
Making It Explicit

Reasoning, Representing,
and Discursive Commitment

Robert B. Brandom

Harvard University Press
Cambridge, Massachusetts
London, England

To Wilfrid Sellars and Richard Rorty

without whom most of it would not even be *implicit*

Copyright © 1994 by the President and Fellows of Harvard College
All rights reserved
Printed in the United States of America
Page xxiv constitutes an extension of the copyright page.
Fourth printing, 2001

First Harvard University Press paperback edition, 1998

Library of Congress Cataloging-in-Publication Data

Brandom, Robert.

Making it explicit : reasoning, representing, and discursive commitment /
Robert B. Brandom.

p. cm.

Includes bibliographical references and index.

ISBN 0-674-54319-X (cloth)

ISBN 0-674-54330-0 (pbk.)

1. Language and languages—Philosophy. 2. Semantics (Philosophy)
3. Pragmatics. 4. Representation (Philosophy) I. Title.

P106.B694 1994

121'.68—dc20

93-50631

. . . both a new world,
and the old made explicit . . .

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

T. S. ELIOT, "Four Quartets"

Contents

Preface xi

PART ONE

1. Toward a Normative Pragmatics	3
I. Introduction	3
II. From Intentional State to Normative Status	7
III. From Norms Explicit in Rules to Norms Implicit in Practices	18
IV. From Normative Status to Normative Attitude	30
V. From Assessment to the Social Institution of Norms	46
VI. From Intentional Interpretation to Original Intentionality	55
<i>Appendix: Wittgenstein's Use of <i>Regel</i></i>	64
2. Toward an Inferential Semantics	67
I. Content and Representation	67
II. The Priority of the Propositional	79
III. Conceptual Classification and Inference	85
IV. Material Inference, Conceptual Content, and Expression	94
V. Circumstances and Consequences of Application	116
VI. Conclusion	132

3. Linguistic Practice and Discursive Commitment	141
I. Intentional States and Linguistic Practices	141
II. Deontic Status and Deontic Attitudes	157
III. Asserting and Inferring	167
IV. Scorekeeping: Pragmatic Significance and Semantic Content	180
4. Perception and Action: The Conferral of Empirical and Practical Conceptual Content	199
I. Assertions as Knowledge Claims	199
II. Reliability	206
III. Observation Reports and Noninferential Authority	213
IV. Rational Agency	229
V. Practical Reasoning: Inferences from Doxastic to Practical Commitments	243
VI. Intentions	253
PART TWO	
5. The Expressive Role of Traditional Semantic Vocabulary: 'True' and 'Refers'	275
I. From Inference to Truth, Reference, and Representation	275
II. Truth in Classical Pragmatism	285
III. From Pragmatism to Prosentences	299
IV. Reference and Anaphorically Indirect Descriptions	305
V. The Function of Traditional Semantic Vocabulary Is Expressive, Not Explanatory	322
6. Substitution: What Are Singular Terms, and Why Are There Any?	334
I. Multivalued Logic and Material Inference	334
II. Substitution, Sentential Embedding, and Semantic Roles	346
III. Subsential Expressions	360
IV. What Are Singular Terms?	367
V. Why Are There Singular Terms?	376
VI. Objections and Replies	384
VII. Conclusion	399
<i>Appendix I: From Substitutional Derivation of Categories to Functional Derivation of Categories</i>	404
<i>Appendix II: Sentence Use Conferring the Status of Singular Terms on Subsential Expressions—An Application</i>	409
7. Anaphora: The Structure of Token Repeatables	413
I. Frege's <i>Grundlagen</i> Account of Picking Out Objects	413
II. Definite Descriptions and Existential Commitments	432
III. Substitution, Token Recurrence, and Anaphora	449
IV. Deixis and Anaphora	459

V. Interpersonal Anaphora and Communication	473
<i>Appendix: Other Kinds of Anaphora—Paychecks, Donkeys, and Quantificational Antecedents</i>	490
8. Ascribing Propositional Attitudes: The Social Route from Reasoning to Representing	495
I. Representation and <i>De Re</i> Ascription of Propositionally Contentful Commitments	495
II. Interpretation, Communication, and <i>De Re</i> Ascriptions	508
III. <i>De Re</i> Ascriptions and the Intentional Explanation of Action	520
IV. From Implicit Attribution to Explicit Ascription	529
V. Epistemically Strong <i>De Re</i> Attitudes: Indexicals, Quasi-Indexicals, and Proper Names	547
VI. The Social-Perspectival Character of Conceptual Contents and the Objectivity of Conceptual Norms	584
<i>Appendix: The Construction and Recursive Interpretation of Iterated Ascriptions That Mix <i>De Dicto</i> and <i>De Re</i> Content Specifications</i>	608
9. Conclusion	614
I. Two Concepts of Concepts	614
II. Norms and Practices	623
III. We Have Met the Norms, and They Are Ours	639
<i>Abbreviations</i>	653
<i>Notes</i>	655
<i>Index</i>	717

PART ONE

Toward a Normative Pragmatics

An ounce of practice is worth a pound of precept.

ENGLISH PROVERB

I. INTRODUCTION

1. *Saying 'We'*

'We' is said in many ways. We may be thee and me. We may be all that talks or all that moves, all that minds or all that matters. Since these boundaries are elastic, we have a task of demarcation: telling who or what we are, distinguishing ourselves from the other sorts of objects or organisms we find in our world. Saying who we are can contract to an empty exercise in self-congratulation—a ritual rehearsal of the endless, pitiable disabilities of clockworks, carrots, cows, and the clan across the river. Such a mean-spirited version of the demarcational enterprise is not forced on us by the way things are, however.

For what we are is made as much as found, decided as well as discovered. The sort of thing we are depends, in part, on what we take ourselves to be. One characteristic way we develop and make ourselves into what we are is by expressing, exploring, and clarifying our understanding of what we are. Arbitrary distinctions of biology, geography, culture, or preference can be and have been seized on to enforce and make intelligible the crucial distinction between us and them (or it). But philosophical thought is coeval with the impulse to understand ourselves according to a more principled, less parochial story—and so to be a more principled, less parochial sort of being.

The wider perspective enjoined by principle poses the question, Who are we? in the form: What would have to be true—not only of the quaint folk across the river, but of chimpanzees, dolphins, gaseous extraterrestrials, or digital computers (things in many ways quite different from the rest of us)—for them nonetheless to be correctly counted among us? Putting the issue this way acknowledges an expansive demarcational commitment to avoid, as far as possible, requiring the sharing of adventitious stigmata of origin or material constitution. In understanding ourselves we should look to conditions at once more abstract and more practical, which concern what we are able to *do*, rather than where we come from or what we are made of. Candidates for recognition as belonging among us should be required to share only the fundamental abilities that make possible participation in those central activities by which we (thereby) define ourselves. How should we think of these?

The most cosmopolitan approach begins with a pluralistic insight. When we ask, Who are we? or What sort of thing are we? the answers can vary without competing. Each one defines a different way of saying 'we'; each kind of 'we'-saying defines a different community, and we find ourselves in many communities. This thought suggests that we think of ourselves in broadest terms as the ones who say 'we'. It points to the one great Community comprising members of all particular communities—the Community of those who say 'we' with and to someone, whether the members of those different particular communities recognize each other or not.

The reflexive character of the proposal that we use self-demarcation as the criterion by which we demarcate ourselves does not suffice to render it purely formal, however. It does not save us the trouble of contentful self-understanding. For until it has been specified in other terms what one must be able to do in order to count as "saying 'we'," demarcation by appeal to such attitudes remains an aspiration tacked to a slogan—empty, waiting for us to fill it. 'We'-saying of the sort that might be of demarcational interest is not a matter merely of the production of certain vocables—indeed perhaps the relevant kind of attitude is not a linguistic matter at all. Nor again does it consist simply in the engendering of warm mammalian fellow-feeling. Making explicit to ourselves who we are requires a theoretical account of what it is in practice to treat another as one of us.

2. *Sapience*

What is it we do that is so special? The answer to be explored here—a traditional one, to be sure—is that we are distinguished by capacities that are broadly cognitive. Our transactions with other things, and with each other, in a special and characteristic sense *mean* something to us, they have a *conceptual content* for us, we *understand* them in one way rather than another. It is this demarcational strategy that underlies the classical iden-

tification of us as *reasonable* beings. Reason is as nothing to the beasts of the field. We are the ones on whom reasons are binding, who are subject to the peculiar force of the better reason.

This force is a species of *normative* force, a rational 'ought'. Being rational is being bound or constrained by these norms, being subject to the authority of reasons. Saying 'we' in this sense is placing ourselves and each other in the space of reasons, by giving and asking for reasons for our attitudes and performances. Adopting this sort of practical stance is taking or treating ourselves as subjects of cognition and action; for attitudes we adopt in response to environing stimuli count as *beliefs* just insofar as they can serve as and stand in need of reasons, and the acts we perform count as *actions* just insofar as it is proper to offer and inquire after reasons for them. Our attitudes and acts exhibit an intelligible content, a content that can be grasped or understood, by being caught up in a web of reasons, by being inferentially articulated. Understanding in this favored sense is a grasp of reasons, mastery of proprieties of theoretical and practical *inference*. To identify ourselves as *rational*—as the ones who live and move and have our being in the space of reasons, and so to whom things can be intelligible—is to seize demarcationally on a capacity that might well be shared by beings quite different from us in provenance and demeanor.

Picking us out by our capacity for reason and understanding expresses a commitment to take *sapience*, rather than *sentience* as the constellation of characteristics that distinguishes us. Sentience is what we share with non-verbal animals such as cats—the capacity to be *aware* in the sense of being *awake*. Sentience, which so far as our understanding yet reaches is an exclusively biological phenomenon, is in turn to be distinguished from the mere reliable differential responsiveness we sentient beings share with artifacts such as thermostats and land mines. Sapience concerns understanding or intelligence, rather than irritability or arousal. One is treating something as sapient insofar as one explains its behavior by attributing to it intentional states such as belief and desire as constituting reasons for that behavior.

Another familiar route to understanding the sort of sapience being considered here for demarcational duty goes through the concept of *truth*, rather than that of *inference*. We are believers, and believing is taking-true. We are agents, and acting is making-true. To be sapient is to have states such as belief, desire, and intention, which are contentful in the sense that the question can appropriately be raised under what circumstances what is believed, desired, or intended would be *true*. Understanding such a content is grasping the conditions necessary and sufficient for its truth.

These two ways of conceiving sapience, in terms of inference and in terms of truth, have as their common explanatory target contents distinguished by their *propositional* form. What we can offer as a reason, what we can take or make true, has a propositional content: a content of the sort that we express by the use of declarative sentences and ascribe by the use of 'that' clauses.

