a reason to have reasons. But this is folly. Reasons aren't intersubjective because we have a moral obligation to recognize the humanity of other agents. They are intersubjective because they function within a public practice of giving and asking for reasons.
The public character of reasons is a conceptual fact about them, not a consequence of a moral imperative. The myth of Jones, which showed us that our very conception of mental episodes is forged in public terms, reminds us that our reasons are not constructed out of private acts of endorsement. We learn to identify our psychological states as reason-giving only by mastering their role and significance in a public network of reasons, in light of which they can be experienced as reliable symptoms of the presence of reasons which are already public. The skeptic already has the reasons which Korsgaard is trying to defend, he is simply being perverse in denying them.

To put this point in slightly different terms, we can see Korsgaard’s transcendental argument as offering a realist solution (of the sort she herself rejects) to the problem of normativity. As she has put it, the transcendental argument supplies a link between the first person and the public in the form of a claim that a commitment to valuing the other’s humanity is implicit in being a reasoner at all. But here Korsgaard moves away from lodging the normativity of reasons in the practice of reasoning to grounding it in a mind independent fact of the sort to which the realist appeals. This is a fact about us as reasoners — a theoretical fact — that our humanity is valuable. This makes the publicity of reasons, rather than being the constitutive feature of the practice of reasoning which she has claimed they are, depend on a theoretical fact about us as reasoners. Korsgaard’s claim that we are committed to valuing each other’s reasons because implicitly committed to valuing their humanity is another version of the realist attempt to answer the normative question by fiat.

Korsgaard’s account of the publicity of reasons reveals the tension in her view, which I have already discussed. If reflective endorsement is doing all of the normative work, then it isn’t
clear how she can get outside of the first person in order to make reasons intersubjective. If, on the other hand, the categorical imperative, as the principle of practical reason, is doing the normative work then she cannot get outside of the first person either. She cannot, because her pared down view the categorical imperative as a law-giving operation internal to reason-giving practice is essentially first-person. We need an account of reasons and reason-giving which is public and social from the start. This will be the task of the next chapter. I will show that practical reasoning, and theoretical reasoning are both irreducibly public activities.
Chapter Five:
The Constitution of Agency and the Public Authority of Reason

In order to set up the argument of this chapter, let me first briefly sketch the path that has led us here. I have argued that although instrumental accounts of practical reasoning claim that reason supplies only means to non-rational ends, these same accounts presuppose normative principles of reason. I have gone on to suggest that in order to answer the normative question posed of practical reasoning, these normative principles of reason must be conceived of as part of the practice of practical reasoning itself. The normative principles of reasoning are, in a sense yet to be specified, constitutive of practical reasoning. A constitutive argument such as Korsgaard’s makes advances over instrumentalist accounts of practical reasoning which simply assume the bindingness of rational principles such as the hypothetical imperative. However, I have cautioned against accepting Korsgaard’s constitutive argument too quickly, arguing that we need to be careful not to invoke the myth of the given at a higher level of abstraction. We need to avoid phenomenalism about reasons, a position according to which reasons are constituted out of their appearances. The problem arises when a public or intersubjective account of the authority of reasons is constructed out of private reasons – reasons whose normative authority is conceptually first-personal. Korsgaard is right to hold that she needs a public conception of
reason, however, it is not clear that her position has the means to provide it.

In this chapter I will argue that constitutive arguments, properly understood, explain the authority of reason. Reason's authority is bound up in the practice of reasoning because reasons function within a public network of inferentially articulated commitments and entitlements. A public conception of the authority of reasons provides the resources to answer the normative question without invoking the myth of the given. I will here present the structure of rational norms in terms of the intersubjective commitments constitutive of the practice of reasoning. I begin by broadening the scope of the discussion to include theoretical reasoning. By posing the normative question of theoretical reasoning we see that problems which are often taken to be exclusive to practical reasoning, are also problems for theoretical reasoning. This suggests that it is the nature of rationality, not morality as such, that deserves our philosophical attention.

Constitutive arguments offer an attractive response to the normative question. A constitutive argument is one which attempts to show that the very nature of a certain kind of thing or property, $A$, depends upon some other thing, $B$, in the sense that without $B$, $A$ wouldn't be what it is. For example, it might be argued that it is constitutive of rational agency that the agent is bound by norms, because if the agent were not bound by rational norms she would not be an agent. If accountability to reasons is constitutive of rational agency, then rational norms are implicit in reasons for action and belief, and so the justification of rational norms is of a piece with the practice of reasoning. In this thesis I am arguing for a constitutive account of the authority of reason. In this chapter we shall see that even after having turned to practice for the justification of reasons, the temptation to mobilize the myth of the given can remain. Constitutive arguments offer the promise of understanding reason's authority in the practice of reasoning itself,
but are misappropriated by those with a foundationalist conception of justification. The search for a non-inferentially justified basic belief on which to ground reason’s authority, distorts the constitutive insight. I argue that constitutive arguments can explain the authority of reason when a public conception of reasons is in place. The practice of taking ourselves to have reasons for belief and action subsists within a larger network of inferentially articulated commitments and entitlements. It is this larger network of public reasons which serves to explain reason’s authority. A public account of the authority of reasons will demonstrate that the distinction between practical and theoretical reasoning can be made in terms of the commitments they entail. This way of marking the distinction suggests that the appeal of instrumentalist accounts of practical reason is, in large measure, the result of a questionable way of distinguishing practical and theoretical reason.

Posing the Normative Question of Theoretical Reason

Just as it was possible to question the rational authority of practical reasons, so is it possible to question the rational authority of theoretical reasons. Admittedly, we seem less disposed to skeptically confront the authority of our theoretical judgements outside of philosophical debate than we do the authority of our practical judgements. Still, the possibility remains. Arguably, the gusto with which philosophers question practical reason’s authority, while finding theoretical reason’s authority uncontroversial, points as much to a tendency to think that the justification of practical and theoretical judgements are unrelated as it does to a human predilection to dispute the authority of intentions more than we dispute the authority of beliefs.
The question 'why believe what you have reason to believe?' demands that we consider not just what led to a particular belief, but what justifies continuing to hold that belief.

In order to get to the heart of the issue regarding theoretical reasoning I will start with a narrowly circumscribed formulation of the problem, a case in which the conclusion of a deductively valid argument is entailed by its premises. By taking up the question in this restrictive way we minimize the opportunities for obscuring the issue. However, by taking up this somewhat austere starting point I do not mean to rule out the possibility that the justification of particular reasons for belief may have complex normative dimensions. We ought not, for example, to foreclose on the possibility that there can be good practical reasons for adopting beliefs.\(^1\) We will not, however, consider cases of this sort here.

Lewis Carroll’s dialogue between Achilles and the Tortoise shows us that an explanation of the normative authority of theoretical reasoning faces challenges similar to those raised by asking the normative question of practical reasoning.\(^2\) In the dialogue, the Tortoise challenges Achilles to force him logically to accept a proposition \(Z\) on the grounds that he has accepted two supporting propositions \(A\) and \(B\). The propositions in question correspond to Euclid’s First Proposition and go as follows:

\[\text{In "Practical Reasoning and Acceptance in a Context," Michael Bratman discusses the possibility that special relations such as friendship might put legitimate constraints on the acceptance of theoretical judgements. In his *Faces of Intention* (Cambridge: Cambridge University Press, 1999) 15-34.}\]

\[\text{Of course, not everyone agrees that this is what Carroll’s fable shows. It has been argued, for example, that Carroll’s puzzle really shows that he lacks an adequate conception of deduction. See Timothy Smiley, "A Tale of Two Tortoises," *Mind*, 104, (1995), 725-735. For our purposes Carroll’s story provides a nice parallel to the normative question asked of practical reasoning.}\]
(A) Things that are equal to the same are equal to each other.
(B) The two sides of this Triangle are things that are equal to the same.
(Z) The two sides of this Triangle are equal to each other.³

Carroll’s belligerent Tortoise claims to accept that $A$ and $B$ are true, and although he doesn’t question the argument’s validity, does not accept that $Z$ is true. Achilles’s task is to show the Tortoise that he must, under the force of logical necessity, accept $Z$ because he has accepted $A$ and $B$. In other words, Achilles’s task is to show the Tortoise that he is rationally committed to $Z$ because he is rationally committed to both $A$ and $B$. Achilles’s first attempt invokes an additional premise, $C$, which he takes to provide the link between the premises which the Tortoise has accepted, and the argument’s conclusion.

(C) If $A$ and $B$ are true, $Z$ must be true.⁴

But even once the Tortoise agrees to accept $C$, having already committed himself to $A$ and $B$, he refuses to accept $Z$ on these grounds. He accepts the conditional (if $A$ & $B$ & $C$, then $Z$) but doesn’t draw the conclusion based on that conditional and the accepted premises. The Tortoise’s strategy is to deny that he is logically forced to infer the conclusion on the basis of the accepted premises, and thereby force Achilles into an infinite regress of justification by way of additional

---

hypothetical premises. Achilles responds by attempting again and again to force the Tortoise to make that inference. By the end of the story, the Tortoise has persevered, and refuses to grant the argument's conclusion on the basis of its now one thousand and one premises.

An examination of just how Achilles fails to demonstrate the logical necessity in accepting the argument's conclusion will help us to see how not to respond to the normative question. Failing to see that being bound by reasons is constitutive of being a reasoner, Achilles attempts to provide a non-constitutive form of argument for the authority of rational norms. The Tortoise has, in effect, demanded a reason why he must be constrained by the relevant rational norms. Achilles's strategy is to provide additional hypothetical premises which he supposes will add up to a categorical reason to accept the conclusion. Achilles's response is understandable. If introductory logic students approached the subject with the Tortoise's attitude they wouldn't get beyond *modus ponens*, and would surely leave their instructor very frustrated. But introductory logic students don't really adopt the Tortoise's attitude, they simply often fail to see the point of a logical rule. The problem is that Achilles's response mistakes the radical character of the Tortoise's position; it doesn't go deep enough. There is a way to respond to the Tortoise which demonstrates his rational commitment to $Z$. As we will see, it will be a constitutive explanation. For now, let us continue with the thought that something can be learned from Achilles's failure.

To clear the way for the kind of constitutive argument which I claim is more promising, I will first briefly rehearse a couple of unsatisfactory explanations of how it is that the Tortoise is committed to the argument's conclusion. One might be tempted to explain the Tortoise's

---

5 Predictably premise $D$ (If $A$ and $B$ and $C$ are true, $Z$ must be true) adds no further support for the conclusion, but merely spells out the form of the inference Achilles already sees as necessary.
mistake by examining the logical properties of the belief set with which he is left. If we take validity to mean that the denial of the conclusion of an argument is inconsistent with the acceptance of the premises then, if the argument is valid, the Tortoise might be accused of courting inconsistent beliefs. Clearly, the Tortoise would have an inconsistent belief set if he were to hold that $A$, $B$ and not-$Z$. Is it possible that, with some assistance, Achilles could force the Tortoise to accept $Z$ on the grounds that he had accepted $A$ and $B$ by showing him that his beliefs are inconsistent?

As a strategy for demonstrating the authority of rational norms, charging the Tortoise with inconsistent beliefs is a non-starter because the normative force of that criticism depends upon the relevant epistemic norms already being in place. Consider two responses which the Tortoise might make. First, our clever Tortoise may be able to escape the inconsistency charge by making the questionable claim that he hasn’t actually committed himself to the truth of not-$Z$, but has only refused to accept $Z$ on the grounds of $A$ and $B$. $A$ and $B$ are the only relevant propositions to which he is verbally committed, and because they are not inconsistent he cannot be charged with inconsistency for not accepting the conclusion they entail. Moreover, the relevant issue is not whether $A$, $B$, and $Z$ form a consistent set of propositions, but rather, when put in terms of validity, whether an agent is logically committed to accept the conclusion of an argument because she has accepted the premises of that argument. This brings us to a second possible reply. Even if the Tortoise were to allow that his refusal to accept $Z$ could legitimately be construed as an acceptance of not-$Z$, and thus that he did have an inconsistent belief set, Achilles would not yet have made his point. The Tortoise’s challenge was for Achilles to force him “logically, to accept $Z$ as true.” The Tortoise isn’t simply being stubborn in his refusal to endorse $Z$. He grants that
Z logically follows - he does accept the conditional (if $A \& B \& C$, then $Z$). But this skeptical Tortoise will take neither the fact that the argument is valid, nor that he has inconsistent beliefs, as a reason to accept its conclusion. He simply denies that a logical rule gives him a reason for action. We can imagine the Tortoise blithely responding to Achilles, 'yes I have an inconsistent belief set, but that is just a fact about me!' Indeed, the Tortoise might even grant that were he disposed to have consistent belief sets, he would likely conclude $Z$, but he doesn't.

Is the Tortoise's demand that Achilles demonstrate the logical necessity of accepting the argument's conclusion ('to force me, logically, to accept $Z$ as true') impossible to meet? We can unpack 'logically force' in two ways. This is important because not only does each unpacking suggest different answers to the problem, but they do so because they reflect two different understandings of the problem posed. Clearly there are both causal and normative strands involved in Achilles's predicament. The causal strand is evident in the Tortoise's refusal to accept the conclusion without offering any reasons for this refusal. Instead he points to the fact that it is possible that one could accept $A$ and $B$ as true, along with the conditional (if $A$ and $B$ are true, $Z$ must be true) without being moved to accept $Z$. The Tortoise gets much mileage out of the fact that accepting $A$ and $B$ doesn't causally necessitate accepting $Z$. The normative strand is evident in the thought that the Tortoise is making a mistake by failing to live up to a logical standard, a standard which binds him even as he fails to live up to it. The focus of the first approach is on getting the Tortoise actually to accept the conclusion - to use the logical means available to persuade him to accept that the conclusion does, indeed, follow from the premises. The other

---

6 As we shall see, Achilles's mistake is to allow that one could accept $A$ and $B$, and the proposition that if $A$ and $B$ are true, $Z$ must be true, and nonetheless "as yet" be under no logical necessity to accept $Z$ as true.
approach is to employ the logical means available to demonstrate the logical necessity involved in accepting the conclusion. A successful demonstration of the logical necessity of accepting the conclusion need not involve the Tortoise's sincerely uttering 'Z is true, on the grounds that A and B!'. On the demonstration model, an explanation of the logical necessity of accepting the conclusion requires that the reasoner's commitment to the conclusion (on the basis of the premises) be made explicit, not that the reasoner be persuaded to accept the conclusion.

The persuasion model, which Achilles seems to adopt, is hopelessly wrong-headed. Persuading an agent that he is rationally bound to a particular proposition is neither necessary nor sufficient for demonstrating reason's bindingness. It is not necessary because reasons's authority cannot be contingent upon the willingness of an agent to grant that authority on a particular occasion -- if it were contingent then it would not be authoritative.\(^7\) It is not sufficient because it is possible to be persuaded that one is rationally bound to a particular conclusion without the episode of persuasion serving as an explanation of how one is bound by that conclusion. For example, I may be persuaded that 'the keys are in the basket' after I have claimed 'the keys are either in the basket or in my coat pocket' and you tell me that they are not in my coat pocket. The authority of modus tollendo ponens is not thereby explained, even though it aptly identifies the logical form of the deduction. Besides, one may be persuaded by the wrong considerations. So let us abandon the persuasive model and proceed to a discussion of how reason's capacity to

\(^7\) In the case of practical reasoning we saw that the persuasive task fails because it misses the point. The norms for which we seek justification are themselves constitutive of the practice of reasoning. Something like the persuasive model appears more intuitive, albeit still wrong-headed, in the practical case because desires are taken to provide a foundation for reasons ('you ought to φ because you really value φ-ing'). To persuade an agent that she really has a reason to act is just to remind her of the desires or values to which she is already committed.
make demands of us might be demonstrated.

Answers to the normative question fall into two categories. The justification of theoretical reasoning will either be found within reasoning itself, or outside of it. A constitutive model of justification explains reason's authority as internal to the practice of reasoning itself. So, a constitutive model will seek to answer the Tortoise's challenge by showing that his participation in reasoning just is a commitment to the rational norms for which he seeks further justification. It is not possible for an agent to engage in the practice of reasoning without being accountable to rational norms since such accountability is constitutive of his agency. A non-constitutive model of justification will explain reason's authority as external to, or independent of, the practice of reasoning itself and thus will seek to answer the Tortoise's challenge by providing a justification of reasoning distinct from its operation.

Achilles adopts a non-constitutive model, holding fast to the idea that logic has the power to compel the Tortoise to accept the conclusion ("Logic would take you by the throat, and force you to do it!"). His mistake is to suppose that norms distinct from the norms of rationality confer authority upon them, thereby justifying them. His is a non-constitutive view because on it the norms of reasoning get their force from the rules of logic. Although it may seem natural to suppose that the authority of logic is part of the practice of reasoning itself, it is important to recognize that there is gap between the rules of logic and reasoning. Without it Carroll's dialogue would make no point. If deductive inference is thought to demand a suppressed hypothetical premise, a justificatory regress is started. And yet Achilles's appeal to the rules of logic as the ultimate regress stopper is not wholly groundless. We often invoke the authority of the rules of

---

logic to authorize conclusions or expose errors. When an introductory philosophy student is told that her argument fails because it violates the principle of non-contradiction, we expect this criticism to carry normative weight. She ought, we assume, to be guided by the rules of logic. Indeed, if we weren't able to make such an assumption, we wouldn't get far. But notice that it is the fact that she ought to be guided by the rules of logic, rather than the rules of logic per se that serve to explain the normative commitment. When we point out that her argument is inconsistent and violates the law of non-contradiction, our appeal to the law makes explicit as a principle the failure in her reasoning. The authority of the rules of logic is not explained by pointing to them. It is possible to again ask how one is bound by them, thus inviting a regress. This is the mistake that Achilles makes.

We might say that Achilles confuses implication and inference. This is how he comes to mistake the source of the authority of reason. It is one thing to say that 'A and B imply Z,' but another to say that 'if you accept A and B, you ought to accept Z'. Another way to put this point is to hold that there is a difference between logic and reasoning. The need for the distinction is clarified once we see that logic can never 'take you by the throat and force you to' accept a conclusion. It is simply false that the 'demands' of logic coincide with, or adequately represent, the demands of reasoning. The rules of logic and good epistemic practice do not always coincide. For example, it may not be reasonable to accept the conclusion of a valid argument. In some cases an agent may see that an argument is valid but have other good reasons for

---


10 For a discussion of such cases see Gilbert Harman, "Rationality," 9-45.
believing that the conclusion is false. In such a case it may be most reasonable to give up a previously accepted premise rather than her independently arrived at conclusion. Entailment doesn’t settle the issue of what it is reasonable to believe. It cannot tell you whether to change your mind about an accepted premise, or give up your reasons for believing that the conclusion is false. Still, one might conclude that entailment forces you to make a choice between conclusion and premises.

The argument involved in “What the Tortoise Said to Achilles” shows that this is not so. It is distinctive precisely because there seems to be no good reason for the Tortoise to deny the conclusion. Because the argument’s conclusion is based on a principle of logic, it avoids the subtle and complicated issues which concern good epistemic practice. It doesn’t leave room for the possibility that there are other good reasons for the Tortoise to deny the conclusion. Carroll’s dialogue is particularly useful in a discussion of the authority of theoretical reason because his logical puzzle puts the question (how rational inference is justified) in its simplest form. We have seen that the puzzle cannot be resolved by pointing to the rules of logic. A successful answer to the normative question will be one in which reason’s bindingness is internal to the practice of reasoning. As reasoners we are bound by rational norms not because they inherit their authority from something external to reasoning, but because to be a reasoner just is to be bound by rational norms.

---

11 This is how G. F. Schueler puts the point to be extracted from Carroll’s fable. In “Why “Oughts” are not Facts (Or What the Tortoise and Achilles Taught Mrs. Ganderhoot and Me about Practical Reason),” *Mind*, 104, (1995), 713–723.
Constitutive Arguments and Agency

According to the constitutive view, the normative authority of reasons is found within the practice of reasoning itself. We, reasoners, are bound by rational standards because to engage in reasoning just is to be accountable to rational standards. An argument that says that some feature is internally related to a thing's being such and so is a constitutive argument. Agents, as opposed to creatures unable to form beliefs and intentions and to act upon them, just are those upon whom rationality makes demands. A constitutive view of reason's authority thus provides a positive answer to the normative question - agents are bound by reasons because they are reasoners. Accordingly, an argument which shows that it is constitutive of agency that agents are bound by reasons, will explicate the concept of agency in such a way that the normative authority of reasons is internal to being an agent. Roughly, my argument will show that agents are constitutively bound by reasons by showing that being responsive to reasons is implicit in having reasons at all.

To mount such an argument we need to ask what the very idea of agency involves. What is it that we say of a person in calling her an agent? For someone to be an agent is, among other things, for her to have beliefs which she takes to be true, upon which she acts and for which she is responsible. A belief is, minimally, a propositional attitude which aims at the truth. For an agent to have a belief is thus for her to be committed to the truth of a propositional content, in other words, to be committed to the world being one way rather than another. Her commitment to the world being one way rather than another includes a commitment to the propriety of her belief. When an agent believes that p, she takes p to be true. To the extent that things can be said for or against believing p in those circumstances, the agent is implicitly committed to responding
to reasons for and against taking \( p \) to be true.\(^{12}\) Thus \( qua \) believer, to be an agent is to be bound by reasons.

It would be a mistake to think that these features of agency can be pried apart. Suppose, for example, a person attempted to adopt the policy of believing whatever first popped into her head, so that lengthy deliberation of any sort was unnecessary. In order to consider her propositional attitudes to be beliefs, and so to aim at the truth, we would have to take seriously the idea that adopting such a policy could lead one to have true, rather than false, beliefs. In short, we would have to entertain the idea that it was an epistemic policy and thus one for which reasons could be given. But this seems silly. There is no obvious connection between a thought popping into one’s head and its truth. Because having a belief includes a commitment to the propriety of that belief, we would wonder at the character of the commitment which such a policy would imply. Such a person is unlikely to able to offer reasons for adhering to her policy and unlikely to be able to offer a defence of any of her claims. As a result, we could not grant that her propositional attitudes were epistemically substantial enough to constitute beliefs. We simply couldn’t make sense of them as claims aiming at the truth since nothing could count against them.

For the sake of argument, let us further suppose that her first thoughts were always true, say, because an Evil Genius set up the world in that way. Even in this case, the question of her agency and the status of her thoughts as beliefs would not be established until we determined the nature of her attitude towards them. Her infallibility might give her status as an oracle, but not yet as a believer. In order to gain status as a believer, an agent, there would have to be deontic

\(^{12}\) For a helpful discussion of the constitutive role of belief in agency see Cheryl Misak, *Truth, Politics, Morality* (New York: Routledge, 2000), especially chapter two.
commitments implicit in her claims, for which we could hold her responsible. Even if her thoughts were infallibly true there would be no reason to attribute deontic commitments to her. After all, her thoughts are just whatever comes into her head first. The pronouncements of oracles, unlike the assertions of agents, need not, given their divine providence, be responsive to reasons. With respect to the truth of their pronouncements we can treat oracles as mechanisms since we don’t hold them responsible -- we would no more demand reasons of an oracle than we would of an alarm clock.

For a believer the matter is different. In addition to merely having thoughts she needs reason to think them worthy of her own endorsement. She needs to be able to take her deontic commitments as sufficient for belief. Since it is her own belief, she must take her own reasons to be adequate grounds on which to make a judgement. In short, she needs to be responsive to reasons even in the first person. So in holding her responsible for her commitments we are not placing new obligations upon her.

This is not to make the justification of beliefs an internal matter, privileging the first person. Many beliefs, like non-inferentially elicited perceptual beliefs, are produced in us by processes which we cannot directly evaluate. Such beliefs implicitly commit us to the reliability of our differential responsive dispositions, about which we can be in error. Other beliefs will not carry this implication. For many beliefs, the commitments accrued depend on inferential relations to other beliefs. No agent could have only one belief. To have beliefs at all is to have many beliefs which stand in inferential relations to each other and jointly constitute a picture of what the world is like. This picture of what the world is like is the world as the believer takes it to be, including the agent’s implicit deontic commitments. Her commitments constitute a stand in the
space of reasons, a stand for which she is responsible both to us and to herself.

Being an agent thus leaves her open to demands for justification. A commitment to the truth of a claim makes the agent responsible in a public way. Since there is something to be said in favour of or against almost any belief, by taking a proposition to be worthy of belief, an agent incurs obligations to defend and support belief in it. This is, of course, a conceptual point, not a claim that all beliefs must be publicly defended in order to be beliefs. Moreover, agents will stand in different evidential relations to a particular belief and hence are in qualitatively different positions from which to confer authority on that belief. Recognition of these differences requires us to be open to the credibility of the epistemic authority of others. Once we see that responsibility to the demands for justification is intrinsic to belief, we are in a position to see how, for example, dogmatism is a degraded form of cognitive attitude. The concept of belief involves a robust form of epistemic responsibility to respond to criticism. The dogmatist may be characterized by her failure to meet this responsibility. She simply refuses to take up the obligation of responding to reasonable criticism or contrary evidence and as a result the products of her dogmatic deliberation provide others no license for belief.

The constitutive argument identifies the source of the authority of reasons in both practical and theoretical reasoning as part of the practice of reasoning. One might think that this argument amounts to a transcendental argument for the correctness of some particular set of rational norms. If this were true the obligation to aim at the truth would be met by following a set of rational norms guaranteed a priori to succeed. However, I have not argued that the very possibility of reasons depends upon a particular conception of substantive rational principles being in place. Instead, the constitutive argument offered here simply makes the claim that it is
constitutive of agency that a believer be responsive to reasons. Nor should this constitutive argument be taken to have the conventionalist consequence that we cannot make mistakes in reasoning. It is constitutive of agency that an agent’s beliefs aim at the truth, but aiming at the truth cannot by itself offer guidance about which direction to aim. The agent can only satisfy the aim by being responsive to reasons, and which reasons the agent should be responsive to is not something that can be determined independently of our experience as agents in the world.

The constitutive model of agency offers a strategic way to respond to the normative question because the very act of questioning the authority of reasoning implicitly relies upon its powers to justify belief. To challenge reason’s authority is to invoke reasoning in order to undermine it. We are now well-placed to respond to the normative question by challenging the idea that the normative authority of reason can simultaneously be invoked and denied. I will argue that because a constitutive argument claims that the very nature of a certain kind of thing or property, $A$, depends upon some other thing, $B$, in the sense that without $B$, $A$ wouldn’t be what it is, it will have the consequence, if true, that claims of the form ‘$A$ but not $B$ obtains’ are conceptually confused. Still, the claim that you are bound by reason because your beliefs must be responsive to reasons is by itself insufficient to meet the normative question. Just how the constitutive view is able to demonstrate the bindingness of reason will be the focus of the remainder of this chapter.

In order to clarify how constitutive arguments work, I’ll first rehearse a series of objections to them and then show how they go wrong. The point of the objections is to show that constitutive accounts of belief and agency cannot be correct because they lack the means to explain how we are bound by reasons. Peter Railton argues that constitutive arguments are
useful, but limited. According to Railton, a successful constitutive argument would have to show the skeptic that standards of reasoning are not optional, but binding irrespective of the agent’s contingent personal ends. Although he grants that an implicit commitment to reason’s authority is evident in the practice of reason, he denies that this is sufficient to demonstrate reason’s authority. An agent’s commitment to purportedly constitutive features of rational agency can always be questioned, he argues, and this suggests constitutive arguments aren’t up to the task of explaining reason’s authority. The failure of constitutive arguments lies in their failure to provide a “self-sufficient non-hypothetical response” to the normative question.