Propositional contents stand in inferential relations, and they have truth conditions. One of the tasks of this work is to explain what it is to grasp specifically propositional contents, and so to explain who we are as rational or sapient beings. A central subsidiary task is accordingly to offer an account of the relation between the concepts of inference and truth, which complement one another and in some measure compete with one another for explanatory priority in addressing the issue of propositional contentfulness, and so of rationality.

3. *Intentionality*

The general self-understanding in view so far identifies us by our broadly cognitive capacities: We are makers and takers of reasons, seekers and speakers of truth. The propositional focus of the approach marks this understanding of intelligible contents as *discursive*. This conception, halloved by ancient tradition, was challenged during the Enlightenment by a rival approach to cognitive contentfulness that centers on the concept of *representation*. Descartes's seminal demarcational story distinguishes us as representers—producers and consumers of representings—from a world of merely represented and representable things. The states and acts characteristic of us are in a special sense *of, about, or directed at* things. They are representings, which is to say that they have representative content. To have such a content is to be liable to assessments of *correctness* of representation, which is a special way of being *answerable* or *responsible* to what is represented.

Another task of this work is accordingly to address the question, How should the relation between representation—the master concept of Enlightenment epistemology—and the discursive concepts of reasons and truth be understood? One of the great strengths representationalist explanatory strategies have developed is the capacity to offer accounts of truth and goodness of inference. There are familiar set-theoretic routes that set out from representational primitives corresponding to subsentential linguistic expressions such as singular terms and predicates, lead to assignments of truth conditions to sentences compounded out of those expressions, and pass from there to determinations of which inferences are correct. While doubts have been raised, perhaps legitimately, about nearly every phase of this construction, no other semantic approach has been worked out so well.

Yet for all that, the primitives involved have never been well understood. Descartes notoriously fails to offer an account either of the nature of representational contents—of what the representingness of representings consists in—or of what it is to grasp or understand such contents, that is to say, of their intelligibility to the representer. He does not tell us what makes a rabbit-idea an idea of (or purporting to be of) rabbits, or of anything at all, nor what it is for the one whose idea it is to understand or take it *as* being of or about something. That things could be represented by and to the mind (have

“objective reality” in it and for it, for the mind to be “as if of” things) is treated as a basic property, an unexplained explainer. But an adequate treatment of the representational dimension of discursive sapience should include an account both of representational *purport*, and of its *uptake*.

The topic to be investigated here, then, is intentionality in the sense of the propositional contentfulness of attitudes, not in the sense (if that should turn out to be different) of the directedness of sense. The aim is to understand ourselves as judges and agents, as concept-users who can reason both theoretically and practically. This is not to say that we should understand ourselves exclusively as sapient rather than sentient, in terms of understanding rather than awareness. ‘We’ is and by rights ought to be said in many ways. The point is just to register and delineate the way that is to be discussed here.

This inquiry is directed at the fanciest sort of intentionality, one that involves expressive capacities that cannot be made sense of apart from participation in linguistic practices. The aim is to offer sufficient conditions for a system of social practices to count as specifically *linguistic* practices, in the sense of defining an idiom that confers recognizably *propositional* contents on expressions, performances, and attitudes suitably caught up in those practices. Looking at this sort of high-grade intentionality accordingly risks being beastly to the beasts—not only by emphasizing sapience over sentience, comprehension over consciousness, but also by unfairly ignoring the sorts of beliefs and desires that are appropriately attributed to non- or pre-linguistic animals.

So it is a further criterion of adequacy of this explanatory enterprise that it have something to say about the lower grades of intentionality: not only as to how the lines should be drawn (corresponding to different senses of ‘we’), but also as to how the advent of the favored sort of linguistic intentionality can be made less mysterious. How can linguistic abilities arise out of nonlinguistic ones? Or to ask a related question, What would sentient creatures have to be able to do in order to count as sapient as well? What is needed is to tell a story about practices that are sufficient to confer propositionally contentful intentional states on those who engage in them, without presupposing such states on the part of the practitioners. The hope is that doing so will offer guidance concerning what would be involved in diagnosing aliens as exhibiting such states, and programming computers or teaching merely sentient animals to exhibit them.

II. FROM INTENTIONAL STATE TO NORMATIVE STATUS

1. *Kant: Demarcation by Norms*

The demarcational proposal being pursued picks us out as the ones capable of judgment and action. Not only do we respond differentially to enviroing stimuli, we respond by forming perceptual judgments. Not

only do we produce behavior, we perform actions. Various ways of talking about this fundamental distinction have been put on the table. It can be made out in terms of *truth*. In perception what we do is responsively *take-true* some propositional content that is intelligible to us. In action what we do is responsively *make-true* some propositional content that is intelligible to us.

The distinction can be made out in terms of *reasons*. The judgments that are our perceptual responses to what is going on around us differ from responses that are not propositionally contentful (and so are not in that sense intelligible) in that they can serve as reasons, as premises from which further conclusions can be drawn. Actions, which alter what is going on around us in response to propositionally contentful intentions, differ from performances that are merely behavior (and so not intelligible in terms of the propositionally contentful intentions that elicit them) in that reasons can be given for them; they can appear as the conclusions of practical inferences.

The distinction can also be made out in terms of the employment of *concepts*. To be a perceiver rather than just an irritable organism is to be disposed to respond reliably and differentially to the perceptible environment by the application of appropriate concepts. To be an agent rather than just a behaver is to be disposed to respond reliably and differentially to applications of appropriate concepts by altering the accessible environment. Intelligibility in the sense of propositional contentfulness, whether the latter is conceived in terms of truth conditions or capacity to serve as a reason, is a matter of conceptual articulation—in the case of perception and action, that the reliably elicited response and the reliably eliciting stimulus, respectively, essentially involve the use of concepts.

So sapience, discursive intentionality, is concept-mongering. What is distinctive of specifically *conceptual* activity? Contemporary thought about the use of concepts owes great debts to Kant. One of his cardinal innovations is his introduction of the idea that conceptually structured activity is distinguished by its *normative* character. His fundamental insight is that judgments and actions are to be understood to begin with in terms of the special way in which we are *responsible* for them.

Kant understands concepts as having the form of *rules*, which is to say that they specify how something *ought* (according to the rule) to be done.¹ The understanding, the conceptual faculty, is the faculty of grasping rules—of appreciating the distinction between correct and incorrect application they determine. What is distinctive about judgments and doings—acts that have contents that one can take or make true and for which the demand for reasons is in order—is the way they are governed by rules. They are conceptually contentful and so are subject to evaluation according to the rules that express those contents. Being in an intentional state or performing an intentional action accordingly has a normative significance. It counts as undertaking (acquiring) an obligation or commitment; the content of the commitment is determined by the rules that are the concepts in terms of which the act or

state is articulated. Thus Kant's version of the sort of demarcation criterion being considered picks us out as distinctively *normative*, or rule-governed, creatures.

2. From Cartesian Certainty to Kantian Necessity

This emphasis on the normative significance of attributions of intentionally contentful states marks a decisive difference between Kantian and Cartesian ways of conceiving cognition and action. For Kant the important line is not that separating the mental and the material as two matter-of-factually different kinds of stuff. It is rather that separating what is subject to certain kinds of normative assessment and what is not. For Descartes, having a mind (grasping intentional contents) is having representings: states that purport or seem to represent something. Some things in the world exhibit this sort of property; others do not. Where Descartes puts forward a *descriptive* conception of intentionality, Kant puts forward a normative, or *prescriptive*, one—what matters is being the subject not of *properties* of a certain kind but of *proprieties* of a certain kind. The key to the conceptual is to be found not by investigating a special sort of mental *substance* that must be manipulated in applying concepts but by investigating the special sort of *authority* one becomes subject to in applying concepts—the way in which conceptually articulated acts are liable to assessments of correctness and incorrectness according to the concepts they involve.

This approach contrasts sharply with Cartesian demarcations of cognition and action according to the presence of items of a certain matter-of-factual kind. The objection is not to the details of Descartes's understanding of the descriptive features required for intentionally contentful states and acts: his conception of mental events that are self-intimating cognitions or infallibly performable volitions, takings-true and makings-true that are minimal in that they cannot fail to be successful. It is, more radically, that what sets off the intentional is its liability to assessments of correctness, its being subject to norms (which are understood as codified in rules), rather than any missing feature that it could be described as having or lacking.

Descartes inaugurated a new philosophical era by conceiving of what he took to be the *ontological* distinction between the mental and the physical in *epistemological* terms: in terms of accessibility to cognition—in terms, ultimately, of certainty. Kant launched a new philosophical epoch by shifting the center of concern from *certainty* to *necessity*. Where Descartes's descriptive conception of intentionality, centering on certainty, picks out as essential our grip on the concepts employed in cognition and action, Kant's normative conception of intentionality, centering on necessity, treats their grip on us as the heart of the matter. The attempt to understand the source, nature, and significance of the norms implicit in our concepts—both those that govern the theoretical employment of concepts in inquiry and knowl-

edge and those that govern their practical employment in deliberation and action—stands at the very center of Kant's philosophical enterprise. The most urgent question for Kant is how to understand the rulishness of concepts, how to understand their authority, bindingness, or validity. It is this normative character that he calls *Notwendigkeit* (necessity).

The nature and significance of the sea change from Cartesian certainty to Kantian necessity will be misunderstood unless it is kept in mind that by 'necessary' Kant means 'in accord with a rule'. It is in this sense that he is entitled to talk about the *natural* necessity whose recognition is implicit in cognitive or theoretical activity, and the *moral* necessity whose recognition is implicit in practical activity, as species of one genus. The key concept of each is obligation by a rule. It is tempting, but misleading, to understand Kant's use of the notion of necessity anachronistically, in terms of contemporary discussions of alethic modality. It is misleading because Kant's concerns are at base normative, in the sense that the fundamental categories are those of deontic modality, of commitment and entitlement, rather than of alethic modality, of necessity and possibility as those terms are used today. Kant's commitment to the primacy of the practical consists in seeing both theoretical and practical consciousness, cognitive and conative activity, in these ultimately normative terms.

So for Kant, concepts are to be understood by the theorist in terms of the *rules* that make them explicit, rules that specify how the concepts are *properly* or *correctly* applied and otherwise employed. Kant's appreciation of this normative significance of concept use is one of the lenses through which he views his relationship to his rationalist and empiricist predecessors. With the wisdom of hindsight, Kant can see a normative strand of concern with *responsibility* as fundamental to the Enlightenment.² Thus the *Meditations* is to be read as motivated by the demand that the meditator take personal responsibility for every claim officially endorsed—be prepared to answer for it, demonstrate entitlement to that commitment by justifying it. This theme remained merely implicit in Descartes's theorizing about us (as opposed to his motivation and methodology), for his explicit theory remains naturalistic (though not, of course, physicalistic).

Leibniz insists, against the empiricists, that inferential transitions between representations ought not to be assimilated to matter-of-factual, habitually acquired causal dispositions. He understands them rather as applications of general principles that must accordingly be available prior to any knowledge of empirical matters of fact. Kant takes over from his reading of Leibniz the general idea of rules as what underwrite cognitive assessments of inferences and judgments. He understands such a priori principles, however, not as very general statements of fact (even metaphysical fact), but as rules of reasoning. They are conceived not as descriptive but as prescriptive—as (in Sellars's phrase) "fraught with ought."

This lesson dovetails neatly with the moral he draws from Hume's

thought. On Kant's reading, Hume's contribution is to see that ordinary empirical discourse involves commitments that reach beyond the sequences of representations, however regular, in which the concepts deployed in that discourse are taken to originate. Kant's Hume recognizes that cognitive experience crucially involves the application and assessment of the correctness of the application of *rules*. For Kant, Hume's inquiry after the nature of the authority for this inferential extension takes the form of a quest for the nature of the necessity, understood as normative bindingness, exhibited by the rules implicit in empirical concepts. It is under this conception that Kant can assimilate Hume's point about the distinction between saying what *happens* (describing a regularity) and saying what is causally *necessary* (prescribing a rule) to his point about the distinction between saying what *is* and saying what *ought* to be. One need not buy the metaphysics that Kant uses to ground and explain his norms, nor accept his answer to Hume, in order to appreciate the transformation of perspective made possible by his emphasis on the normativeness of the conceptual, and hence of cognition and action—the latter distinguished in the first instance as what we are responsible for.

3. Frege: Justification versus Causation

Kant's lesson is taken over as a central theme by Frege, whose campaign against psychologism relies on respecting and enforcing the distinction between the normative significance of applying concepts and the causal consequences of doing so. For Frege, it is possible to investigate in a naturalistic way acts of judging or thinking (even thinking conceived in a dualistic way), but such an investigation inevitably overlooks the normative dimension that is essential to understanding the propositional contents that are judged or thought. Sometimes this point is put in terms of *reasons*, invoking inferential relations among judgeable contents, as when he complains that psychologism "loses the distinction between the grounds that justify a conviction and the causes that actually produce it,"³ or again when he argues that "the laws in accordance with which we actually draw inferences are not to be identified with the laws of correct [*richtigen*] inference; otherwise we could never draw a wrong inference."⁴ Sometimes the point is put in terms of *truth*, as when he says, "It is not the holding something to be true that concerns us, but the laws of truth. We can also think of these as prescriptions for making judgments; we must comply with them if our judgments are not to fail of the truth."⁵ Put either way, the point is that concern with the propositional contents that are thought or judged is inseparable from the possibility of assessments of *correctness*. Besides empirical regularities, there are also proprieties governing inferring and holding-true. Besides questions of which judgeable contents *are* held true, and under what circumstances, there is the question of which ones *ought* to be, and when. Besides the question of what consequences holding-true or making a judgment with

a certain content in fact leads to, there is the question of what those consequences ought or must rationally be. Psychology can study the matter-of-factual *properties* of contentful acts of judging and inferring, but not the semantically determined *proprieties* that govern them, the norms according to which assessments of truth and rationality are to be made.

Psychologism misunderstands the pragmatic significance of semantic contents. It cannot make intelligible the applicability of norms governing the acts that exhibit them. The force of those acts is a prescriptive rather than a descriptive affair; apart from their liability to assessments of judgments as true and inferences as correct, there is no such thing as judgment or inference. To try to analyze the conceptual contents of judgments in terms of habits or dispositions governing sequences of brain states or mentalistically conceived ideas is to settle on the wrong sort of modality, on causal necessitation rather than rational or cognitive right. Such natural processes "are no more true than false; they are simply processes, as an eddy in the water is a process. And if we are to speak of a right, it can only be the right of things to happen as they do happen. One phantasm contradicts another no more than one eddy in water contradicts another."⁶ Contradiction, correct inference, correct judgment are all *normative* notions, not *natural* ones.

The laws of nature do not forbid the making of contradictory judgments. Such judgments are forbidden in a normative sense. It is *incorrect* to endorse incompatible contents: rationally incorrect, incorrect according to rules of reason, prescriptions governing what is *proper* in the way of inferring and judging. The 'must' of justification or good inference is not the 'must' of causal compulsion. But the possibility of expressing each in terms of rules or laws, so central to Kant's enterprise, misleads if these two different sorts of laws are not kept distinct, as they are not by psychologism and associationism. "What makes us so prone to embrace such erroneous views is that we define the task of logic as the investigation of the laws of thought, whilst understanding by this expression something on the same footing as the laws of nature . . . So if we call them laws of thought, or, better, laws of judgment, we must not forget we are concerned here with laws which, like the principles of morals or the laws of the state, prescribe how we are to act, and do not, like the laws of nature, define the actual course of events."⁷

Frege expresses his views about the normative character of judgeable contents, which he understands as having truth conditions, and so about the application of concepts, which he understands as functions whose values are truth-values, by talking about the nature of logic, which he understands as the study of the laws of truth.

Logic, like ethics, can also be called a normative science.⁸

The property 'good' has a significance for the latter analogous to that which the property 'true' has for the former. Although our actions and

endeavours are all causally conditioned and explicable in psychological terms, they do not all deserve to be called good.⁹ Discussion of just how these remarks about the normative or prescriptive character of logic relate to a commitment to the normative or prescriptive significance of the exhibition of conceptual content by judgments must await more detailed consideration of Frege's *Begriffsschrift* theory of logical vocabulary as expressive of conceptual contents, in Chapter 2.

4. Wittgenstein on the Normative Significance of Intentional Content

Frege emphasizes that concern with the contents of concepts and judgments is inseparable from concern with the possibility of the concepts being correctly or incorrectly applied, the judgments correctly or incorrectly made, whether this correctness is conceived in terms of truth or of the goodness of inference. Understanding this point requires distinguishing normative from causal modalities. Beyond enforcing this distinction, however, Frege has little to say about the nature of the norms that matter for the study of conceptual contents, and so for logic—though of course he has a great deal to say about the structure of such contents (some of which will be rehearsed in subsequent chapters). His concerns are at base semantic rather than pragmatic. In the twentieth century, the great proponent of the thesis that *intentionally* contentful states and acts have an essentially *normative* pragmatic significance is the later Wittgenstein.

The starting point of his investigations is the insight that our ordinary understanding of states and acts of meaning, understanding, intending, or believing something is an understanding of them as states and acts that *commit* or *oblige* us to act and think in various ways. To perform its traditional role, the meaning of a linguistic expression must determine how it would be *correct* to use it in various contexts. To understand or grasp such a meaning is to be able to distinguish correct from incorrect uses. The view is not restricted to meaning and understanding but extends as well to such intentionally contentful states as believing and intending. This is one way of developing and extending Kant's point that to take what we do as judging and acting is to treat it as subject to certain kinds of assessments as to its correctness: truth (corresponding to the world) and success (corresponding to the intention). A particular belief may actually relate in various ways to how things are, but its content determines how it is *appropriate* for it to be related, according to the belief—namely that the content of the taking-true should be true. A particular intention may or may not settle how one *will* act, but its content determines how it is *appropriate* to act, according to the intention—namely by making-true that content. To say this is in no way to deny that occurrences of intentional states of meaning, understanding, intending, and believing have causal significances. It is simply to point out that

understanding them as contentful involves understanding them as also having normative significances.¹⁰

The issue constantly before us in Wittgenstein's later works is how to understand these normative significances of intentional contents—the way in which they incorporate standards for assessments of correctness. Many of his most characteristic lines of thought are explorations of the inaptness of thinking of the normative 'force', which determines how it would be appropriate to act, on the model of a special kind of causal 'force'. The sense in which understanding or grasping a meaning is the source of the *correct* use is quite different from the sense in which it is the source of what one in fact goes on to do.¹¹ Enforcing the Kantian and Fregean distinction between grounds in the order of justification and grounds in the order of causation is what is behind talk of the "hardness of the logical 'must'"¹² and the picture of the dominion or compulsion intentional states exercise over what counts as correct performance as a machine whose "super-rigid" construction precludes any sort of malfunction. "The machine as symbolizing its action . . . We talk as if these parts could only move in this way, as if they could not do anything else. How is this—do we forget the possibility of their bending, breaking off, melting, and so on? Yes; in many cases we don't think of that at all . . . And it is quite true: the movement of the machine-as-symbol is predetermined in a different sense from that in which the movement of any given actual machine is predetermined."¹³

The relation between the content of an intention and the performances that would fulfill that intention does not leave any room for misfire, corresponding to the melting or bending of the parts of a mechanism, for it is already a *normative* relation. The state is to settle what *ought* to be done, what *must* be done if it is to be realized. What actually does or would happen is another matter. The images of superrigidity—of being guided by rails that one cannot fall away from—are what one gets if one assimilates normative compulsion to causal compulsion, ignoring the Kantian distinction. That is, if the normative 'must' were a kind of causal 'must', it would have to be a puzzling, superrigid sort—but the point is not to start with this sort of naturalistic prejudice.

In fact, by contrast, "The laws of inference can be said to compel us; in the same sense, that is to say, as other laws in human society."¹⁴ They determine, in a sense yet to be specified, what one *ought* to do. Being compelled in this sense is entirely compatible with failing to act as one 'must'. Indeed, the physical or causal possibility of making a mistake, or doing what one is obliged, by what one means, intends, believes, and desires, not to do, is essential to the conception of such states and shows the essentially normative nature of their significance. "'But I don't mean that what I do now (in grasping a sense) determines the future use causally and as a matter of experience, but that in a queer way, the use itself is already present'.—But of course it is, 'in some sense'! Really the only thing wrong with what you say

is the expression 'in a queer way'. The rest is all right; and the sentence only seems queer when one imagines a different language game for it from the one in which we actually use it"¹⁵—the different game, namely, of attributing *natural states* and properties, rather than *normative statuses* such as commitments. What is determined is not how one *will* act but how one *ought* to, given the sense or content grasped, or the rule one has endorsed. "How am I able to obey a rule?"—if this is not a question about causes, then it is about the justification for my following the rule in the way I do."¹⁶ That is, it is a question about what actions accord with the rule, are obliged or permitted by it, rather than with what my grasp of it actually makes me do.