By exploiting the gap between what one can believe and what one can say one believes Railton constructs a skeptical reply to the constitutive view. This raises the issue of what it is to doubt something. Clearly, an agent can utter the words ‘I question...’ without believing them. One can mentally frame the thought ‘I do not exist,’ but it is not at all clear that one can have this thought in the mode of belief. Questioning involves placing a certain thought in doubt in the sense of withdrawing one’s commitment to it. We shall see that on Railton’s view, acts of questioning at times seem like rationally unconstrained free choices. On the constitutive view propositional attitudes like believing and questioning are not unconstrained free choices.

Self-reflection may falsely suggest that the authority of reason can be denied or repudiated in this first-personal way. Since one can say anything, it might seem that one can deny the authority of a reason simply by uttering a challenge. But our constitutive argument shows that

13 Peter Railton, “On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action,” Ethics and Practical Reason, eds. Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997) 53-80. Railton is careful to note that to “show that a norm or reason is non-hypothetical is not to show that it is utterly without condition. It is only to show that it would necessarily apply to any agent as such, regardless of her contingent personal ends.” 58.
the authority of reason is non-repudiable in the first person. The skeptic's attempt to disown the authority of reason by denying it is either a genuine statement of belief, or empty words. If the latter, it isn't a serious challenge. If the former, it invokes and relies upon the authority it attempts to deny. A genuine statement of belief of this sort commits the believer to its inferential consequences, and this is precisely what the skeptic is attempting to deny. He is in a position of practical contradiction, where the intent of his thought (to deny the authority of reasoning) is ruled out by the intent of his thinking it (to come to a conclusion which is correct and can so bind us to believe it true). The demands of reason are thus not genuinely repudiable in the first person, although as we will see, Peter Railton thinks they are.

Railton's view that a constitutive argument must, but cannot, provide a self-sufficient response to the normative question explains why he reads Carroll's tale of Achilles and the Tortoise as a paradox. According to Railton, the lesson to be learned is that it is a mistake to see rules of inference in a valid argument as hypothetical premises. Inferences are not best understood as hypothetical propositions because they serve a different role in argument than do premises. In order to avoid the regress of hypothetical propositions with which Achilles is faced, a non-hypothetical reason for belief must be found. Railton argues that constitutive arguments are limited because they fail to provide this kind of reason. Reason's authority remains elusive because it is always possible to question the requirement that one participate in reasoning. Hence, on Railton's view, Carroll's fable generates a paradox.

In the remainder of this section I will outline Railton's view of constitutive arguments and the two objections he raises, along with my responses. In the next section I'll argue more generally against his assumptions about how constitutive arguments must work. As I see it, the
insight of the constitutive arguments is lost when it is employed in the search for a basic belief on which to ground reason's authority. Railton argues that constitutive arguments are limited by this inability to supply a non-hypothetical reason for belief. Limitations, however, are as much a function of what one expects to achieve, as they are a function of what one is able to achieve. Where Railton finds failure, I think he has mistaken the role of the constitutive argument. His view is instructive because what Railton considers a limitation on constitutive arguments is really a limitation of the foundationalist model of justification which he presupposes. If we don't presuppose a foundationalist justification for reasoning we don't run into the regress problem Railton identifies as key to resolving the normative question. But more importantly, we see that the force of the constitutive argument lies in an implicit appeal to the publicity of reasons. Or so I shall argue.

According to Railton, agents deliberate about what to believe and make claims on the strength of those beliefs. Because we have beliefs, and beliefs imply doxastic commitments, we (agents) are governed by epistemic norms:

as an agent you must possess beliefs; as a believer you must represent certain of your propositional attitudes as accountable to truth and as disciplined by truth-orientated norms (at least, in the limit); therefore, as an agent you must so represent at least some of your attitudes, irrespective of what other goals this might or might not serve.  

So far, Railton is presenting the authority of reasons as constitutive of the practice of

---

reasoning. On Railton's version of the constitutive argument, the agent's commitment to rational standards is of a piece with the business of reasoning. He puts the point in terms of the agent's self-representation as a truth-seeker, not in terms of the agent's actual or attempted conformity with epistemic norms. Accordingly, he states that a "self-representation of certain of one's attitudes as 'aiming at' truth is partially constitutive of belief, which in turn is partially constitutive of agency."15 Here 'self-representation' is important because it makes the relevant constitutive feature depend upon the agent's normative attitude, not his success in producing true or justifiable beliefs. Railton's reply to the normative question is that it is constitutive of agency that agents take their beliefs to be responsive to epistemic constraints. Agents, insofar as they are agents, must understand themselves as engaged in a practice constrained by epistemic norms. But at this point he begins to see problems for the constitutive view.

According to Railton, the purportedly necessary connections which give constitutive arguments their edge are also the source of their problems. He argues that they are either too weak or too strong. Railton puts the point by wondering whether the necessity relation involved is merely linguistic or substantive.16 If linguistic, so that having a belief just means that one takes one's beliefs to be constrained by epistemic norms, then it doesn't really do the work of justifying the practice of being constrained by epistemic norms. It is arbitrary because it fails to incorporate a sense of why beliefs are rightly accountable to truth - not just that they are. To explain the


16 Railton only canvasses the problems for the constitutive argument in these terms as it applies to practical reasoning, although he states that the weakness of constitutive arguments is the same in both cases. I am here constructing what I believe the parallel would be in the theoretical case.
authority of reason, the connection needs to be more substantive. However, if it is a "substantively necessary, non-analytic" connection, then it is too strong. Railton provides an example intended to illuminate this point.

If someone has a propositional attitude toward $p$ which involves representing $p$ as true, then we ought to call it a belief. "But suppose further that she sees no relevance to this attitude of admitted evidence against $p$, even evidence she recognizes to be conclusive. When challenged, she is not defensive and produces no elaborate rationale, but simply points out that she is quite indifferent as to whether her attitude toward $p$ is responsive to the truth of $p$." On the strong reading, we can't accuse her of having an irrational belief because, according to the constitutive view, she has no belief at all. Railton finds this result unsatisfactory. The person in question might insist that her attitude towards $p$ is a belief, and in response we could only search for another label ('supposition' perhaps). This possibility, he suggests, shows that the strong reading strays from an explanation of the authority of reason to a position from which it is unable to criticize badly formed judgements.17

Railton's objection to the strong reading is incoherent. On the constitutive account of agency argued for here the person in Railton's example would not have a belief. If she sees no relevance to the attitude of admitted evidence against it, her propositional attitude wouldn't even

17 Railton, "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," 70.

18 It is not clear why Railton takes this to be a problem for the constitutive view. Constitutive arguments need not serve as criticisms of badly formed judgements in order to explain the authority of reasons. Consider an analogy. I may tell you that your soup is not a chowder because it isn't fish-based. The quality of your soup is beside the point. In order to demonstrate that your soup isn't a chowder I don't need to be able to show you that it is bad soup.
constitute dogmatism. Railton simply asserts that her propositional attitude involves representing $p$ as true, but we can give no sense to the claim that this is so if her epistemic attitude is unresponsive to the truth of $p$. Her insistence that it is a belief carries no weight. Yet, we might suppose that Railton’s example is intended to make a more subtle point. On this supposition, he proposes that a person’s belief producing mechanisms may be in place, and responsive to the truth, without that person concerning herself with the functioning of these mechanisms. Her belief producing mechanisms operate anonymously, and so her feeling of indifference towards them is irrelevant. Here the relevant thought would be that responsiveness to the truth is a subconscious psychological operation and so one’s indifference to this operation is a mere sentiment without epistemic consequence. Even this reading of Railton’s account doesn’t undermine the constitutive view. The constitutive view need not suppose that a disposition to be accountable to truth begets an appropriate sentiment, particularly if that sentiment plays no part in responsiveness. Railton seems to be thinking of ‘responsiveness to truth’ as a purely causal process which might not be reflected in consciousness. But if this process is to qualify the agent’s state as a belief, it must nonetheless serve as a normative (epistemic) commitment. This is why he thinks that indifference poses a problem for the constitutive view. Railton’s conception of ‘responsiveness to truth’ as both merely causal and yet functionally normative is a clear instance of the myth of the given in the form of a mongrel concept.

Railton’s second objection to constitutive arguments makes a similar point at a more general level. According to Railton, even at one level of remove, the constitutive argument fails. A higher-order constitutive argument might claim that a being who fails generally to form propositional attitudes which she takes to be responsive to evidence fails to possess beliefs and
so fails to be an agent. There is surely something to this idea. We may deny that all or most non-human animals have beliefs on the grounds that the practices surrounding the forms of their behaviour which could count as propositional attitudes give us no reason to suppose that they are accountable to the truth. If a being is such that it fails completely to respond to evidence challenging the correctness of its purported belief, then we can reasonably conclude that it is not a believer. This seems to fit with what we do. We tend to treat beings who demonstrate accountability for their beliefs and actions differently than we treat those who do not.

But Railton is uncomfortable with this. The more general form of constitutive argument fails, he claims, because it doesn’t explain the authority of the relevant deontic constraints. Our tendencies to categorize things into agents and non-agents is a fact about what we are like, not a justification of it: “If we rely on these ascending constitutive arguments, we quickly reach a point in which the only thing left to say of someone is to dismiss him as not one of us. This is xenophobia, not criticism.”

Railton holds that even though we may distinguish between believers and non-believers in this way, it is not adequate to the task at hand. An explanation of the authority of reason is not furnished by pointing to the power of agents to assert of non-agents ‘you aren’t one of us unless you manifest feature x.’ Non-agents (those unable to participate in the game of giving and asking for reasons) haven’t thereby been provided with a reason to adopt the epistemic values that go along with agency. It is always open to the agent, Railton suggests, to question why he must be or remain an agent. As we shall see, this is confused. By asking why he must remain an agent, the questioner betrays the fact that he is an agent. Of course, one

---

cannot provide non-agents with reasons. If you are talking to an agent and he refuses to listen to your reasons, you have not necessarily failed, since success at providing someone with a reason does not require that she take it up, but merely that what you provide is something which she can or ought to take up.

According to Railton, constitutive arguments only go so far and then turn into discussions of whether or not there exists a non-hypothetical reason for being an agent. Because they fail to provide a non-hypothetical reason for remaining an agent, constitutive arguments fail to explain how any agent is bound by rational norms. Railton offers two examples of cases in which it seems reasonable to opt out of agency as support for his position. First, one may reasonably give up one's agency through euthanasia in order to avoid a long, painful and imminent death. Second, one may reasonably give up one's agency temporarily in order to avoid complicity in a serious harm. For example, one might render oneself unconscious in order to avoid revealing important information under torture. He claims that cases such as these show that agency itself doesn't have a non-hypothetical justification; this marks the limitation of constitutive arguments.

Railton makes a similar error in this more general case. In order to succeed, constitutive arguments need not keep us from committing suicide. They are supposed to show us how rational standards bear on agency, not how rational standards demand continued agency. Agents who find there to be good reasons to give up agency are, in the doing of it, of necessity agents. An agent's decision to euthanise herself entails that the very feature which made that decision possible will no longer exist (because the agent making the decision will not). This, however, doesn't serve as a refutation of the constitutive argument. The constitutive point is just that it is because you are an agent that you are bound by standards of rationality, not because you are
bound by standards of rationality that you are bound to be an agent.

We have come some distance from Achilles and the Tortoise, and from the suggestion that reason's justification is internal to the practice of reasoning. Railton agrees that reason gets a normative grip on agents because they engage in it to form beliefs and intentions. But he denies that this serves as a sufficient explanation. The constitutive view is limited because it cannot provide a non-hypothetical reason for belief or agency. He is right, of course, that it does not provide a non-hypothetical ground for reason's authority which is in principle unquestionable, if by unquestionable one means an unavoidable (causally necessary) disposition to agree. Our dispositions to live up to the demands of reason are notoriously limited, and partly as a result, the normative question can be asked. But why presume that an answer to the normative question requires a non-hypothetical, regress-stopping answer of this sort? As we have seen, the demand that a normative commitment also be an unavoidable disposition is a mongrel concept motivated by the myth of the given. Railton attempted to pack the norms of rationality into the norms of agency, so that a commitment to the norms of rationality came as a consequence of meeting the conditions of agency. On his view, the constitutive strategy backfires because it is possible for an agent to question the rationality of remaining an agent. If we don't assume that a foundationalist answer is required, we will find that Railton's argument is without force.

In the next section I turn to a discussion of the proper role of constitutive arguments. I will argue that what Railton considers a limitation on constitutive arguments is an artifact of his foundationalist presuppositions. On the condition that the intersubjective validity of reasons is taken to be a part of what constitutes them, a constitutive argument will show that our beliefs and intentions have normative authority because of the public conditions for reasoning.
Constitutive Arguments and the Myth of the Given

The insight of the constitutive argument – that we are bound by the authority of reason because we are agents – was undermined, according to Railton, because it failed to provide a non-hypothetical reason for being guided by standards of rational belief formation. The insight need not be lost however. Railton has fallen prey to the myth of given, not because he argues for a particular feature of belief or agency that will serve as a ground for reason’s authority, but because he thinks an explanation of reason’s authority requires some such thing. First he adopts the persuasive model of justification. This is evident in the strategy he adopts in searching for an answer to the normative question. He attempts to provide the skeptic with a reason why he must accept reason’s authority, but finds that the skeptic (like Carroll’s belligerent Tortoise) always has room to refuse the authority of any particular epistemic norm. It is one thing to prove to the skeptic that he is committed to rational norms in the sense of persuading him to repudiate his skepticism, and another to show that insofar as the skeptic engages in reasoning with us, he is committed to rational norms whether or not he likes it, or admits it.\(^\text{20}\) I have already argued against the former approach. Railton has confused the conceptual necessity of commitment with ‘forcing the skeptic to cry uncle’.