5. Norms and Intentional Explanation

Although Wittgenstein often uses specifically linguistic examples, and some commentators have focused exclusively on these cases, the normative phenomena he highlights are part and parcel of intentional attribution generally, whether or not language is in the picture. *Ceteris paribus*, one who believes that it is raining, and that moving under the tree is the only way to stay dry, and who desires to stay dry, *ought* to move under the tree. The intentional states make the action appropriate. Indeed, the concept of rationality achieves its paradigmatic application in just such circumstances, as conduct warranted by the attributed intentional states is characterized as *rationally* appropriate. The qualification marks the option being reserved to deny that the conduct is, say, morally, politically, or aesthetically appropriate. (It is a further question whether this explanatory role of rationality justifies conceiving of what is rationally appropriate as reducible to what is prudentially or instrumentally appropriate.) Taking the category of rationality to be essentially involved in intentional explanation, as Dennett and Davidson for instance do, is one way of recognizing the normative dimension of intentionality.

It is important to keep this acknowledgment distinct from further theses one may then want to endorse concerning that normative dimension. If one keeps one's eye resolutely on the causal dimension of intentional explanation, the normative aspect can be masked. For instance, Dennett conjoins his recognition of the constitutive role of rationality in intentional explanation with the claim that such explanation involves a substantive "rationality assumption," the assumption in effect that the system will by and large in fact act as it ought rationally to act. There is nothing wrong with considering explanations of this sort as intentional explanations, but it is important to distinguish normative intentional explanation from causal intentional explanation. The former explains only what the subject of the intentional states ought or is obliged or committed (rationally) to do in virtue of its exhibition of the attributed states. The latter makes the substantive rationality assumption and goes on to explain what in fact happens. Normative intentional

explanations are more fundamental; they are presupposed and built upon by causal ones.

The same normative considerations arise if one approaches intentional states from the direction of their functional roles in mediating perception and action. Where Wittgenstein, Sellars, and Dummett, for instance, look at meaning, understanding, believing, or intending something in terms of mastery of the public proprieties governing the use of linguistic expressions, others would see intentional states as already definable by their role in accounting for the conduct of rational agents, whether linguistically adept or not. Both views are functionalist, in a broad sense. They differ over how to draw the boundaries around the functional systems within which alone something can have the significance of an intentional state. The issue of the extent to which mastery of linguistic social practice is a prerequisite for possession of intentional states of various sorts is of course an important one. But one need not have settled on one or the other of these explanatory approaches in order to appreciate that intentional states belong on the normative side of the Kantian divide. For talk of functional roles is itself already normative talk.

Specifying the functional role of some state in a system is specifying how it *ought* to behave and interact with other states. It is with reference to such a role that one makes sense of the notion of a *malfunctioning* component, something that is not behaving as it is supposed to. One may go on to offer various stories about the source of the correctnesses involved in functional roles: invoking the intentions of the designer, the purposes of the user, or the way the system must function if it is to realize the evolutionary good of survival, or even the cognitive good of accurately representing its environment. The point is that all of these are accounts of the source of the norms of proper functioning that are an integral part of functional explanations. The job of a designer's drawing of a machine is to specify how the machine is supposed to work, how it ought to work, according to the intentions of the designer. It is for this reason that "we forget the possibility of [the pieces] bending, breaking off, melting, and so on." Wittgenstein is of course concerned to understand how it is possible to understand such normative roles. But the current point is just that the roles one seeks to specify, in explaining the significance of intentional states, must, to begin with, be understood in normative terms of proper or correct functioning. Once again, this is not to deny that the fact that some component or system ought (functionally) to behave in a certain way may under many circumstances have a causal significance regarding how it will in fact behave. The issues are in principle distinct, however, and causal functional accounts presuppose normative functional ones.

The recognition that the consequences of attributing intentionally contentful states must be specified in normative terms may be summed up in the slogan, "Attributing an intentional state is attributing a normative

status." This is one of the leading ideas to be pursued in the present investigation. Intentional states and acts have contents in virtue of which they are essentially liable to evaluations of the "force of the better reason." It is this mysterious "force"—evidently the core of the social practices of giving and asking for reasons—that Greek philosophy investigated and appealed to in demarcating us from the nonrational background of items that we can think and find out about but that cannot themselves think or find out about other things. This "force of the better reason" is a normative force. It concerns what further beliefs one is *committed* to acknowledge, what one *ought* to conclude, what one is *committed* or *entitled* to say or do. Talk of what is a reason for what has to do in the first instance not with how people do or *would* act but with how they *should* act, what they should acknowledge. The sophist may not in fact respond to this "force," but even the sophist *ought* to. To understand rationality and states whose contents are articulated according to their role in reasoning, one must understand the force of such 'ought's. The relevance of reasons to the attributing and undertaking of intentional states and acts is *prima facie* reason to employ a normative metalanguage in analyzing such activity.

The normative dimension of intentional attributions is equally apparent if the propositional contents of the states and acts that are attributed, exhibited, or performed are conceived, not in terms of their accessibility to *reasons*, but in terms of there being circumstances under which they would be *true*. Assessments of truth, no less than assessments of rationality, are normative assessments. Truth and rationality are both forms of correctness. To ask whether a belief is true is to ask whether it is in some sense proper, just as to ask whether there are good reasons for it is to ask whether it is proper in a different sense. The business of truth talk is to evaluate the extent to which a state or act has fulfilled a certain kind of responsibility. This normative aspect of concern with truth can be masked by offering a descriptive, matter-of-factual account of what truth consists in. But doing so should be understood as offering a theory about this variety of semantic correctness, not as a denial that correctness is what is at issue. Thus Dummett argues that one does not understand the concept of truth when one has only a method for determining when it correctly applies to a claim or belief—a practical mastery of its circumstances of application. One must also know the *point* of applying it, must understand that truth is the *proper goal* of assertion and belief, that the language game of assertion and belief implicitly but essentially involves the injunction that one *ought* to speak and believe the truth. That is what one is supposed to be trying to do. Without an appreciation of this normative significance of application of the concept *truth*, one does not understand that concept.

Raising the question of what a belief or claim *represents* or is *about* can be understood as treating it as in a special way answerable for its *correctness* to what is represented, what it is about.¹⁷ Thus the claim that semantically

or intentionally contentful states and acts have, as such, pragmatic significances that should be specified in normative terms does not depend upon what particular model (for instance, reasons, truth conditions, or representation) is employed in understanding such contents. The theoretical task of the intentional content of a state or act is to determine, in context, the normative significance of acquiring that state or performing that act: when it is appropriate or correct to do so and what the appropriate consequences of doing so are. The content is to determine proprieties of use, employment, or performance for states, acts, and expressions that exhibit or express such contents. The content must (in context) settle when it is *correct* to apply a concept in judging, believing, or claiming, and what correctly follows from such an application. Correctnesses of application are discussed under the general headings of assessments of truth or representation; correctnesses of inference are discussed under the general heading of assessments of rationality.¹⁸ To pick out intentional states and acts as ones to which any of these sorts of assessments—truth, accuracy of representation, or reasonability—are in principle appropriate is to treat their normative articulation as essential to them. For this point, it does not matter which sort of assessment is treated as fundamental, whether the goodness of claiming of the sort concepts of truth try to capture, the goodness of representation that concepts of correspondence try to capture, or the goodness of reasoning of the sort concepts of rationality try to capture. All are *prima facie* normative or evaluative notions.

III. FROM NORMS EXPLICIT IN RULES TO NORMS IMPLICIT IN PRACTICES

1. *Regulism: Norms as Explicit Rules or Principles*

The first commitment being attributed to Wittgenstein, then, is to taking the significance of attributing intentional states to be normative, a matter of the difference it makes to the correctness or justification of possible performances (including the adoption of other intentional states). The second commitment he undertakes concerns how to understand the normative statuses of correct and incorrect, justified and not justified, which this approach to intentionality concentrates on. The question of how the normative significances of intentional states are to be taken to be related to the matter-of-factual consequences of those states, which would be one way into this issue, can be put to one side for the moment. It is a question Wittgenstein is much interested in, but it ought to be seen as arising at a different point in the argument. For an account of the normative pole of the Kantian dualism need not take the form of a specification of how the normative is related to the nonnormative. Instead, Wittgenstein considers, and rejects, a particular model of correctness and incorrectness, roughly Kant's, in which what makes

a performance correct or not is its relation to some explicit rule. To understand his argument and the lesson he draws from it, it is necessary to see what this model of the normative is, and for what sort of explanatory role he claims it is unsuitable.

According to this more specific Kantian view,¹⁹ norms just *are* rules of conduct. Normative assessments of performances are understood as always having the form of assessments of the extent to which those performances accord with some rule. Reference to proprieties of performance is taken as indirect reference to rules, which determine what is proper by explicitly *saying* what is proper. On this account, acts are liable to normative assessments insofar as they are governed by propositionally *explicit* prescriptions, prohibitions, and permissions. These may be conceived as rules, or alternatively as principles, laws, commands, contracts, or conventions. Each of these determines what one may or must do by *saying* what one may or must do. For a performance to be correct is, on this model, for the rules to permit or require it, for it to be in accord with principle, for the law to allow or demand it, for it to be commanded or contracted. It is because Kant is someone for whom the normative always appears in the explicit form of rules, laws, and commandments that he could see the rationalists' insistence on the essential role of principles in cognition and action as a dark appreciation of the fundamentally normative character of those faculties. It is for this reason that when Kant wants to say that we are creatures distinguished from others by the normative dimension of our conduct (both cognitive and practical), he puts this in terms of our being bound by *rules*.

On an approach according to which normative assessment of conduct—whether prospectively, in deliberation, or retrospectively, in appraisal—always begins with the question of what rule is followed in producing the performances in question, norms are likened to laws in the sense of statutes. For conduct is legally appropriate or inappropriate just insofar as it is governed by some explicit law that says it is. Assessments of legal praise and blame must at least implicitly appeal to the relation of the performance in question to some law. In this way, the model appeals to a familiar institutional context, in which the norms most in evidence clearly take the form of explicit principles, commands, and the like.

The influence of the jurisprudential analogy is evident in Kant's conception of the normative aspect of cognition and action in terms of following rules. Kant inherits the Enlightenment tradition, handed down from Grotius and Pufendorf, which first studied the normative in the form of positive and natural laws, conceived as the explicit commandments of sovereigns or superiors of one sort or another. As a result, Kant takes it for granted that it is appropriate to call a 'rule' or a 'law' whatever it is that determines the propriety or impropriety of some judgment or performance. For him, as for most philosophers before this century, explicit rules and principles are not simply one form among others that the normative might assume. Rules are

the form of the norm as such. This view, that proprieties of *practice* are always and everywhere to be conceived as expressions of the bindingness of underlying *principles*, may be called *regulism* about norms.²⁰

According to this intellectualist, platonist conception of norms, common to Kant and Frege, to assess correctness is always to make at least implicit reference to a rule or principle that *determines* what is correct by explicitly *saying* so. In the best-known portion of his discussion of rule-following in the *Philosophical Investigations*, Wittgenstein argues that proprieties of performance that are governed by explicit rules do not form an autonomous stratum of normative statuses, one that could exist though no other did. Rather, proprieties governed by explicit rules rest on proprieties governed by practice. Norms that are *explicit* in the form of rules presuppose norms *implicit* in practices.

2. Wittgenstein's Regress Argument

Norms explicit as rules presuppose norms implicit in practices because a rule specifying how something is correctly done (how a word ought to be used, how a piano ought to be tuned) must be applied to particular circumstances, and applying a rule in particular circumstances is itself essentially something that can be done correctly or incorrectly. A rule, principle, or command has normative significance for performances only in the context of practices determining how it is correctly applied. For any particular performance and any rule, there will be ways of applying the rule so as to forbid the performance, and ways of applying it so as to permit or require it. The rule determines proprieties of performance only when correctly applied.

If correctnesses of performance are determined by rules only against the background of correctnesses of application of the rule, how are these latter correctnesses to be understood? If the regulist understanding of all norms as rules is right, then applications of a rule should themselves be understood as correct insofar as they accord with some further rule. Only if this is so can the rule-conception play the explanatory role of being the model for understanding *all* norms. A rule for applying a rule Wittgenstein calls an "interpretation" (*Deutung*). "There is an inclination to say: every action according to the rule is an interpretation. But we ought to restrict the term 'interpretation' to the substitution of one expression of the rule for another."²¹ The question of the autonomy of the intellectualist conception of norms, presupposed by the claim that rules are the form of the normative, is the question of whether the normative can be understood as "rules all the way down," or whether regulist proprieties depend on some more primitive sort of *practical* propriety. Wittgenstein argues that the latter is the case. Rules do not apply themselves; they determine correctnesses of performance only in the context of practices of distinguishing correct from incorrect applications of the rules. To conceive these practical proprieties of application as themselves rule-governed is to embark on a regress. Sooner or later the theorist will have to

acknowledge the existence of practical distinctions between what is appropriate and what not, admitting appropriatenesses according to practice as well as according to rules or explicit principles.

This regress argument shows that the platonist conception of norms as rules is not an autonomous one, and so does not describe the fundamental form of norm. "What does a game look like that is everywhere bounded by rules? whose rules never let a doubt creep in, but stop up all the cracks where it might?—Can't we imagine a rule determining the application of a rule, and a doubt which it removes—and so on?"²² In each case the doubt is the possibility of a mistake, of going wrong, of acting incorrectly, for instance in applying a rule. The point is to be that a rule can remove such a doubt, settle what is correct to do, only insofar as it is itself correctly applied. "But how can a rule show me what I have to do at this point? Whatever I do is, on some interpretation, in accord with the rule.'—That is not what we ought to say, but rather: any interpretation [*Deutung*] still hangs in the air along with what it interprets [*dem Gedeuteten*], and cannot give it any support. Interpretations by themselves do not determine meaning."²³ No sequence of interpretations can eliminate the need to apply the final rules, and this is always itself subject to normative assessment. Applied incorrectly, any interpretation misleads. The rule says how to do one thing correctly only on the assumption that one can do something else correctly, namely apply the rule.

This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here. It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us for at least a moment, until we thought of yet another standing behind it. What this shows is that there is a way of grasping a rule [*eine Auffassung einer Regel*] which is not an interpretation, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.²⁴

Absent such a practical way of grasping norms, no sense can be made of the distinction between correct and incorrect performance—of the difference between acting according to the norm and acting against it. Norms would then be unintelligible.

3. Wittgenstein's Pragmatism about Norms

The conclusion of the regress argument is that there is a need for a *pragmatist* conception of norms—a notion of primitive correctnesses of performance *implicit* in *practice* that precede and are presupposed by their *explicit* formulation in *rules* and *principles*. "To use the word without a

justification does not mean to use it wrongfully [*zu Unrecht gebrauchen*]."²⁵ There is a kind of correctness that does not depend on explicit justifications, a kind of correctness of practice.

And hence also 'obeying a rule' is a practice [*Praxis*].²⁶

—To obey a rule, to make a report, to give an order, to play a game of chess, are customs [*Gepflogenheiten*] (uses, institutions).²⁷

The regress argument does not by itself provide such a conception of proprieties of practice; it just shows that without one we cannot understand how rules can codify the correctnesses that they do.

This argument shares its form with the regress Lewis Carroll invokes in "The Tortoise and Achilles"²⁸ but takes that line of thought one level deeper. That story depends on the fact that in a formal logical system, statements are inferentially inert. Even conditionals, whose expressive job it is to make inferential relations explicit as the contents of claims, license inferential transitions from premises to conclusions only in the context of rules permitting detachment. Rules are needed to give claims, even conditional claims, a normative significance for action. Rules specify how conditionals are to be used—how it would be correct to use them. It is the rules that fix the inference-licensing role of conditionals, and so their significance for what it is correct to do (infer, assert). 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. So one cannot express all of the rules that govern inferences in a logical system in the form of propositionally explicit postulates within that system.

Carroll uses the regress of conditionals that results from the attempt to replace the rule of conditional detachment by explicitly postulated conditionals as an argument to show this. Wittgenstein's regress-of-rules argument shows further that, while rules can codify the pragmatic normative significance of claims, they do so only against a background of practices permitting the distinguishing of correct from incorrect applications of those rules. 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. Wittgenstein's point is then that conceiving such significances in regulist terms, as the invocation of rules of inference does, is not the whole story. Rule-based proprieties of performance depend on practice-based ones. The regulist, platonist, intellectualist conception of norms must be supplemented by that of the pragmatist.

Two commitments have now been attributed to Wittgenstein. The first is a *normative* thesis about the *pragmatics* of intentionality. The second is a *pragmatic* thesis about the *normativeness* of intentionality. In the first case, pragmatics is distinguished from semantics, as the theory of the significance of contentful states and performances from the theory of their contents. In

the second case, pragmatic theories of norms are distinguished from platonist theories, in treating as fundamental norms *implicit* in *practices* rather than norms *explicit* in *principles*. The first point enforces attention to the significance of intentional states for what it is *correct* to do. The second point is that proprieties of practice must be conceivable antecedently to their being expressly formulated into propositionally explicit governing rules or principles. For performances can be rule-governed only insofar as they are governed as well by practices of applying rules.

It is useful to approach the sort of understanding that is involved in mastering a practice, for instance a practice of applying or assessing applications of a rule, by means of Ryle's distinction between knowing *how* and knowing *that*.²⁹ Knowing how to do something is a matter of practical ability. To know how is just to be reliably able. Thus one knows how to ride a bicycle, apply a concept, draw an inference, and so on just in case one can discriminate in one's practice, in the performances one produces and assesses, between correct and incorrect ways of doing these things.

The explicit knowing-that corresponding to such implicit knowing-how is a theoretical formulation or expression of that practical ability, in a rule or principle, that *says* what is correct and what not. The intellectualist picture underwrites every bit of know-how with a bit of knowledge-that, which may be only implicit in practical discriminations. "Compare knowing and saying: how many feet high Mont Blanc is—how the word 'game' is used—how a clarinet sounds. If you are surprised that one can know something and not be able to say it, you are perhaps thinking of a case like the first. Certainly not like the third."³⁰ What Wittgenstein shows is that the intellectualist model will not do as an account of the nature of the normative as such. For when applied to the norms governing the application of rules and principles, it generates a regress, which can be halted only by acknowledging the existence of some more primitive form of norm. The regress is Wittgenstein's master argument for the appropriateness of the pragmatist, rather than the regulist-intellectualist, order of explanation.³¹

4. *Sellars against Regulism*

Another thinker who, like Wittgenstein, takes his starting point from Kant's and Frege's appreciation of the normative character of intentionality (for him, coeval with language use) is Wilfrid Sellars. He takes up this theme in one of his earliest papers, published in 1947. The opening section of that paper is entitled "Behavior, Norm, and Semantic Meta-Language" and makes this point:

The psychologistic blunder with respect to 'means' is related to another fundamental error, that, namely, of confusing between (1) language as a descriptive category for which symbols are empirical classes to which

certain events belong (and hence are symbol-events) by virtue of performing an empirical function, with (2) language as an epistemological category for which the relation of type to token is not that of empirical class to member. . .

For the moment it will help clarify the epistemological distinction between symbol-types and symbol-tokens, if we think of the former as norms or standards, and of the latter as events which satisfy them. We can therefore, for the moment at least, contrast the above two senses of 'language' as the descriptive and the normative respectively. Making use of this distinction, we argue that 'meaning' or, better, 'designation' is a term belonging to language about languages in the second sense. Its primary employment is therefore in connection with linguistic expressions as norms.³²

Like Wittgenstein, Sellars sees that an adequate conception of these norms must move beyond the pervasive regulist tradition, which can understand them only in the form of rules.

Focusing on linguistic intentionality, Sellars in another paper examines the regulist conception as it applies to the linguistic norms in virtue of which it is possible to say anything at all. "It seems plausible to say that a language is a system of expressions, the use of which is subject to certain rules. It would seem, thus, that learning to use a language is learning to obey the rules for the use of its expressions. However, taken as it stands, this thesis is subject to an obvious and devastating objection."³³ The objection is that taking 'correct' to mean 'correct according to a rule' generates a familiar sort of regress:

The refutation runs as follows: Thesis. Learning to use a language (*L*) is learning to obey the rules of *L*. But, a rule which enjoins the doing of an action (*A*) is a sentence in a language which contains an expression for *A*. Hence, a rule which enjoins the using of a linguistic expression (*E*) is a sentence in a language which contains an expression for *E*—in other words, a sentence in a metalanguage. Consequently, learning to obey the rules for *L* presupposes the ability to use the metalanguage (*ML*) in which the rules for *L* are formulated. So that, learning to use a language (*L*) presupposes having learned to use a metalanguage (*ML*). And by the same token, having learned to use *ML* presupposes having learned to use a metmetalanguage (*MML*) and so on. But, this is impossible (a vicious regress). Therefore, the thesis is absurd and must be rejected.³⁴

The metalanguage expresses rules for the proper application of concepts of the object language. But these rules, too, must be applied. So the

metametalinguage expresses rules for applying the rules of the metalanguage, and so on.