Second, while the promise of the constitutive argument lies in its appeal to the practice of reasoning – that is, in its appeal to the features reasoning must have if it is to be reasoning –

Railton’s search for a non-hypothetical reason for the adoption of rational standards sidetracks him, leading him to search for something which is not internal to the practice of reasoning. We see this in his attempt to stop the regress of justification with a curious normative fact – a fact which every agent must of necessity acknowledge. By putting it this way, we can see that Railton’s search for a foundation leads him away from a discussion of the practice of reasoning to a (private mental) state to which the agent must acquiesce. But a constitutive argument is not an argument of causal necessity. If the authority of reasoning is bound up in the practice of reasoning, then it is a mistake to turn away from the practice of reasoning to search for an indubitable private mental state. Undeniable facts about agency can serve as a foundation for theoretical reasoning no better than they serve as a foundation for practical reasoning. Railton thus implicitly identifies the essential feature of reasoning with a fact about the particular instantiation of that reasoning in the psychology of the agent (which is presumably a causal process).

We are in familiar territory. The attempts to provide a ground for reason, which we have identified as forms of the myth of the given, run together causal and normative orders of explanation in order to explain reason’s authority. These mongrel concepts function as justificatory regress stoppers. The foundation which they purport to provide must both be simply a brute (undeniable) fact and supply a justification for belief. Because the foundationalist holds that infinite chains of justification are unacceptable, and that circular reasoning is unacceptable (the belief systems of rational agents aren’t circular), he needs to have some beliefs be non-
inferentially justified (known directly). Basic beliefs are not themselves capable of further justification, they are non-inferentially justified. Once this is established, other beliefs can be justified inferentially in reference to these basic, non-inferentially justified beliefs. Railton assumes that for the constitutive argument to work, it must provide a foundation for reason of this sort. In one sense, Railton avoids the myth of the given because he argues that constitutive arguments fail to provide a suitably non-hypothetical ground for reason. But in another sense, he doesn’t avoid the myth of the given, because he presumes that a theoretical postulate, in the form of the given, is what is needed. He finds no justification for theoretical reasons because he cannot find an immediate and indubitable feature of agency which compels assent to the standards of rationality.

Constitutive arguments will invoke the myth of the given if they presuppose foundationalism about justification. In the last chapter I argued that Korsgaard’s reflective endorsement is guilty of a looks/is mistake and thus guilty of the myth of the given because on it the authority of reasons is built out of the psychological appearance of reasons. Reflective endorsement turns out to be a mongrel concept which conflates normative and description forms of explanation. The authority of reasoning is derived from the act of giving oneself of reason. The problem with this analysis is that the fact that we are able to give ourselves reasons – to make judgements normative – depends, not on heroic singular acts of willing, but on the system of inferential relations which is already in place given our position in the space of reasons. Public

reasons must be, conceptually speaking, prior to private reasons. Reasons are constitutively public in the sense that when someone offers a reason the question of its authority is a public matter. This is a point which Korsgaard acknowledges but, I argued, given her psychologized account of reflective endorsement, lacks the means to realize.

We are now well-placed to see how it is that constitutive arguments invoke the myth of the given if a foundationalist model of justification is presupposed. Constitutive arguments which presuppose foundationalism will invoke the myth of the given in their attempt to find a feature of reason or agency which itself justifies reason. In Railton's case this turned out to be a non-hypothetical reason for remaining an agent. A foundationalist interpretation of the constitutive argument is bound to fail. A non-inferentially authoritative brute fact cannot serve as a *reason* for accepting rational standards. In order for it to serve as a reason, it would have to function within a public network of inferential relations. But if it were to serve as a *reason* in this way, then its authority wouldn't be merely non-inferentially given and so would presuppose what it sets out to show.

Let's return to Railton's xenophobia objection to the constitutive argument for clarification of this point. According to Railton, higher order constitutive arguments hold that a being who fails generally to form propositional attitudes responsive to evidence fails to have beliefs and thus agency. These arguments fail because the denial of agency to such beings is nothing more than xenophobia. Railton means that it is insufficient for the purposes of explaining the authority of reason, to charge that a certain being, or class of beings, fails to meet our standards of rationality. Who is to say whose standards should prevail? It won't serve as a *criticism* of their standards to say that they fail to meet our standards, nor as a justification of our
own. In other words, it would be question-begging against their standards. Railton takes this to count as evidence that a non-hypothetical justification for reason is unlikely. There is no categorical reason to accept the standards of rationality.

As we have seen Railton is mistaken. The constitutive argument defends reason’s authority without appealing to the correctness of substantive principles of reasoning. To deny a being status as rational is not simply to charge that the being fails to meet our standards of rationality. To be an agent is (constitutively) to have standards for making choices and epistemic standards for having beliefs. A creature who fails to be responsive to reasons fails to be an agent; such a creature does not have the resources to represent its propositional attitudes as aimed at the truth and so accountable to reasons because it doesn’t have propositional attitudes. To make the distinction between reasoners and non-reasoners we need not presuppose the correctness of our own standards, which we can acknowledge to be fallible. To be a creature who fails to be responsive to rational standards is not the same as to be a creature who fails to be responsive to the standards we accept. Railton’s argument requires that these be the same.

Imagine a creature who has different standards than ours. Since such a creature has standards to which it is responsive, and since standards are constitutively public, this creature’s standards constitute different opinions about which inferences are available or required in the space of reasons. Consequently, it makes no sense to think of it as occupying a different space of reasons. Understood in these terms the situation is not an unusual one. It is the familiar one of meeting someone who has different beliefs than you do. This happens all the time, and we

---

22 This is, of course, not to deny the logical possibility of creatures whose rational standards are as a matter of fact unknowable to us. Science fiction examples abound.
know how to bridge the differences — we present evidence, give arguments, and so on. The creature who had radically different standards would be like Wittgenstein’s example of a talking lion which we would be unable to understand — a nonsense experiment. So, Railton’s suggestion that constitutive arguments implicitly appeal to substantive rational standards is wrong. The commitments constitutive of agency are not like those constitutive of being a 19th century English gentleman, who can talk of savages or inscrutable Orientals as failing to meet his standards of civilized conduct. It is true that sometimes, when we make a constitutive claim, we mean to indicate that we recognize instances of kind $K$, by means of some feature $F$, essential to it. But we need not, in that case, think of $F$ as the ground of something’s being $K$, but only as a way to recognize that certain standards are in play. For example, if you have lost your king’s knight you may use a spool of thread in its place. When I see you move the spool two squares forward and one to the right, I see by the way you move it, that it is your knight. Railton’s xenophobia argument makes no sense in these terms. Either the strange creature (the one with different standards) has standards that are difficult to recognize at first because they seem odd to us, in which case it is a believer and is reasonable (in the categorical sense of being responsive to the requirements of reason) or doesn’t have standards and isn’t the kind of creature which has mental states of the right sort to be an agent.

Railton is correct that the possibilities for convincing skeptical agents are limited, but wrong to conclude that this demonstrates a limitation of constitutive arguments. The constitutive argument doesn’t fall prey to his xenophobia rebuttal if the rational standards in question are

---

23 Wittgenstein, *Philosophical Investigations*, #223. Clearly, I am committed to the view that there could not be in principle incommunicable differences between communities of rational beings.
intersubjective, not subjective. Railton’s complaint was that our rational standards cannot be justified by asserting that membership in our community demands adherence to those standards. But our project wasn’t to justify any particular set of rational standards. An answer to the question ‘why believe what I take myself to have reason to believe?’ doesn’t depend upon, or presuppose, the truth of a particular rational standard. Our project was to explain reason’s normative authority, its power to command, not its particular commandments. We are concerned with the accountability to rational standards constitutive of belief, not with the particulars of the rational standards themselves. Of course we must make use of the standards we have. When, having made an inference, we formulate the standard we have followed explicitly, that standard will summarize what we take to be our reason for belief. That standard could be criticized, we could even be convinced by argument or example to give it up. If we do, we will judge that our initial belief wasn’t justified after all. But that will give us no reason to give up on the project of believing itself – as though we could!

The problem with Railton’s xenophobia argument lies in its foundationalist presuppositions and the tendency of those presuppositions to construe reasons as private mental entities. Once we see this, Railton’s xenophobia argument loses its force. Indeed, if the xenophobia objection worked it would start a regress having as a consequence relativism about rational standards. The foundationalist mistake is to suppose that a feature of belief, intention, or agency, once identified, will serve as an explanation of reason’s authority. This is to take reason’s authority outside of the practice of reasoning and put it into the agent in virtue of a curious normative fact about him. The foundationalist strategy promotes a private conception of reason by proposing that in order for a belief to function as a reason, it has to be accepted by
the agent as authoritative. On such a view the authority of reasons would be subjective, their authority would depend upon a subjective affirmation. Moreover, the authority of reasons would be private in the sense that their authority would not be transferrable, precisely because it would depend upon a subjective affirmation. But as we saw in our discussion of the myth of the given, the demand that authority have this double character is incoherent.

To better see how reasons must be conceptually public let us suppose, for the sake of argument, that contrary to Railton there is a non-hypothetical reason for belief. Let us further suppose that a skeptic such as the Tortoise confronted by indubitability, is compelled to grant that he is bound by norms of reasoning. Even in such a case it isn’t clear that the authority of reason would have been demonstrated. In such a case, a fact about that agent – that he can no longer doubt the truth of a proposition – does the justificatory work. Reason’s authority would not be thereby justified, but at best only justified in this case, at this time, for this agent. Compelled to accept reason’s justification on these grounds, our belligerent Tortoise might continue to complain ‘sure, I am rationally compelled now by reason’s authority, but I may not be tomorrow, nor do I see why anyone else is!’ This is a low grade form of justification at best. Moreover, on such a view, it isn’t open to Achilles to respond that ‘the very reason why you have to accept reason’s authority is the same reason any agent has to accept reason’s authority’. If Achilles, or the foundationalist, appeals to ‘what any agent has reason to accept’ then it is the thought that any agent must be constrained by epistemic norms, and not any agent’s compunction to accept them, that is doing the normative work.

Recall that Achilles is trying to prove to the Tortoise that he is rationally required to accept the argument’s conclusion. It wouldn’t count as an explanation of the justification of
reason if Achilles tricked the Tortoise into accepting the conclusion with a specious argument. Nor would it count if Achilles slipped the Tortoise a drug which produced the desired effect. The sort of reason which Achilles must offer the Tortoise is one whose form is intersubjective, or public. As doxastic commitments, reasons for belief represent commitments to rational standards which are in principle shareable. So a reason for belief incorporates a rationally constrained intersubjective form of commitment. The form of the reason necessary in order to answer the normative question is thus public, and so even if a foundationalist succeeds in producing an indubitable proposition, it will be the intersubjective form of that proposition (understood as a reason) which does the normative work. In other words, the foundationalist’s ‘private reason’ presupposes the public character of reason, and it is the public character which serves to provide the explanation. If the Tortoise were in some mental state which simply compelled him to assent to the claim ‘reason requires me to infer Z,’ he would not be rationally justified in the claim but merely the helpless puppet of an internal compulsion. It is a confusion to grant that reason’s authority is bound up in its practice and then to propose that confirmation of this authority must be found outside of practice, in the agent’s willingness to acknowledge his implicit commitment to reason’s authority. The confusion is particularly clear when we notice that ‘private reasons’ of this sort either fail to explain adequately how reason is justified, or presuppose the sort of justification they set out to explain.

We may be reflectively aware of our reasons as reasons when we make judgements and form intentions. This reflective awareness is something like noticing that you have taken up a position in the game of giving and asking for reasons. We may identify our reasons as having the inferential roles which they do in virtue of one’s being in that position in the space of reasons.
We represent our propositional attitudes as responsive to reasons and as disciplined by truth-oriented norms. Consider an analogy. You are playing chess and at a certain point in the game you attend to your king’s knight. When you do you will notice which moves are available to that piece. If you are being reflective about those moves (or if you are unsure which moves are permitted) you might rehearse the rule which expresses the moves available (two spaces in any direction and ...). By ‘representing the knight as accountable to that rule’ you are implicitly committed to the rules of the game. You cannot play chess without employing its rules. Of course, you can decide not to play chess, but cannot play chess by holding the knight accountable to a different set of rules. To do so would treat the lump of plastic in your hand as something other than a knight; its being a knight and its being accountable to that rule are just two different ways of saying what kind of moves it can make. Of course you can make up a new game, or decide not to play, but that is different matter. And of course, chess is optional for agents, but reasoning is not.

I have argued that what Railton proposes is a limitation on constitutive arguments is better understood as a limitation on foundationalism of the sort he presupposes. A non-inferentially justified belief cannot establish reason’s authority. On the one hand, if a basic belief serves as a reason for accepting reason’s authority, then it does so in a context in which reasons can be exchanged. Beliefs function to authorize and provide reasons for other beliefs within a public space in which some judgements serve as reasons for other judgements. But in this case, the private, basic belief presupposes what it sets out to explain. On the other hand, if the basic belief doesn’t serve as a reason for action, but merely as a fact, it invokes the myth of the given.

In the next section I will argue that constitutive arguments can explain the authority of
reasons when a public conception of reasons is in place.