If any talk is to be possible, there must be some meta . . . metalevel at which one has an understanding of rules that does not consist in offering another interpretation of them (according to rules formulated in a metalanguage) but which consists in being able to distinguish correct applications of the rule in practice. The question is how to understand such practical normative know-how. Although he, like Wittgenstein, uses 'rule' more broadly than is here recommended, Sellars is clearly after such a notion of norms implicit in practice: "We saw that a rule, properly speaking, isn't a rule unless it lives in behavior, rule-regulated behavior, even rule-violating behavior. Linguistically we always operate within a framework of living rules. (The snake which sheds one skin lives within another.) In attempting to grasp rules as rules from without, we are trying to have our cake and eat it. To describe rules is to describe the skeletons of rules. A rule is lived, not described."³⁵

This line of thought, common to Wittgenstein and Sellars, raises the key question of how to understand proprieties of practice, without appealing to rules, interpretations, justifications, or other explicit claims that something is appropriate. What does the practical capacity or 'know-how' to distinguish correct from incorrect performances (for instance—but this is only one example—applications of a rule) consist in? This is to ask what it is to take or treat a performance as correct-according-to-a-practice. It should also be asked, What is it for an act to be correct-according-to-a-practice? Both questions are important ones to ask: In what sense can norms (proprieties, correctnesses) be implicit in a practice? and What is it for someone to acknowledge those implicit norms as governing or being binding on a range of performers or performances?

The answers to these questions may be more intimately related to one another than at first appears. To foreshadow: On the broadly phenomenalist line about norms that will be defended here, norms are in an important sense in the eye of the beholder, so that one cannot address the question of what implicit norms are, independently of the question of what it is to acknowledge them in practice. The direction of explanation to be pursued here first offers an account of the practical attitude of *taking* something to be correct-according-to-a-practice, and then explains the status of *being* correct-according-to-a-practice by appeal to those attitudes. Filling in a story about normative attitudes as assessments of normative status, and explaining how such attitudes are related both to those statuses and to what is actually done, will count as specifying a sense of "norms implicit in practice" just insofar as the result satisfies the criteria of adequacy imposed on the notion of practice by the regress-of-rules argument.

Another central explanatory criterion of adequacy for such a conception of implicit practical normative knowing-how is that it be possible in terms

of it to understand explicit knowing-*that*. The effect is to reverse the regulist-intellectualist order of explanation. The regulist starts with a notion of norms explicit in principles and is obliged then to develop an account of what it would be for such things to be implicit in practices. The pragmatist starts rather with a notion of norms implicit in practice and is obliged then to develop an account of what it would be for such things to become propositionally explicit, as claims or rules. One of the primary tasks of this book is accordingly to offer an account of what it is to take some propriety that is implicit in a practice and make it explicit in the form of a claim, principle, or rule.

5. *Regularism: Norms as Regularities*

The regress-of-rules or regress-of-interpretations argument common to Wittgenstein and Sellars sets up criteria of adequacy for an account of contentful states that acknowledges their essentially normative significance, their characteristic relevance to assessments of the correctness of acts (including the adoption of further states). It must be possible to make sense of a notion of norms implicit in practice—which participants in the practice are bound by, and can acknowledge being bound by—without appeal to any explicit rules or capacities on the part of those participants to understand and apply such rules. Since the regress arises when the rule-following model of being bound by norms is applied to the agent, one strategy for avoiding it is to shift to a different model. Perhaps rules are relevant only as describing regularities, and not as being followed in achieving them.

Sellars (who does not endorse it) introduces such an approach this way: "Now, at first sight there is a simple and straightforward way of preserving the essential claim of the thesis while freeing it from the refutation. It consists in substituting the phrase 'learning to conform to the rules . . .' for 'learning to obey the rules . . .' where 'conforming to a rule enjoining the doing of *A* in circumstances *C*' is to be equated simply with 'doing *A* when the circumstances are *C*'—regardless of how one comes to do it . . . A person who has the habit of doing *A* in *C* would then be conforming to the above rule even though the idea that he was to do *A* in *C* never occurred to him, and even though he had no language for referring to either *A* or *C*."³⁶ What generates the regress is the demand that each practical capacity to act appropriately be analyzed as following an explicit rule that *says* what is appropriate, since understanding what is said by such a rule turns out to involve further practical mastery of proprieties.

If the practices in which norms are implicit are understood simply as regularities of performance, then there is nothing the practitioner need already understand. If such regularities of performance can be treated as practices governed by implicit norms, then there will be no regress or circularity in appealing to them as part of an account of knowing-*that*, of expressing

norms explicitly in rules and principles. For the only one who needs to understand how to apply correctly the rule conforming to which makes performances count as regular is the theorist who describes the regularity in terms of that rule. The norms implicit in regularities of conduct can be expressed explicitly in rules, but need not be so expressible by those in whose regular conduct they are implicit.

The view that to talk about implicit norms is just to talk about regularities—that practices should be understood just as regularities of behavior—may be called the simple regularity theory. It is clear how such a *regularist* account of the normative avoids the regress that threatens *regulist* accounts. The proposal is to identify being correct according to (norms implicit in) practice—in the sense required to avoid the regress of rules as interpretations that plagues fully platonist accounts—with conforming to (norms explicit in) a rule, where ‘conforming to a rule’ is just producing performances that are regular in that they count (for us) as correct according to it. The immediate difficulty with such a proposal is that it threatens to obliterate the contrast between treating a performance as subject to normative assessment of some sort and treating it as subject to physical laws.

For this reason simple regularity theories seem to abandon the idea that the significance of contentful states is to be conceived in normative terms. No one doubts that actions and linguistic performances are subject to laws of the latter sort and so conform to rules or are regular. The thesis of the normative significance of intentional states sought to distinguish intentional states from states whose significance is merely causal, and that distinction seems to be taken back by the simple regularity account. After all, as Kant tells us, in this sense “everything in nature, in the inanimate as well as the animate world, happens according to rules . . . All nature is actually nothing but a nexus of appearances according to rules; and there is nothing without rules.”³⁷ Everything acts regularly, according to the laws of physics. In what special sense do intentional states then involve specifically normative significances?

For a regularist account to weather this challenge, it must be able to fund a distinction between what is in fact done and what ought to be done. It must make room for the permanent possibility of mistakes, for what is done or taken to be correct nonetheless to turn out to be incorrect or inappropriate, according to some rule or practice. The importance of this possibility to the genuinely normative character of the force or significance associated with contentful states is a central and striking theme in Wittgenstein’s later works. What is correct or appropriate, what is obligatory or permitted, what one is committed or entitled to do—these are normative matters. Without the distinction between what *is* done and what *ought* to be done, this insight is lost.

The simple regularity approach is committed to identifying the distinction between *correct* and *incorrect* performance with that between *regular* and

irregular performance. A norm implicit in a practice is just a pattern exhibited by behavior. To violate that norm, to make a mistake or act incorrectly according to that norm, is to break the pattern, to act irregularly. The progress promised by such a regularity account of proprieties of practice lies in the possibility of specifying the pattern or regularity in purely descriptive terms and then allowing the relation between regular and irregular performance to stand in for the normative distinction between what is correct and what is not. Wittgenstein explicitly considers and rejects this approach. Where his master argument against regulism has the form of an appeal to the regress of interpretations, his master argument against regularism has the form of an appeal to the possibility of gerrymandering.

The problem is that any particular set of performances exhibits many regularities. These will agree on the performances that have been produced and differ in their treatment of some possible performances that have not (yet) been produced. A performance can be denominated 'irregular' only with respect to a specified regularity, not *tout court*. Any further performance will count as regular with respect to some of the patterns exhibited by the original set and as irregular with respect to others. For anything one might go on to do, there is some regularity with respect to which it counts as "going on in the same way," continuing the previous pattern. Kripke has powerfully expounded the battery of arguments and examples that Wittgenstein brings to bear to establish the point in this connection.³⁸ There simply is no such thing as *the* pattern or regularity exhibited by a stretch of past behavior, which can be appealed to in judging some candidate bit of future behavior as regular or irregular, and hence, on this line, as correct or incorrect. For the simple regularist's identification of impropriety with irregularity to get a grip, it must be supplemented with some way of picking out, as somehow *privileged*, some out of all the regularities exhibited. To say this is to say that some regularities must be picked out as the ones that *ought* to be conformed to, some patterns as the ones that *ought* to be continued. The simple regularity view offers no suggestions as to how this might be done and therefore does not solve, but merely puts off, the question of how to understand the normative distinction between what is done and what ought to be done.

One might respond to the demand that there be some way to pick out the correct regularity, from all the descriptively adequate but incompatible candidates, by shifting what one describes, from finite sets of performances to the set of performances (for instance, applications of a concept) the individual is *disposed* to produce. This set is infinite, in that any bearer of an intentional state is disposed to respond, say by applying or refusing to apply the concept *red* or *prime*, in an infinite number of slightly different circumstances. Kripke argues that this appeal to dispositions nevertheless does not suffice to rule out regularities that agree in all the cases one has dispositions with respect to, and differ in others so remote (perhaps, in the case of *prime*, because the numbers involved are so large, and in the case of *red* because surrounding

circumstances are so peculiar) that one does not have dispositions to treat them one way rather than another.

This last argument is controversial, but it is not a controversy that need be entered into here; however it may be with the finiteness objection to a dispositional account of the regularities that, according to the line of thought being considered, are to play the role of norms implicit in practice, there is another more serious objection to it. No one ever acts incorrectly in the sense of violating his or her own dispositions. Indeed, to talk of 'violating' dispositions is illicitly to import normative vocabulary into a purely descriptive context. Understanding the norms implicit in practice as descriptively adequate rules codifying regularities of disposition (even if a unique set of such rules is forthcoming) loses the contrast between correct and mistaken performance that is of the essence of the sort of normative assessment being reconstructed. If whatever one is disposed to do counts for that reason as right, then the distinction of right and wrong, and so all normative force, has been lost. Thus the simple regularity view cannot be rescued from the gerrymandering objection by appealing to dispositions in order to single out or privilege a unique regularity.

The problem that Wittgenstein sets up, then, is to make sense of a notion of norms implicit in practice that will not lose either the notion of implicitness, as regulism does, or the notion of norms, as simple regularism does. McDowell puts the point nicely: "Wittgenstein's problem is to steer a course between a Scylla and a Charybdis. Scylla is the idea that understanding is always interpretation. We can avoid Scylla by stressing that, say, calling something 'green' can be like crying 'Help' when one is drowning—simply how one has learned to react to this situation. But then we risk steering on to Charybdis—the picture of a level at which there are no norms . . . How can a performance be nothing but a 'blind' reaction to a situation, not an attempt to act on interpretation (thus avoiding Scylla); and be a case of going by a rule (avoiding Charybdis)? The answer is: by belonging to a custom (*PI* 198), practice (*PI* 202), or institution (*RFM* VI-31)."³⁹ The Scylla of regulism is shown to be unacceptable by the regress-of-rules argument. The Charybdis of regularism is shown to be unacceptable by the gerrymandering-of-regularities argument.