The Public Authority of Reason

Having seen that the authority of reasoning is inherent in the practice of reasoning itself we need to ask what this claim entails. We have seen that a private conception of reasons is a dead end. It cannot supply a justification of rational authority and leads to a vicious relativism about rational standards. A public conception of reasons, however, allows for a positive answer to the normative question without invoking the myth of the given. The aim of this section is to explore the consequences of this insight. In order to demonstrate the authority of reason we needed to show that reason itself provides ends, not merely means to ends arrived at by something other than reasoning. We needed to show that there are theoretical equivalents of categorical imperatives – that there are ends of reasoning to which one is rationally bound irrespective of one’s interests or goals. Let us return to Achilles and the Tortoise to see how a public conception of reason’s authority speaks to the normative question.

In order to make any headway with the Tortoise, the enthymematic understanding of the argument must be circumvented. Railton’s approach showed us that it is folly to try to persuade the Tortoise. He will not be satisfied that a non-hypothetical foundation for reasoning exists if the conclusion hangs on his further belief in a conditional (if $A$ and $B$ and $C$ and $D$ are true, $Z$ must be true). He will only balk at accepting the conclusion on the rational authority of that conditional. So Railton’s approach doesn’t work. Instead, we need to build the authority of the conditional into reasoning itself.
According to Robert Brandom, the lesson of Achilles and the Tortoise is that judgements, logical claims included, must be understood in terms of their inferential consequences, their conceptual role in a language game. As Brandom sees it, Carroll's puzzle arises because we tend to think of the formal rules of logic in non-normative terms, and it can be resolved by recognizing that they too must have pragmatic implications. Carroll's puzzle depends on the fact that in a formal logical system, statements are inferentially inert....Rules are needed to give claims, even conditional claims, a normative significance for action. Although particular rules can be traded in for axioms (in the form of conditional claims), one cannot in principle trade in all rules for axioms. Carroll's point is that the significance of claims for what it is correct to do must somehow be secured. Logical claims, like others, must have some normative pragmatic significance.24

We can think of the normative pragmatic significance to which Brandom refers in terms of inferential relations which we authorize, and for which we may be held responsible. Assertions are claims (propositional contents) which one authorizes and thus makes oneself responsible for.

Saying or thinking that things are thus-and-so is undertaking a distinctive kind of inferentially articulated commitment: putting it forward as a fit premise for further inferences, that is, authorizing its use as such a premise, and undertaking responsibility to entitle oneself to that commitment, to vindicate one's authority, under suitable circumstances, paradigmatically by exhibiting it as the conclusion of an inference from other such commitments to which one is or can become entitled.25

24 Brandom, Making it Explicit, 22.

By looking at things this way, we can avoid the tendency to think of reasons as events occurring inside the heads of reasoners. And still, we can make sense of the space of reasons as a logical space, by understanding the authority and responsibility involved in assertion as inferentially constrained. Stands taken in the space of reasons can be monitored along two normative dimensions -- commitment and entitlement. In order to understand theoretical judgements as propositionally contentful they must have inferential consequences and inferential antecedents. Inferential consequences mark the commitments entailed by the commitment an agent makes in advancing a judgement. Inferential antecedents serve as a base from which an entitlement can be inherited. One is entitled to infer that the apple is coloured, for example, from the propositional content 'the apple is red'. These two sorts of normative content are conceptually connected. If commitment to one judgement ('it is red') precludes commitment to another judgement ('it is not coloured'), those judgements are incompatible. So, by telling me 'the apple is red' you authorize in me that belief -- I am entitled to take it and its implications as true on that basis.

We don't need to go into the details of a theory as complex as Brandom's inferentialism to see the sense of, and advantage in, conceiving of reasons as conceptually public. If we approach theoretical judgements in this way we quickly see that the Tortoise is only living up to one part of his commitment. He authorizes A and B by accepting them, but refuses to take responsibility for doing so. Indeed, the form of his challenge presupposes that his commitment

26 Brandom explains this by identifying three sorts of inferential relations: commitment preserving, entitlement preserving and incompatibility entailments. Articulating Reasons, 194.

27 "In making an assertion, one lends to the asserted content one's authority, licensing others to undertake a corresponding commitment to use as a premise in their reasoning." Brandom, Articulating Reasons, 165.
to \( A \) and \( B \) has no inferential consequences. Unless we assume that assertions carry with them consequential commitments, moves within the game of giving and asking for reasons are not moves at all. They lack the structure required for us to understand them as rational stances. In short, they cannot be made sense of as contentful assertions. Unless we grant this, we cannot make sense of what is going on between Achilles and the Tortoise. That they are playing a language game with a rational structure is clear in the Tortoise’s insistence that Achilles live up to the demands of rationality.

The Tortoise’s commitment to the conclusion is necessitated by the premises which he has already endorsed. He is responsible for the conclusion, because it is entailed by the premises to which he is committed. We might say that he fails to be responsible for his earlier commitments by failing to sincerely engage with Achilles. It is a mistake to suppose that in order to be rationally bound by the conclusion of a piece of reasoning, one has to undergo a particular sort of mental episode – an act of accepting. The act of accepting the conclusion of an argument possesses a sort of normative status because it signals the authorization of a conclusion, not because it authorizes it. It was also a mistake to suppose that \textit{modus ponens} rationally bound the Tortoise to the conclusion of the argument. \textit{Modus ponens} here formally signals, or explains, but doesn’t provide, the authority of the inference to the conclusion.

The Tortoise is rationally bound to accept the conclusion of the argument because he has accepted the premises. By accepting the premises he has committed himself to their truth and taken a stand in the space of reasons, a stand identifiable in terms of the inferential commitments
it makes.\textsuperscript{28} The relevant inferential relations in this case commit the Tortoise to the conclusion. But they don’t commit him to the conclusion because \textit{modus ponens} is a suppressed premise in the argument. Nor do they commit him because he has, through a private mental act, granted the truth of the conclusion. What he is willing to grant is not at issue. Commitment is not, in this way, a first-person notion. Reasons aren’t private mental property. Simply put, his commitment to the conclusion arises out of the fact that someone in his position, or even someone who overheard his assertions, would be entitled to draw the inference and accept the conclusion on the strength of his assertion. In this sense the authority of reason is public because it is publicly transferrable.

If we say to the Tortoise ‘given \(A\) and \(B\), you ought to accept \(Z\),’ rather than ‘given your commitment to \(A\) and \(B\) you are already committed to \(Z\),’ we are not offering him a motive to accept the conclusion, wrapped up in an imperative. We use the imperative to indicate a commitment to be made, and to be made as a result of the implications manifest in the content which is communicated. This is not the same as ordering, or persuading, or forcing it to be made – we may be completely indifferent about whether he will make it.

If we think of judgements as conceptual commitments, and unpack their content in a conceptually public terms, then we are on our way to explaining the normativity of inferential commitments.\textsuperscript{29} But also, we will have adopted a way of thinking about reasoning which

\textsuperscript{28} I am here assuming that the Tortoise doesn’t simply accept the premises for the sake of the argument, but that he believes them to be true. However, even if he only accepted them for the sake of the argument, he would still be undertaking normative commitments, and so while the case would be more complicated, it would not be essentially different.

\textsuperscript{29} Brandom explains the “social articulation” of the commitments involved in judgements by employing a score-keeping metaphor. Keeping score of inferentially articulated commitments
individuates acts of reasoning (theoretical and practical) by the shape of their commitments. Beliefs will have a different set of commitments than intentions, beliefs implying other beliefs, intentions implying other intentions. The question occupying so much of 20th century moral philosophy (how moral judgements can be both motivating and truth-apt) seems less perplexing once practical and theoretical judgements are understood in terms of the commitments they entail. If beliefs are not understood as inert states of the agent, then the assumptions of the instrumentalist starting point will no longer seem attractive. The Tortoise has reason to accept the conclusion of the argument and we need not, indeed (if we have learned anything from Achilles) we ought not, presume that he needs a motivating force in order to really have a reason to make the inference.

In this chapter I have presented a constitutive account of agency and gone on to consider some objections to it. As I have shown, the problem with the objections is that if reasons are not constitutive of agency they must be lodged in private mental states which exemplify the mongrel concept of the myth of the given. The authority of reasons isn’t the authority of some indubitable propositional content. The way in which reasons are constitutive of agency requires them to be conceptually public. In the next chapter I return to a discussion of the normative question as it applies to practical reasoning. I will argue that a public conception of the authority of reasons provides the means to explain the objectivity of reasons without looking to a foundation outside of the practice of reasoning.

and entitlements conceptually is clearly a social notion. Articulating Reasons, 165.
Chapter Six: Intersubjectivism

Summary Of The Argument Thus Far

Let me begin with a summary of the argument thus far. I have claimed that constitutive arguments provide an explanation of how we are bound by norms of rationality. Rational norms are constitutive of reasoning itself. You are not bound by reason because I can give you a particular reason to which you must acquiesce. Indeed, to assume that there could be such a justifying reason is to presuppose that reasons already have the capacity to obligate agents; it is to suppose that in order to show an agent that she is bound by rational norms we must find the right reason and confront her with it. But this, I have argued, gets things the wrong way around. An approach like Railton’s assumes that in order for there to be an answer to the normative question, there has to be a reason whose propositional content alone has the power to obligate (in the form of an undeniable, regress-stopping principle or fact). However, as we saw, facts per se are not the right kind of thing to obligate us. The idea that a propositional content could justify the normative authority of reason by simply describing it is an instance of the myth of the given.

In the case of practical reasoning, I argued that Korsgaard’s position was an advance on the instrumentalist’s. By understanding practice as the source of normativity, Korsgaard avoids
the instrumentalist's mistake—she avoids pursuing a curious normative fact as a means to answering the normative question. Because the instrumentalist is committed to the idea that reasoning about action is theoretical reasoning in the service of preference satisfaction, he is unable to help himself to a practical account of the normative authority of practical reason. If reason's authority is a function of the practice of reasoning, the gap between theory and practice cannot be bridged without abandoning either the conception of theoretical reason as inert (only indirectly able to constrain action) or the thought that practical reasoning is theoretical reasoning in the service of preference satisfaction. In either case, the problem for the instrumentalist is clear. In order to avoid the myth of the given an explanation of how it is that practical reasons bind us must possess an irreducibly normative dimension. Korsgaard's approach is an advance over the instrumentalist's because it holds some promise for avoiding the myth of the given. By turning to practice as the source of authority, she appears to reject a foundationalist model of justification. She denies that the authority of reasons lies in curious normative facts which we can grasp if we deliberate long and hard enough. Nonetheless, I argued, she is guilty of exploiting the myth of the given at a more abstract level.

Korsgaard argued that practical reasons bind us because practical reasoning is inherently normative. We give ourselves reasons, and in doing so confer authority upon intentions or rules of conduct. Thus, we are the source of the normative authority of reasons. This seems like a move in the right direction. The question remains, however, in just what sense we are the source of normative authority—what feature of our agency or practice supplies normative authority. On this point, Korsgaard's position is ambiguous between a private and public conception of reasons. At times she seems to suggest that it is our tendency to maintain a unified sense of self which
provides the normative authority of reasons – we are bound to act on our reasons because unless we do so we will transgress important parts of our personal identity; we will no longer be who we claim to be. So it seems that our tendency toward a unified self-conception is operating as the basis for the normative authority of reasons. On this view, it is a psychological fact about us that we are bound by our reasons. Or more precisely, we are bound by our reasons because, as agents, we feel the psychological force of our practical commitments. Because, for Korsgaard, reasons are a product of reflective endorsement while obligation is a product of reflective rejection, having a reason and being bound by that reason come apart.

The problem with this private account of reason and obligation is that according to it, the normative authority of reason is just the tendency of agents to feel obligated. Here the given appears as a psychological state of the agent, which is immediately given in her experience and yet normatively charged. Put in familiar terms, the problem is that the feeling of obligation is either unanalyzable, in which case it is a brute fact about the agent (a fact about how she feels) or it is analyzable. If it is unanalyzable, its normative powers remain a mystery. If it is analyzable, its normative significance depends not on the feeling state per se, but will have been informed by considerations of the propriety of that response which signal its significance. In other words, in order to make sense of the feeling, or appearance, of obligation as one which signifies the existence of an obligation, a background story explaining how we distinguish between appropriate and inappropriate feelings of obligation must be already in place. Agents may feel obligated when they are not. For example, a deferential housewife may feel obligated to forego some basic

---

1 That she elsewhere explicitly argues against the claim that reasons are contingent on the presence of motivations in this way, gives us reason to suppose that the psychological force in play is normatively charged. See her “Skepticism about Practical Reason.”
personal good so that her spouse will benefit by acquiring a non-basic personal good, something frivolous. It seems just as likely that she has a poorly developed sense of obligation as it does that her feeling of obligation signals a reason. Of course, she may feel that she has a reason, but the question of whether or not she does will not be settled merely by examining how she feels. When we challenge her claim to be obligated we are not disputing how she feels, but rather the reason indicating function of that emotional response in these circumstances.

Other times Korsgaard comes closer to avoiding the myth of the given. Her suggestion that the categorical imperative, understood as a purely formal principle of practical reason, is the source of the authority of practical reasons might seem promising. On this interpretation of her position, the practical feature of agency through which reasons are made normative is reflective endorsement, the will's self-legislation. The trouble with her position is that it attempts to establish a public conception of reasons out of an essentially first-person model of reasons. Korsgaard notes that a public conception of reasons is needed for her argument to work, but lacks the means to build an account of public reasons out of what, I argued, is an essentially private conception of them. One of the many virtues of Korsgaard's account is that it attempts to make sense of much of the phenomenology of forming and even failing to live up to reasons. But the focus on the first person, and the search for the justificatory regress-stopper in the form of a mental state, lead her astray. Korsgaard's constitutive argument depends upon a public conception of reasons — one which goes beyond the limits of her own account of reasons — already being in place. Public reasons must be prior to the private sorts of reasons she describes. Her constitutive argument points us in the right direction but misses its mark because it supposes that the answer to the question of the justification of practical reasons will be found in a
conceptually private reason.

In the last chapter we saw that we can make sense of the idea that the normative authority of reasons is constitutive of agency when reasons themselves are understood as public transactions, not as private mental events, or psychological states. We also saw that the rational commitments implicit in the public exchange of reasons go beyond what their own (private) psychological states authorize, to include the inferential moves authorized by the agent's commitment to a propositional content. Some of the commitments we acquire in practical reasoning are related to the judgements for which we have explicitly taken responsibility. Others are not. We saw that by endorsing $A$ and $B$, the Tortoise committed himself to $Z$. He was rationally bound to accept $Z$ because he had already accepted $A$ and $B$. But it was not simply because he verbally or psychologically assented to $A$ and $B$ that he was committed to the conclusion. Our rational commitments aren't reducible to consistency across those beliefs to which one has already assented. After all, the conclusion is authorized even for those agents who have not explicitly considered that syllogism. In the case of practical reasoning we must be careful to avoid the parallel mistake: the idea that verbal assent to propositional contents, or comparable psychological acts, delimit the scope of our rational commitments. The question of how to delimit the scope of our practically rational commitments is the subject of the remainder of this chapter.

Making Room For Objectivity

There are really two modes in which the normative question may be posed. First, how
do reasons bind at all? I have already argued that their power to bind is constitutive of the practice of reasoning. There is no single, foundational reason to take reasons seriously. On pain of a destructive justificatory regress there couldn’t be. Because we are in the business of reasons they have authority whether we grant it or not. Second, we can go on to ask how are we bound by particular reasons, reasons of which we might not be aware, or willing to accept were we aware of them. How can we make sense of the idea that substantive practical reasons have normative force for us? I attend to this second question here, keeping in mind that my answer must account for reasons without looking to a foundation outside of the practice of reasoning to do the work. The account of intersubjectivity I will be outlining allows for an explanation of the justification of reasons for action by leaving room for error in practical judgement and by making room for an account of getting things right which remains practical, not theoretical.

Intersubjectivism about reasons is one approach to their justification. Here, I am not providing a detailed account of the intersubjective validity of practical norms. I am not making substantive claims about which reasons are good ones. Rather, I am trying to make room for such an account by showing, albeit in general terms, that intersubjectivism has the means to explain how rational constraints on action can be justified. On my view, what we have reason to do is a function of what good practical discourse would vindicate. This doesn’t mean that for someone to have a reason there must have been a public discourse about her reasoning. Nor does it mean that an agent must be, as a matter of fact, answerable to others for all her reasons. Indeed, some of our reasons may be best thought of as personal standards or choices which it would not be legitimate for others to hold us to. For example, my reason for pursuing a career in philosophy, or reading all of Tolstoy’s novels, is not public in the sense that others are entitled
to demand of me that I do so. My reason is public in the sense that it is a commitment, the consequences of which are independent of my endorsement of it.

Intersubjectivism provides an admittedly open-ended account of the intersubjective validity of practical norms. It doesn't yield a single substantive, or even formal, principle. It keeps practical reasoning practical by conceiving of reasons as defensible positions within a social space, but nonetheless provides the means for a normatively robust account of reasons (one which outruns 'what we take ourselves to have reason to do') by insisting that the social space in which reasons are constructed is one in which commitments to objective practice are implied. In the remainder of this chapter, I will show how my account of public reasons has the means to answer the normative question without recourse to the myth of the given. To make the case that a public, or intersubjective, account of reasons is the right one, I must show that intersubjectivism about reasons has the means to distinguish between having reasons and merely appearing to have them. We need to know, in other words, how intersubjectivism about reasons avoids the myth of the given and still explains how reasons authorize action without turning reasons into psychological dispositions and without construing them as mysteriously normative facts.

In order to do justice to our experience of reasons, intersubjectivism must allow that we are sometimes mistaken about what we have reason to do. Our reasons for action extend beyond the set of our actual intentions. Clearly, we can be mistaken about our reasons in a variety of ways. We get things wrong by making bad judgements, employing faulty standards, being misinformed or overcome by our passions. Reasons outrun our individual perspectives and desires. The normativity of reasons thus couldn't be a product of the fact of their endorsement. Consider the assertion: 'I would be crazy to buy a lottery ticket. If I won the lottery, I'd be
kicked off the dole!' The speaker has given herself a reason - she has made herself responsible for a claim - and so we are entitled to predict that she will not buy a lottery ticket and entitled to challenge her claim. She does have a reason insofar as that term picks out a form of commitment. But her reason is a bad one. She has, other things being equal, mistakenly concluded that she has a reason not to buy a lottery ticket in order to preserve her financial security. Whether or not she has a good reason to buy a ticket is neither settled by her speech act, nor her act of willing. But it also seems clear that reasons must have some sort of normative grip on us to count as reasons for action. When we hold that an agent has a reason say, to drink only one beer with dinner, we are suggesting that were that agent to act rationally, she would drink only one beer with dinner.

We have already considered a number of attempts to reconcile these two demands and found them wanting. We saw that realists like Smith focus on the truth-aptness of reasons at the expense of explaining their normativity. A more subjectivist approach (such as Korsgaard's) is so keen to explain their normativity that the question of 'getting things right' drops out of the picture in an important way. Both of these views attempt to build normative commitments into reasoners, albeit in different ways. My strategy instead is to show that the normative commitments are built into reasoning.

The appeal of realism about reasons lies, to a large extent, in the idea that we have reasons to act which surpass both our immediate desires and our immediate awareness of what we ought to do, an idea which, as we have seen, intersubjectivism already satisfies. In important ways, what we have reason to do is not limited by particularities of our psychology or the adequacy of our beliefs. Realism about reasons is tempting because it seems to explain how moral reasons can be right or wrong, good or bad independently of our views about them. It distinguishes between
reasons and what we take ourselves to have reason to do by making the question of what we have reason to do answerable by appeal to a fact, in the form of a definition of rational action. This makes the justification of practical reasons a purely theoretical matter. The problem with realism is that what we have reason to do becomes a fact to which our normative allegiance is mysterious. By trying to give a normatively robust account of reasons, realism sacrifices the connection between agents and reasons. It sacrifices the normative authority which is constitutive of having reasons. Realism cannot explain how reasons bind us because it conceives of reasons as fundamentally separated from our practice. If intersubjectivism about reasons is to be an advance over realism of this sort, it must not sever the connection between reasons and practice so completely. I will argue that the intersubjectivist has the means to explain both the normative authority and the objectivity of reasons. She need not give up on the idea that there are objective moral reasons, or that we can be mistaken about our reasons for action. Because normative authority is built into practice at a basic level, reasons are irreducibly normative. Their normativity is not reducible to some fact about us. Neither can the authority of reasons be reduced to the fact of what we happen to endorse.

We need a distinction between reasons and the mere appearance of reasons, which remains tied to practice, but not tied so closely that what we have reason to do amounts to a matter of fact about what we take ourselves to have reason to do. If we try to get objective reasons out of the reasons we actually endorse, we end up with an unsatisfactory conventionalist account of their justification. In summary then, we need to avoid two unattractive options. First, we need to avoid the idea that what we have reason to do is entirely determined by what we take there to be reason to do (conventionalism about reasons). While such a view has the virtue of
making reasons practical; the reasons such a view has the means to explain fail to do justice to our experience of what reasons are like. Reasons have normative authority beyond what we happen to be persuaded by. If reasons are conceptually intersubjective they will have normative authority beyond what we happen to be persuaded by because they will be public expressions of defensible stands from the start. Their publicity, that they speak to anyone who can listen, is a feature of their independence. The authority of a reason surpasses the fact that a propositional content has been endorsed because the commitments which accrue are there to be inferred independently of whether one is moved by them. The second unattractive option we need to avoid is the idea that what we have reason to do is determinable by theoretical reason alone. We need to maintain a connection between reasons for action and practice which will explain how reasons rationally bind us. What makes something a reason must remain a practical, normative issue, and not the product of a merely theoretical, non-normative fact.

Intersubjectivism

I have used the phrase public conception of reasons to distinguish it from conceptions which make having a reason a mental state which gets its authority from some private feature of the person. I have argued that the latter view is incoherent. By understanding reasons as public and irreducibly normative commitments we see that their capacity to bind agents operates in the activity of reasoning itself. What needs to be shown is how this public, intersubjective account of reasons allows for reasons to be objective. Roughly put, intersubjectivism holds that the standards which constitute rational constraints on action are ones which are in principle
shareable. There are different ways to conceive of how reasons may be shared. One way is to conceive of the shareable reasons in terms of what we, as a matter of fact, accept so that shareable reasons are the ones that we, as a matter of fact, share. This is the mistake of the conventionalist. Instead, we must conceive of shareable reasons as reasons which are shareable in principle, in the sense of being acceptable to rational agents. The intersubjectivist does not look to what we accept to define what is rational, nor does she look to a definition of what is rational in order to explain what is acceptable. For the intersubjectivist, objectivity, as that which is in principle acceptable to rational agents, is a concept which must be approached negatively by ruling out forms of dependence on factors which would render judgements biased. In the practice of reasoning agents attempt to form objective judgements by meeting challenges to the purported objectivity of their own claims.

Reasons have authoritative force because they are commitments across reasoners. The nature and extent of the commitments involved in having particular moral reasons will be determined by practical, not theoretical, reason. Particular reasons serve as moves within a larger network of reasons – a conceptually public practice of justifying and explaining judgements with reasons. Agents cannot have reasons without these reasons having consequences for other judgements – judgements which other agents may have reasons to accept or reject. Let me expand upon the intersubjective conception of reasons by further distinguishing it from the competing conceptions of reasons.

Let us call conventionalism about reasons the view that a reason is justified just in case it is communally agreed upon. Conventionalism in ethics is “the view that acts are judged to be right merely by appeal to the de facto practices that govern them...roughly, an act is right if and
only if most members of the society believe it to be right." According to conventionalism, one has a reason to φ in circumstance C iff φ-ing in C is sanctioned by one's community. Of course, the question of just who constitutes one's community demands an answer, but the details need not worry us here. The conventionalist will have to choose from the usual array of possibilities: one's nation, one's culture, and so on. All that is required is that the conventionalist stick to the thought that it is the fact of agreement (not hypothetical agreement) which provides the justification. On this view, morality is based on patterns of conduct, conventions arrived at through implicit or explicit agreement, which have been endorsed by the community. What we really have reason to do is a function of what members of the community agree to as a matter of fact.

The strongest objection to conventionalism is that it fails to offer a justification of reasons. In virtue of what could it be an account of justification? We don't typically think of justification as a mere consequence of our social practice. Instead, we think of justification as identifying a class of judgements which have the virtue of being connected to the truth, or to right reasoning. Our beliefs aren't made true or justified simply by being believed. Why then, should we think of reasons for action as justified because they accord with social convention? Why think of them as justified because they are taken to be justified?

However, conventionalism can appear appealing, especially when focussing on the differences between judgements of value and judgements of fact. Philosophers have been much more likely to argue for something like conventionalism in the case of moral judgements than they have been in the case of, say, judgements about what there is, although conventionalism is

---

possible in that case as well. It seems more obvious just how we can get claims about the world wrong, than it is obvious how we could get morality wrong. Quite possibly, the appeal of conventionalism grows out of the thought that whatever true moral judgements are like, they will be informed by our psychological make-up in ways that the truth of other sorts of judgements will not. Conventionalism in ethics appears attractive because it may seem that making correct moral judgements and being an objective moral practitioner, are not fixed by human nature or the world in any deep way, but are simply a consequence of our common practices. Conventionalism may seem an attractive justificatory option to some because it appears to them to be the only form of justification available. The rules of morality, like the rules of etiquette, are legitimate rules not because they mark a fundamentally deep feature of human nature, but because they represent social convergence.

Moreover, the conventionalist might add, the justificatory value of the agreement involved in establishing social conventions should not be underestimated when assessing conventionalism's power to justify moral judgements. Agreement, after all, can be thought to be the force which legitimately confers normative authority upon judgements. Moral reasons have the authority to command because we have given them that authority. According to conventionalism, the normative authority of reasons is a consequence of the practice of real moral agents making agreements within a community. For example, once it is established that same sex marriages are permissible, then marriage to a member of the same sex is condoned. No appeal to curious normative facts is required, and yet we seem to come to an explanation of how reasons are made

---

3 Conventionalism about scientific theories promotes the idea that differences in scientific theories simply indicate different conventions for describing the world.
practical. Moral reasons are constructed out of the reasons we give ourselves and others in order to get along in the world. Social conventions mark the establishment of practices which we take to be the right ones. That we are persuaded of their rightness, indicates some measure of their success. If we are persuaded through our moral deliberation and practice that there is reason to \( \phi \) in circumstance \( C \), then \( \phi \)-ing in \( C \) is justified.

While conventionalism has the virtue of connecting the authority of reasons to the public practice of reasoning, it is not an adequate account of their justification. The conventionalist holds that justification remains simply a matter of fact, where the matter of fact is just what there is agreement about. The arguments we have canvassed against the myth of the given should make it clear that the appeal to bare facts about what agents do cannot provide a satisfactory foundation for the authority of reasons. Yet, it may seem that I have tied myself to a conventionalist account of justification by arguing for a public conception of the authority of reasons. It may seem that to hold that reasons are publicly authorized is to hold that our practice of sanctioning particular acts (those consistent with our conventions) just is their justification. This accounting, however, would overlook an important feature of the intersubjective position which I have been advancing.