If anything is to be made of the Kantian insight that there is a fundamental normative dimension to the application of concepts (and hence to the significance of discursive or propositionally contentful intentional states and performances), an account is needed of what it is for norms to be implicit in practices. Such practices must be construed both as not having to involve explicit rules and as distinct from mere regularities. Wittgenstein, the principled theoretical quietist, does not attempt to provide a theory of practices, nor would he endorse the project of doing so. The last thing he thinks we need is more philosophical theories. Nonetheless, one of the projects pursued in the rest of this work is to come up with an account of norms implicit in

practices that will satisfy the criteria of adequacy Wittgenstein's arguments have established.

IV. FROM NORMATIVE STATUS TO NORMATIVE ATTITUDE

1. *Kant: Acting According to Conceptions of Rules*

Two theses have so far been attributed to Kant. First, the sort of intentionality characteristic of us, exhibited on the theoretical side in judgment and on the practical side in action, has an essential *normative* dimension. Second, norms are to be understood as having the form of *explicit rules*, or principles. The first of these has been endorsed, as expressing a fundamental insight. The second has been rejected, on the basis of Wittgenstein's argument from the regress of rules as interpretations of rules. The conclusion drawn was that norms that are explicitly expressed in the form of rules, which determine what is correct according to them by *saying* or describing what is correct, must be understood as only one form that norms can take. That form is intelligible only against a background that includes norms that are *implicit* in what is *done*, rather than *explicit* in what is *said*.⁴⁰ At least the norms involved in properly understanding what is said by rules, or indeed in properly understanding any explicit saying or thinking, must be construed as norms of practice, on pain of a vicious regress.

In Kant's account of us as normative creatures, however, these two theses are inseparably bound up with a third. As has already been pointed out, Kant takes it that everything in nature happens according to rules. Being subject to rules is not special to us discursive, that is concept-applying, subjects of judgment and action.⁴¹ What is distinctive about us as normative creatures is the *way* in which we are subject to norms (for Kant, in the form of rules). As natural beings, we act according to rules. As rational beings, we act according to our *conceptions* of rules.⁴² It is not being bound by necessity, acting according to rules, that sets us apart; it is being bound not just by natural but by rational necessity. Kant's whole practical philosophy, and in particular the second Critique, is devoted to offering an account of this distinction between two ways in which one can be related to rules. Most of the details of his way of working out this idea are special to the systematic philosophical setting he develops and inhabits and need not be rehearsed here. Two fundamental features of his idea, however, must be taken seriously by any attempt to pursue his point about the normative character of concept-users.

The first of these has already been remarked on in connection with Frege. It concerns the distinction between the causal modalities and the more properly normative 'ought's whose applicability to us is being considered as a criterion of demarcation. This is the phenomenon distinguishing the force of causal 'must's from the force of logical or rational 'must's that Wittgen-

stein invokes in connection with his discussion of misunderstandings of the 'hardness' of the latter in relation to the former. It is an essential feature of the sort of government by norms that Kant is pointing to that it is compatible with the possibility of *mistakes*, of those subject to the norms going *wrong*, *failing* to do what they are obliged by those norms to do, or doing what they are *not* entitled to do. The 'ought' involved in saying that a stone subject to no other forces *ought* to accelerate toward the center of the earth at a rate of 32 feet per second per second shows itself to have the force of an attribution of natural or causal necessity by entailing that the stone *will* so act. The claim that it in this sense ought to behave a certain way is incompatible with the claim that it does not do so. In contrast, no such entailment or incompatibility is involved in claims about how we intentional agents ought to behave, for instance what else one of us is committed to believe or to do by having beliefs and desires with particular contents. Leaving room for the possibility of mistakes and failures in this way is one of the essential distinguishing features of the 'ought's that express government by norms in the sense that is being taken as characteristic of us, as opposed to it. The sense in which we are *compelled* by the norms that matter for intentionality, norms dictating what we are under various circumstances obliged to believe and to do, is quite different from natural compulsion.

The second feature of Kant's idea addresses precisely the nature of this normative compulsion that is nevertheless compatible with recalcitrance. For he does not just distinguish the sense in which we are bound by these norms from the sense in which we are bound by natural necessity in the purely formal terms invoked by this familiar point about the possibility of our going wrong. He characterizes it substantively as acting according to a *conception* or a *representation* of a rule, rather than just according to a rule. Shorn of the details of his story about the nature of representations and the way they can affect what we do, the point he is making is that we act according to our *grasp* or *understanding* of rules. The rules do not immediately compel us, as natural ones do. Their compulsion is rather mediated by our *attitude* toward those rules. What makes us act as we do is not the rule or norm itself but our *acknowledgment* of it. It is the possibility of this intervening attitude that is missing in the relation between merely natural objects and the rules that govern them. The slippage possible in our acting according to our conception of a rule is made intelligible by distinguishing the sense in which one is bound by a rule whose grip on us depends on our recognition or acknowledgment of it as binding from the sense in which one can be bound by a rule whose grip does not depend on its being acknowledged. This explanatory strategy might be compared to Descartes's invocation of intervening representations in explaining the possibility of error about external things—though Kant need not be understood as following Descartes's path from an implicit appeal to the regress that threatens such representationalist pictures of cognition to a diagnosis of the relation be-

tween the subject and those mediating, error-enabling internal representations as itself immediate and hence immune to the possibility of error.

The idea underlying the demarcational strategy Kant introduces when he defines us, denizens of the realm of freedom, as beings that are capable of acting according to a conception of a rule by contrast to the denizens of the realm of nature, is that natural beings, who merely act according to rules, that is, regularly, are capable of acknowledging norms *only* by obedience to them. We rational beings are also capable of grasping or understanding the norms, of making assessments of correctness and incorrectness according to them. Those assessments play a role in determining what we go on to do—the phenomenon Kant denominates “the rational will.” But for us, in contrast to merely natural creatures, the assessment of the propriety of a performance is one thing, and the performance itself is another. The possibility of not doing what we nevertheless count as bound or obliged to do arises out of this distinction. What is special about us is the sort of grasp or uptake of normative significance that we are capable of. To be one of us in this sense of ‘us’ is to be the subject of normative *attitudes*, to be capable of acknowledging proprieties and improprieties of conduct, to be able to treat a performance as correct or incorrect.

2. Practical Normative Attitudes

It is a challenge to retain this insight about the significance of our normative attitudes while accommodating Wittgenstein’s pragmatist point about norms (and so jettisoning the intellectualist insistence on the explicitness of norms that colors Kant’s treatment). In order to do so, it must be possible to distinguish the attitude of acknowledging *implicitly* or *in practice* the correctness of some class of performances from merely exhibiting regularities of performance by producing only those that fall within that class. Otherwise, inanimate objects will count as acknowledging the correctness of laws of physics, and the distinction Kant points out is lost. As before, the challenge is to reject intellectualist regulism about norms without falling into nonnormative regularism.

Consideration of this third thesis of Kant’s sharpens the point, however, by focusing it on the capacity to adopt a normative practical attitude—to act in such a way as to attribute a normative significance, without doing so by *saying* that that is what one is doing. The question now becomes, What must one be able to *do* in order to count as *taking* or *treating* a performance as correct or incorrect? What is it for such a normative attitude—attributing a normative significance or status to a performance—to be implicit in practice? The importance of this question is a direct consequence of Kant’s point, once his rendering has been deintellectualized by replacing grasp of principles with mastery of practices.

It is a consequence of this criterion of adequacy that the practical performances that are assessments cannot be just the same performances that are assessed. Addressing the simplest case first, treating a performance as correct cannot be identified with producing it. For according to such an identification, the only way in which a norm can be acknowledged in practice is by obeying it, acting regularly according to it. But then it is impossible to treat performances as *incorrect*: the norms one counts as acknowledging simply are whatever regularities describe what one does. Such an account would collapse Kant's distinction between the way in which we are governed by norms we acknowledge and the way in which we are governed by natural laws independently of our acknowledgment of them.

Kant's principle that we are the ones who act not only according to rules but according to a conception of them is the claim that we are not merely *subject* to norms but *sensitive* to them. This principle has been taken over here by saying that we are characterized not only by normative *statuses*, but by normative *attitudes*—which is to say not only that our performances are correct or incorrect according to various rules but also that we can in our practice *treat* them as correct or incorrect according to various rules. Using 'assessment' to mean an assignment of normative significance—in the most basic case taking as correct or incorrect—the point may be put by saying that Kant's principle focuses demarcational interest on the normative *attitudes* exhibited by the activity of *assessing*, rather than just on the normative *statuses* being *assessed*. In order to respect the lessons of Wittgenstein's pragmatism about the normative, assessing must be understood as something *done*; the normative attitude must be construed as somehow implicit in the practice of the assessor, rather than explicit as the endorsement of a proposition. Construed in these practical terms, a consequence of Kant's distinction is that mere conformity to a norm is not even a candidate as a construal of the normative attitude of assessing conformity, which expresses the sort of sensitivity to norms that characterizes us.

For brutes or bits of the inanimate world to qualify as engaging in practices that implicitly acknowledge the applicability of norms, they would have to exhibit behavior that counts as treating conduct (their own or that of others) *as* correct or incorrect. Of course such things do respond differentially to their own and each other's antics. The question is what role such a response must play in order to deserve to be called a practical taking or treating of some performance *as* correct or incorrect, perhaps in the way in which eating something deserves to be called a practical taking or treating of it as food. Any sort of reliable differential responsive disposition can be understood as inducing a classification of stimuli. Iron rusts in some environments and not others, and so can be interpreted as classifying its environments into two sorts, depending on which kind of response they tend to elicit. Such responsive classification is a primitive kind of practical taking *of* something *as*

something. It is in this sense that an animal's eating something can be interpreted as its thereby *taking* what it eats *as* food. The issue of current concern is what must be true of a behavioral response-kind for it to be *correct* or *appropriate* that something is taken as by being responded to in that way, rather than *food* or *wet*.