Justification is not just a matter of fact about our agreements. While what we have reason to do is determined by practice, it is not as the conventionalist supposes. I have already argued that it is constitutive of being an agent that one be open to demands for justification. A commitment to the truth of a claim makes the agent responsible in a public way to respond to the challenges and claims of others. These challenges and responses to them form a system of commitments and entitlements which express the reasons we share. Agreements are important because they express what we take to be good reasons but they are not what makes our reasons
good. This helps us to see why the conventionalist cannot adequately explain how we can be wrong about what we have reason to do.

One problem for conventionalism is that it can’t explain how agents unmoved by reason-producing agreements must be making errors in practical judgement. If conventionalism is true, those who fail to recognize the bindingness of social convention (dissenters in the community) are making rational mistakes. But the conventionalist has no way to explain what makes those who disagree members of the community who err, rather than members of a distinct community which disagrees. Mere social dominance of the majority over the minority cannot explain their error. The conventionalist must hold that you really do have a reason to φ because others have decided that you do. If even a minority agree that you don’t have reason to φ then conventionalism is faced with opposing agreements with no criteria for deciding which form of agreement is binding. Agreements about what we have reason to do make those reasons explicit but do not constitute them. Another problem for conventionalism is that it gets wrong what the disagreement is about. In the same-sex marriage example, the community’s lack of agreement reflects disagreement about whether or not same-sex marriages are permissible. Conventionalism cannot make sense of this disagreement however, because according to it ‘same sex marriages are permissible’ implies that agreement has been established. Conventionalism must understand this disagreement as disagreement about whether an agreement has been reached, not disagreement about what there is reason to do.

The conventionalist might object that he is better able to account for disagreement than I have allowed. The disagreement expressed by dissenters in a community need not be construed as members of a distinct community with their own standards once we make room for
disagreement within the system of practical discourse. On this objection, our shared commitments to moral norms can be consistent with some measure of disagreement about how those norms ought to be realized in practice. In this case the conventionalist does have the means to understand disagreement as disagreement about what there is reason to do. So, for example, agreement about the claim that 'cruelty to non-human animals is morally impermissible' is consistent with the existence of differing standards as to what constitutes cruelty to non-human animals. Strong adherents of animal rights need not be thought of as dissenters who are making rational mistakes, detectable only through an analysis of what members of our society in fact take to be impermissible treatment of animals. In this way the conventionalist can allow for some level of disagreement in his account of what we have moral reason to do. Our ongoing commitment to the badness of cruelty to non-human animals is consistent with disagreement about what amounts to cruelty in any particular case. We may agree that we have a reason to φ without agreeing on all of the details about how to φ. However, while this would make conventionalism about reasons more attractive, because descriptively more sensitive to the possibilities of disagreement within shared agreements, it would not avoid the problem which disagreement poses for the conventionalist. The view could only explain disagreement as disagreement about reasons relative to general agreements already presumed to be rationally justified. It could not offer us any help in understanding how the more general agreements are justified without presupposing an account of normativity beyond the power of conventionalism to supply.

Michael Bratman’s account of shared cooperative activity shows how disagreement can be part of a system of shared intentions and thus might offer the conventionalist the means to advance such an objection. See “Shared Cooperative Activity,” “Shared Intention,” “Shared Intention and Mutual Obligation,” in his Faces of Intention (Cambridge: Cambridge University Press, 1999) 93-108, 109-129, 130-141.
That A has a reason to φ is determined, according to the conventionalist, by the end state of the practice of reasoning. But the simple fact of agreement, without reference to the structures of practical reasoning which make it good, fails to explain how reasons are justified across persons. What makes conventionalism initially seem appealing as an account of shareable reasons is that agreement often signals a form of success in rational discourse. However, what is valuable is not agreement itself, but those forms of rational exchange which, in producing agreement, manifest normative virtues which make the agreement valuable (for example, objectivity, openness to divergent points of view, consistency across beliefs). Our commitment to the virtues of rational discourse informs our judgement that agreement is valuable in the first place. The intersubjectivist is able to embrace the virtues of practical discourse, indeed she must. Having reasons for action implies a commitment to justifying those reasons to others by appealing to other reasons for support. Hence, the commitment to justification enters at the ground level and remains practical.

This also helps us to show how conventionalism fails because it has the consequence that all established social conventions (even those which are repugnant but widely accepted) are justified in virtue of social agreement. Conventionalism has no way to distinguish between agreements which legitimate social practices and those which fail to because they are biased. For example, the social agreement that denied women the right to vote was arrived at by half of the population of reasoners. Of course, at issue was in large part the status of the capacity of women to be counted as reasoners. An agreement arrived at through political discourse which aimed, say, to maintain standards of the then privileged could not serve to justify the judgement that women ought not to be allowed to vote. The question of whether or not women should have the
right to vote isn’t answered by considering the impact of that right on privileged members of the society. Once we see that the real value of agreement lies in the suggestion that through non-biased public discourse we are able to reject bad reasons, conventionalism is no longer a threat.

Nonetheless, I am not arguing that moral agreement, appropriately arrived at, constitutes justification. My aim is simply to show that an intersubjectivist account of the authority of reasons requires justification to be more than what we actually do. Procedures for reaching agreement are themselves normatively constrained. Although agreement often plays an important role in justification it may be impossible to reach. It is possible that agreement about a particular issue may never be reached without calling the procedures for producing that agreement into question.5 Moreover, agents seeking agreement regarding incommensurable goods may be prevented from reaching a consensus without this discord indicating that the public moral discourse has failed in some significant way. We ought to leave open the possibility that moral reasons can conflict. The real point is not, as conventionalism suggests, that it is the fact of reaching agreement that matters. The quality of agreement matters, even where agreement is not in fact reached. The question of the quality of agreement, for example whether it is unbiased, is itself irreducibly normative and so a matter for practical reasoning to determine. To the extent that we want agreement, we want it because reasons are already intersubjective or shareable. In many ways, conventionalism is confused about the importance of agreement.

---

5 Thomas Scanlon perhaps gets around some of these difficulties by arguing that taking a judgement to be true just is believing that it wouldn’t be rejected by people trying to find agreement. He goes on to note that this model gets different results than one which assumes that what we have reason to do is derived from the agreements reached by self-interested maximizers. See his “Contractualism and Utilitarianism,” eds. Amartya Sen and Bernard Williams, *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982) 103-128.
According to intersubjectivism, what we have objective reason to do will be determined intersubjectively through moral discourse. This means that the commitments undertaken in reasoning extend beyond the first person and are not themselves resolvable by an appeal to a special kind of foundational fact, but only through the practice of reasoning itself. What we have a reason to do is a matter for practical, not theoretical reason, to uncover. This is an important point. It commits us to a certain open-endedness in our response to the normative question. Because reasons are in principle shareable, the question of which reasons hold for all will be settled by sorting out those practical commitments to which we are entitled to hold all persons. If you have a reason to φ, it is a reason which others are entitled to hold that you have. Their entitlement will not derive from the fact that there is a consensus (a fact about a relation in which they stand to their community) but will derive from their practical competence as reasoners. The fact that we agree in our use of the term 'red' doesn't make colour conventional. Our entitlement in using the word authoritatively does not derive from a consensus in behaviour but in a shared competence which that consensus reflects.

It may be helpful to consider an example of theoretical reasoning. You have reason to believe that the Earth is much more than 6000 years old, whether or not you are persuaded by my attempts to convince you that it is true. The community of reasoners is entitled to hold you to that reason (they may challenge your claims, point to conceptual incompatibilities, and so on). We are so entitled not because we constitute the majority, or the most vocal part of the population. Rather, we are entitled to hold that you have reason to believe that the Earth is much more than 6000 years old because there are good reasons related to what we all know, to take this proposition to be true. Part of the point is that justification is not just a matter of what the
facts are (a position which both the conventionalist and realist espouse). We need to see that there can be objective reasons to believe which transcend what we take to be reasons for belief, without making them mind-independent. How can I defend the claim that, for example, Stockwell Day has reason to believe the Earth is much more than 6000 years old (even though he claims that the Earth is just 6000 years old) without appealing simply to what we take to be true, or to a realist position in which facts conceal normative demands?

According to intersubjectivism, the question of what there is reason to believe (what reasons are justified) is not settled by an account of the nature of reasoning, but by the practice of reasoning. Substantive issues will have to be resolved by an investigation of the commitment and entitlement relations implicit in judgements we make. What we really have reason to believe and to do will be settled by the relevant forms of inquiry, not by a meta-theory. When Stockwell Day says that the Earth is 6000 years old he makes a move in the space of reasons. His assertion must be reason-sensitive in order to count as a belief. In other words, he is committed (in virtue of his status as a believer) to defending his claim by offering reasons for it and by responding to reasons against it. These reasons will be other beliefs which he holds or rejects and which are implied by or inconsistent with what he believes. The belief that the Earth is much more than 6000 years old is one upon which specialized epistemic practices such as geology and physics have expert claim. By claiming to know that belief to be false, Stockwell Day implicitly claims to repudiate the authority of these practices (presumably on the basis of some expertise of his own) or declares himself an expert in them. In disagreeing with him, I point out his lack of expertise to dispute the vast body of scientific evidence which indicates that the Earth is much older than he claims it is. After pointing to contrary evidence I may go on to show how the denial
of the proposition poses other epistemic problems. Sciences like geology, which cannot be seriously doubted, cannot obviously be reconciled with a denial of the proposition he claims is true. Is he willing to give up geology? If so, on the basis of what competency does he base his willingness? When agents disagree their disagreements are only rationally significant, in the sense of making moves in the space of reasons, when they are attempting to offer each other reasons for their respective beliefs which are acceptable (good) apart from our happening to believe them to be true. The exchange of reasons commits agents to standards for belief and these standards are intrinsically part of the practice of reasoning.

What we are entitled to believe will depend, in part, on what is available to us in the form of the actual body of theory implicit in the systematic beliefs that we (mostly) share. Epistemic expertise will be socially distributed in communities of inquiry so that what non-experts can know on authority will depend upon the rational standards shared by members of those expert communities. Our complaint about Stockwell Day’s belief is likely to come down to a complaint about his lack of credentials as a member of an expert community of knowers upon whose judgement we rely regarding matters such as the age of the Earth. It would be silly to say that prior to the existence of theoretically sound reasons for believing that the Earth is much more than 6000 years old, that we were entitled to do so. So, our entitlements are in this respect contingent upon our place in history. We can acknowledge this without falling back into conventionalism, however, by recognizing that objectivity is not freedom from contingency, but rather a form of practice which aims at getting things right by avoiding acknowledged forms of
bias in judgement. Aiming to make one's beliefs consistent with a particular reading of the Bible may be an appropriate theological pursuit, but in geological discourse it is a mistake. These considerations show that our reason to believe the Earth is much more than 6000 years old cannot simply be the mind-independent fact that the Earth is much more than 6000 years old. Our reason to so believe must be the consequence of unbiased inquiry about the age of the Earth. To say this is not to make a claim about the theory of truth, but to make a point about the nature of justification. If our belief that the Earth is much more than 6000 years old is true, then by any theory of truth it is a fact that it is. 'Is a fact' is not the name of a justification relation.

If intersubjectivism is not a form of conventionalism, neither is it a species of realism. If what we have reason to believe and do is not determined by what we take ourselves to have reason to believe and do then having a reason does transcend practice in a way that the realist can make sense of. Why then isn't intersubjectivism just realism about reasons? To answer this we need to know what we mean by realism, as the term is clearly used in divergent ways. If all it meant were that having a reason transcends actual practice, then any adequate view would be a form of realism. What more does the realist demand of the justification of reasons?

The realist holds that moral reasons are moral facts, and that moral facts can be known through theoretical reason. Michael Smith puts it this way: "moral realism is simply the metaphysical (or ontological) view that there exist moral facts." According to Smith's account, moral reasons are moral facts with motivational consequences. As we have seen, he tries to make

---


theoretical reasons (matters of fact) practical by supplying these motivational consequences through an act of theoretical reasoning. Practical reasons are thus motivationally and rationally authoritative, he argues, because in having beliefs about what one would want were one fully rational, one makes it rational for oneself to have the corresponding desires. Our reasons are thus the products of theoretical reason.

The intersubjectivism I have been describing will not be content with the realist move away from practical reason. In my discussion of his position I showed that the sort of realism that Smith espouses has trouble explaining how we are bound by reasons because it conceives of reasons as theoretical facts, products of theoretical reason. I showed that he is unable to supply the needed normative commitment without moving away from the claim that the rational standard which gives reasons their authority is a function of the practice of reasoning itself. The heuristic 'the hypothetical fully rational believer' really serves as a reason — 'you should do this because if you were fully rational you would.' The authority of reasons thus remains practical and intersubjective. If Korsgaard's persuasive model was too first-personal, the realist's model is too third-personal. It cannot explain, furthermore, how you might have reason to do something you don't even take yourself to have reason to do. Realism fails because it makes the authority of reasons independent of them, so that they are curious normative facts instantiating mongrel concepts.

Other realist accounts claim to be less metaphysically based. Thomas Nagel, for example, denies that moral realism is necessarily a metaphysical thesis. According to Nagel, the realist need only hold (and should only hold) that moral questions are answerable, and that these moral facts
are not reducible to anything else. There are objective reasons which are not mind-independent, but neither are they constructed out of our best attempt to get things right. For Nagel, being able to get things right implies that there is something right that is already there to be found. "I do not assume that the correct answer is just whatever will result from consistent application of deliberative methods – even assuming perfect information about the facts. In deliberation we are trying to arrive at conclusions that are correct in virtue of something independent of our arriving at them." But the justification of reasons through contact with some immediately given feature of experience is a classic form of the myth of the given. We see this clearly in the fact that Nagel's realism leads him to claim that primitive pains and pleasures provide reasons for action from an objective point of view: "primitive pleasures and pains provide at least agent-relative reasons for pursuit and avoidance – reasons that can be affirmed from an objective standpoint and that do not merely describe the actual motivation of the agent." He goes on to argue that the avoidance of suffering has objective value because suffering is bad regardless of whose suffering it is: "if I lacked or lost the conception of myself as distinct from other possible or actual persons, I could still apprehend the badness of pain, immediately." Nagel believes, I think, that his realism is not the objectionable metaphysical sort because the reality of our reasons is immediately present to us in experiences such as suffering. But, of course, the reality of such reasons is not

---


10 Nagel, The View From Nowhere, 158.

11 Nagel, The View From Nowhere, 161.
immediately given. Even reasons to avoid suffering are defeasible (as anyone who has gone through a painful exercise routine recognizes). It is a mistake to disconnect justification from inquiry in this way.

If realism is appropriately shorn of its independence from practice, it begins to resemble intersubjectivism.\(^{12}\) The important point, for my purposes, is to note that intersubjectivism about reasons denies that the question of what we really have reason to do is answerable independently of our practical inquiry into the question. Any position stronger than this risks being a form of the myth of the given. Intersubjectivism respects the thought that there really are reasons – that there are rational constraints on action – while insisting that these reasons be understood as rational entitlements arrived at through the practice of reasoning. It can explain how we are bound by rational norms without employing mongrel concepts.

Agent-Relative and Agent-Neutral Reasons

I have been focussing on an explanation of how intersubjectivism accounts for the justification of reasons. In concluding, I will briefly turn to an explanation of two different categories of reasons (agent-relative and agent-neutral reasons) and a mistaken way of understanding intersubjectivism about reasons. We will see that intersubjectivism does not make

\(^{12}\) Peter Railton offers an account of moral realism in which objectivity in moral discourse is not to be understood as the property ‘mind-independent’, but rather as a sort of ‘opinion-independence.’ See his “Moral Realism: Prospect and Problems,” *Moral Knowledge? New Readings in Moral Epistemology*, eds. W. Sinnott-Armstrong and Mark Timmons (Oxford: Oxford University Press, 1996) 49-81. On the strength of this claim, one might take his view to reduce to a form of intersubjectivism.
the distinction in metaphysical terms. This is important because it helps to undercut the thought that reasons originate in the first-person, as private mental episodes, and that from them interpersonal reasons are constructed.

One means of answering the question of what we have reason to do, is to consider how different sorts of reasons apply to persons. Some reasons seem to get a grip on us through particular features of our agency while others get a grip on us through more generic features of our agency. One way to think of the intersubjectivity of reasons, is to think of is as that part of our moral lives in which we construct objective reasons out of subjective ones. On this view, intersubjective reasons are derived or created through the practice of exchanging subjective reasons. We can all come to have a reason to defend civil liberties, for example, because we value them first for ourselves. Moral discourse allows for the sharing of reasons around that value, and this allows for the construction of an objective value, a value which every person has reason to promote. It is important to see how this way of conceiving of intersubjective reasons is mistaken. In Chapter Four, I argued that although Korsgaard promotes the view that reasons are public, her position has it that public reasons are built out of private ones – that the subjective becomes objective, or intersubjective. Here I want to return to this issue as a means of further clarifying our understanding of intersubjectivism.

The agent-neutral, agent-relative distinction is motivated by the need to make sense of the limits of moral demands. I will argue that although it is motivated by a genuine need, the distinction presupposes that reasons are subjective. Korsgaard attempts to construct intersubjectivity out of subjective reasons through her discussion of this distinction. It is instructive to see how her attempt fails.
Arguably, the moral reasons which bind us must not be so demanding that they make impossible the pursuit of personal projects and continued attention to intimate relations. An account of reasons must respect the difference between reasons which apply to all persons, and those which apply only because an agent possesses a particular property. For example, it makes sense to say that because I am an Akira Kurosawa fan, I have reason to watch *Seven Samurai*. But because you are neither impressed by Kurosawa, nor intrigued by the ancient code of the Japanese warriors, you don’t have a reason to watch it. Kurosawa fans have reasons which non-Kurosawa fans do not. Still, there are reasons for action which do not depend upon the presence of a certain desire. For example, you have reason to get off of my foot, once I tell you that you are standing on it, whether you want to or not. Roughly, agent-neutral reasons are those applying to agents irrespective of their desires and interests. Nagel puts the distinction this way:

If a reason can be given a general form which does not include an essential reference to the person who has it, it is an *agent-neutral* reason. For example, if it is a reason for anyone to do or want something that it would reduce the amount of wretchedness in the world, then that is a neutral reason. If, on the other hand, the general form of a reason does include an essential reference to the person who has it, it is an *agent-relative* reason. For example, if it is a reason for anyone to do or want something that it would be in *his* interest, then that is a relative reason.

Korsgaard argues against the agent-neutral, agent-relative distinction, making a case for intersubjective reasons (“the only reasons that are possible are the reasons we can share”). I will


claim that she is right to argue that the agent-neutral, agent-relative distinction isn’t a good way to conceive of the normative force of reasons, but wrong to argue that objective reasons are constructed out of subjective reasons. According to Korsgaard, we ought to think of morality as the “practical project of bringing our subjective motives into the impersonal point of view and conferring objective normative force or value upon them as far as that can be consistently done. The result would be Intersubjectivism...”15 From an objective point of view we will all value some things, and so these things will have objective (agent-neutral) value. Other things will be valued from a point of view which depends upon subjective features of the agent, so these will have subjective (agent-relative) value. Korsgaard’s approach makes sense of Nagel’s suggestion that we think of ourselves as conferring authority on reasons by forming values. Accordingly, the question of what we have reason to do, what we can legitimately hold each other responsible for, will be answered by sorting out which form of authority it is reasonable to extend in one’s valuations.

But this is a strange way of talking. It suggests that there exist in each of us two sorts of value which we can bestow on states of affairs – objective or subjective – and that once we sort out what we have a right to confer objective value on, we will have figured out what there is objective reason to do. But objective and subjective reasons cannot be part of each person’s make-up from the start. As we have seen, reasons (even agent-relative reasons) are already intersubjective commitments. So, it doesn’t make sense to say that we construct intersubjective reasons out of subjective reasons; intersubjective reasons must already be in place for subjective

---

15 Christine Korsgaard, “The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values,” in her Creating the Kingdom of Ends (Cambridge: Cambridge University Press, 1996) 280.
reasons to exist. Yet according to Korsgaard, intersubjectivism is the view that “objective values are derived or, better, constructed from subjective ones. Our individual, subjective interests become intersubjective values when, because of the attitude we take towards one another, we come to share each other’s ends....neutral reasons are shared, but they are always initially subjective or agent-relative reasons.”\(^{16}\) It is the direction of her argument (from subjectivity to objectivity) that is objectionable.

On Korsgaard’s view, the intersubjectivity of reasons is not really a conceptual position about the nature of reasons, but a norm to which we ought to aspire: “you ought to be committed to the view that another could explain to you what is good about the world as she sees it through the eyes of her ambition.”\(^{17}\) Perhaps, for moral reasons, we ought to be open to the reasons of others as she recommends. But this is an argument for endorsing a particular norm, an argument for being open to reasons, not an argument which establishes that reasons are essentially intersubjective. Korsgaard’s argument doesn’t go deep enough.

The problem is that her position makes a commitment to intersubjectivity a feature of good moral reasoners – a consequence of a moral point of view – instead of building it into reasoning itself. She asserts: “To share another’s ends, or at least to grant they could be shared, is to see them as expressions of that capacity, and so as expressions of our common humanity. The Intersubjectivist sees the other as human, and therefore shares or tries to share the other’s

\(^{16}\) Korsgaard, “The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values,” 278-279.

\(^{17}\) Korsgaard, “The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values,” 290.
ends." But this is surely overstated. It sounds as though the commitments acquired through reason giving and taking are grounded in a recognition of the humanity of persons. The argument gets things the wrong way around. It tries to get rationality, in effect, out of morality. Of course there is something to the thought that seeing others as human is the source of intersubjectivism about reasons. Persons (at least most persons) are the bearers of reasons, and so, in a sense their source. As I have shown, reasons are in principle shareable from the start, they come with implicit commitments. When I give you a reason I am committed to it being a standard which you could share. You may not share it of course, but in seeing you as a person, as a reasoner, I am taking you to be susceptible to reasons, to be 'one of us reasoners.'

The view of intersubjectivism I have developed in this thesis makes the publicity of our commitments as reasoners constitutive of agency itself. Korsgaard’s intersubjectivism is more like a moral principle which we have reason to adopt than an account of the ways in which the commitments made in reasoning are essentially public. As a moral principle Korsgaard’s intersubjectivism is unobjectionable. She is right to suppose that we can develop better judgements and more fully-informed attitudes by being open to the claims which others make on us. We bring our moral perspectives and values under scrutiny when we make claims on ourselves and others. The practical project of discovering which ends we can share is the project of managing amidst a diversity of claims and points of view. We will need to sort out which reasons are reasons for everyone, and which reasons are contingent upon an agent’s possession

---

18 Korsgaard, “The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values,” 290.

19 Of course, this fits with the transcendental argument Korsgaard advances in The Sources of Normativity, but this compatibility makes it no more convincing.
of a desire or ambition. Sorting this out surely involves openness to divergent points of view. We are, after all, aiming at truth not homogeneity. However, none of this is an account of what makes reasons intersubjective.

The confusion in Korsgaard’s account of the intersubjectivity of reasons arises because she doesn’t lodge intersubjectivity deeply enough in the practice of reasoning. What makes reasons intersubjective is not the fact that there are good moral reasons for them to be. The question of which reasons are reason for everyone is not settled by establishing the limits of an agent’s ‘right to confer agent-neutral or objective value on a state of affairs.’ Once we move away from talk of authority being conferred through acts of willing, and remember that having a reason is making a move in the space of reasons, we see that intersubjectivity is built into the reasons and reasoning and not built out of the reasoner. The commitment to be responsive to other’s reasons is part of what it is to have a reason, and not simply a response to valuing humanity in the correct way.

The intersubjectivity of reasons is a conceptual, not a moral feature of reasons. If intersubjectivity were a moral principle it would require a substantive defence. And if subjective reasons provided our starting point, there would be no obvious way to mount that defense. On my version of intersubjectivism, the distinction between agent-relative and agent-neutral reasons does no work to explain the authority which reasons have. The virtue of intersubjectivism lies in the fact that it is able to make sense of the objectivity of reasons without presuming an untenable private account of reasons from the start.

In this chapter I have argued that a public conception of reasons, which I have called intersubjectivism, respects features of rational discourse while leaving room for an account of the
objectivity of reasons. I have also distinguished it from conventionalism, realism and subjectivism. We have seen that intersubjectivism is superior to conventionalism in its ability to make genuine claims to objectivity. We have further seen that realism, to the extent that it extends beyond the sensible bounds of intersubjectivism, makes unwarranted metaphysical claims and falls foul of the myth of the given. And finally, I argued that intersubjectivism cannot be understood as a moral imperative which serves to make subjective reasons objective.
Chapter Seven: Conclusion

In this thesis I have been principally occupied with two related tasks. I have argued that a questionable assumption about the distinction between practical and theoretical reasoning colours the debate about what practical reasons would have to be like in order to be reasons at all. This is the assumption that an explanation of the normative authority of theoretical reasoning is unproblematic, while an explanation of the normative authority of practical reasoning is mysterious, if not impossible. What is at issue is the form of the rational imperatives constraining agents in making judgments about what to believe and do. Because many philosophers have felt that explaining the normative authority of practical reasons would require an appeal to categorical imperatives, the debate has centred around whether these imperatives are possible.

Instrumentalists have argued that only hypothetical imperatives constrain action because their authority is transmitted through ends whose authority is not given by reason. Hypothetical imperatives, supplying means to ends, are unproblematic because they have no authority apart from the ends they serve. Philosophers who have wanted to defend categorical imperatives have thus been put in the difficult position of having to defend the idea of imperatives whose rational authority is independent of the ends agents actually have. This way of framing the distinction
between practical and theoretical reasoning obscures the irreducible normativity involved in theoretical reasoning itself. Once we see that both practical and theoretical reasoning are in the same boat with respect to owing an account of the categorically imperative force of reasons, the instrumentalist cannot simply appeal to the unproblematic character of his conception of theoretical reasoning. I argued that irreducibly normative commitments are a feature of reasons, both theoretical and practical, thereby undermining the suggestion that practical reason bears a special burden in the justification of rational imperatives.

Having shifted the debate from the question of how practice could be rational, to the question of how reasons themselves could be authoritative, my second task was to provide an argument for intersubjectivism about reasons. My method was partly critical and partly constructive. By showing how views which treat theoretical reason as unproblematically authoritative violate the myth of the given, because they suppose that justification could be a private mental state, I supported the thesis that reasons must be public. By showing how views which make the publicity of reasons a product of private reasons also violate the myth of the given, I supported the thesis that reasons must not only be public but must be irreducibly public. This was the first stage in my argument in favour of intersubjectivism.

I then offered a constitutive argument according to which reasons bind us because we are reasoners. The authoritative force of reasons is fundamentally a feature of reasoning, not a feature of ends which reasoners contingently possess. I defended intersubjectivism against the claim that constitutive arguments can only provide hypothetical justification of the authority of reasons by showing this claim to depend upon a foundationalist conception of justification. Finally, I distinguished intersubjectivism from two rival conceptions of the justification of reason.
I argued that conventionalism, by making the bindingness of reasons an artifact of agreement, cannot make sense of error and hence of the objectivity of reasons. Realism, by making the objectivity of reasons a function of facts independent of the practice of reasoning thereby severs the connection between reasons and that practice, and thus cannot explain how reasons are binding. That completed my argument for intersubjectivism. My account of intersubjectivism, if true, shows that there can be no proof of the authority of reasons outside of the practice of reasoning: reasons can only be vindicated in practice.
Bibliography