3. Sanctions

In part because much of the tradition of thought about normative status and attitudes has taken its departure from a legal model, it is natural to answer this question by invoking the notion of *sanctions*: of reward and punishment. According to such a *retributive* approach to assessment, one treats a performance as correct or appropriate by *rewarding* it, and as incorrect or inappropriate by *punishing* it. Such an account can take many forms, depending on how sanctions are construed. In the simplest case, applying a negative sanction might be understood in terms of corporal punishment; a prelinguistic community could express its practical grasp of a norm of conduct by beating with sticks any of its members who are perceived as transgressing that norm. In these terms it is possible to explain for instance what it is for there to be a practical norm in force according to which in order to be entitled to enter a particular hut one is obliged to display a leaf from a certain sort of tree. The communal response of beating anyone who attempts to enter without such a token gives leaves of the proper kind the normative significance, for the community members, of a license. In this way members of the community can show, by what they do, what they *take* to be appropriate and inappropriate conduct.

One example of this approach is Haugeland's account of practical norms in terms of social constellations of dispositions having a structure he calls "conformist." He asks us to imagine under this heading creatures who not only conform their behavior to that of other community members in the sense of imitating each other, and so tending to act alike (normally in the sense of typically) in similar circumstances, but also sanction each other's performances, making future behavior more likely to conform to ("cluster around") the emergent standards. "The clusters that coalesce can be called 'norms' (and not just groups or types) precisely because they are generated and maintained by censoriousness; the censure attendant on deviation automatically gives the standards (the extant clusters) a de facto normative force."⁴³ According to such an account the normative attitudes of taking or treating some performance as correct or incorrect are understood in terms of behavioral *reinforcement*, in the learning-theoretic sense. The advantage of such a way of putting things is that reinforcement is a purely functional descriptive notion, definable in abstraction from the particular considerations about familiar animals, in virtue of which beating them with a stick is

likely to function as negative reinforcement. Treating a performance as correct is taken to be positively sanctioning it, which is to say positively reinforcing it. Positively reinforcing a disposition to produce a performance of a certain kind as a response to a stimulus of a certain kind is responding to the response in such a way as to make it more likely in the future that a response of that kind will be elicited by a stimulus of the corresponding kind. Treating a performance as incorrect is taken to be negatively sanctioning it, which is to say negatively reinforcing it. Negatively reinforcing a disposition to produce a performance of a certain kind as a response to a stimulus of a certain kind is responding to the response in such a way as to make it less likely in the future that a response of that kind will be elicited by a stimulus of the corresponding kind.

The approach being considered distinguishes us as norm-governed creatures from merely regular natural creatures by the normative *attitudes* we evince—attitudes that express our grasp or practical conception of our behavior as governed by norms. These normative attitudes are understood in turn as *assessments*, assignments to performances of normative significance or status as correct or incorrect according to some norm. The assessing attitudes are then understood as dispositions to *sanction*, positively or negatively. Finally, sanctioning is understood in terms of *reinforcement*, which is a matter of the actual effect of the sanctioning or reinforcing responses on the responsive dispositions of the one whose performances are being reinforced, that is sanctioned, that is assessed.

Such a story is a kind of regularity theory, but not a simple regularity theory. It does not identify a norm wherever there is a regularity of behavior. In keeping with Kant's insight (as transposed from an intellectualist to a pragmatist key), norms are discerned only where attitudes—acknowledgments in practice of the bindingness of those norms—play a mediating role in regularities. Only insofar as regularities are brought about and sustained by effective assessments of propriety, in the form of responsive classifications of performances as correct or incorrect, are regularities taken to have specifically normative force. The possibility of incorrect, inappropriate, or mistaken performances—those that do not accord with the norm—is explicitly allowed for. Thus there is no danger of this sense of 'obligatory' collapsing into the sort of causal modality that governs merely natural happenings. Acknowledging a norm as in force is distinguishable in these terms from simply obeying it. A cardinal advantage of these theories is that while to this extent countenancing Kant's distinction between genuinely norm-governed and merely regular activity, they make intelligible how conduct that deserves to be called distinctively norm-governed could arise in the natural world.

The fundamental strategy pursued by such a theory is a promising one. As here elaborated, it involves three distinguishable commitments. First, Kant's distinction between acting according to a rule and acting according to a

conception of a rule is taken to express an important insight about the special way in which we are normative creatures. Second, the pragmatist regress-of-rules argument is taken to show that in order to make use of this insight, it is necessary that the sort of normative attitude that Kant takes to play an essential mediating role in our government by norms be understood as involving implicit acknowledgment of norms in practice. Specifically, it is necessary to make sense of the idea of practically taking or treating performances as correct or incorrect. Third, taking or treating performances as correct or incorrect, approving or disapproving them in practice, is explained in terms of positive and negative sanctions, rewards and punishments. This tripartite strategy is endorsed and pursued in the rest of this work. There are reasons not to be happy with the regularist way of working it out that has just been sketched, however.

Even the version of a regularity theory that Haugeland presents, which appeals only to patterns of positive and negative reinforcement to fund the notion of sanctions and thereby that of practical normative attitudes, merely puts off the issue of gerrymandering. Just as there is no such thing as *the* regularity of performance evinced by some actual course of conduct—because if there is one way of specifying it, there are an infinite number of distinguishable variants that overlap or agree about the specified performances and disagree about what counts as “going on in the same way”—so there is no such thing as *the* regularity that is being reinforced by a certain set of responses to responses, or even dispositions to respond to responses. The issue of gerrymandering, of how to privilege one specification of a regularity over equally qualified competitors, arises once more at the level of the reinforcing regularity. Again, simple regularity theories are subject to the objection that they conflate the categories of what is in fact done and what ought to be done, and hence that they fail to offer construals of genuinely normative significances of performances at all.

This is a way of failing to take sufficiently seriously Kant's distinction between acting according to a rule and according to a conception of a rule. Sanctions theories fund this crucial distinction by means of the distinction between producing a performance and assessing it. But assessing, sanctioning, is itself something that can be done correctly or incorrectly. If the normative status of being incorrect is to be understood in terms of the normative attitude of treating as incorrect by punishing, it seems that the identification required is not with the status of *actually* being punished but with that of *deserving* punishment, that is, being *correctly* punished. Of course sanctioners can be sanctioned in turn for their sanctioning, which is thereby treated as itself correctly or incorrectly done. Nonetheless, if actual reinforcement of dispositional regularities is all that is available to appeal to in making sense of this regress, it might still be claimed that what is instituted by this hierarchy of regularities of responses to regularities of responses ought not to count as genuinely normative.⁴⁴

4. Regularities of Communal Assessment

There are other dimensions along which it is instructive to consider sanctions theories of implicitly normative practice. Accounts that endorse the tripartite strategy by rendering practical acknowledgment of normative significance in terms of reward and punishment understand normative *status*, paradigmatically the significance of a performance as correct or incorrect, in terms of normative *attitudes*, paradigmatically *taking* or *treating as* correct or incorrect. Their characteristic suggestion is that this sort of practical attitude of assessment—responsive classification as correct or incorrect—is to be understood in terms of the practice of sanctioning. Giving pride of place in this way to normative attitudes in the understanding of normative statuses involves emphasizing the distinction of perspective between *assessing* a performance and *producing* that performance. Any theory that reconstructs Kant's distinction by appealing to this difference of perspectives on normative significance—the difference, namely, between consuming and producing normatively significant performances—is in one important sense a *social* theory of the sort of norm-governedness distinctive of us. Haugeland's censorious herd animals shape *each other's* behavior by their capacity not only to perform but to censure performance. Each animal in the community that is thereby constituted may (and perhaps to be a full-fledged member must) be able to do both, but as he conceives it, each act of censure involves two organisms, the censoring and the censored, the reinforcer and the reinforced.⁴⁵

There is another sort of theory that combines the idea that normative statuses cannot be understood apart from normative attitudes with the idea that the relation between them involves social practices in a different but perhaps even more robust sense of 'social'. It has often been noticed that simple regularity theories of implicit norms gain no ground by shifting from concern with regularities characterizing the behavior of individuals to regularities characterizing the behavior of groups.⁴⁶ As a response to this concern, the leading idea of this sort of the construal of norms as implicit in social practices is that of *communal assessment*. On this approach, the key to the importance of the social is taken to lie in the possibility that the performances individual community members produce are assessed, responded to, or treated in practice as appropriate or inappropriate by the community to which the individual belongs. An individual might be taken implicitly to endorse or treat a performance as correct simply by producing it. The community, unlike the individual, need not be counted as having taken up a practical attitude regarding the propriety of the performance just in virtue of that performance's having been produced by one of its members. The class of performances produced by its members, rather, determines which fall within the scope of communal attitudes, which are liable to communal endorsement or repudiation.

Looking to assessments by the community does provide further resources for regularity theories. For a regularity can govern (that is, be displayed or conformed to by) the assessments of the community without each individual performance counting as correct according to it, and in that case a distinction between correct and incorrect performances by individuals is underwritten. In this way, the communal assessment theories are just like the theories already considered, with Haugeland's as an example, in which norms are taken to consist in regularities of practical appraisal. (It is irrelevant for the present point whether such appraisals are then understood in terms of sanctions, and the sanctions in terms of reinforcement.) The difference is that where the theories previously considered look to regularities of appraisals by individuals, the approach now on the table looks to regularities of appraisals by the community as a whole.⁴⁷

There are two sorts of objections not yet considered to accounts that construe the norms implicit in practices in terms of regularities or dispositions regarding communal assessments or attributions of normative significance. First, the idea of communal performances, assessments, or verdicts on which it relies is a fiction. Second, the approach smuggles normative notions illicitly into what purports to be a reductive, nonnormative regularity theory. On the first point, communal assessment theorists have a tendency to personify the community, to talk about it as though it were able to do the same sorts of things that individual community members can do—perform additions, apply rules, assess performances, and so on. Thus to pick one page almost at random, Kripke talks about “the community's accepting” conditionals codifying relations between attributions of intentional states and commitments to act, what “the community regards as right,” what “the community endorses,” and so on.⁴⁸ This is a typical passage from Wright: “None of us can unilaterally make sense of the idea of correct employment of language save by reference to the authority of securable communal assent on the matter.”⁴⁹

The difficulty with this way of talking is that assenting, endorsing, accepting, and regarding as right are in the first instance things done by *individuals*. It is not the community as such that assesses applications of the concept *yellow*, say, but individual members of that community. Any account that seeks to extend these notions to include cases where the subject or agent is a community should say explicitly what sort of performance or speech act is envisaged. Some communities have meetings, authorized representatives, or other ways of officially settling on a communal view or act, for instance of disapproval or endorsement. But this sort of thing is the exception and could in any case hardly be what is wanted for explaining either norms in general or conceptual, intentional, or linguistic norms more particularly.

This tendency to talk of the community as somehow having attitudes and producing performances of the sort more properly associated with individuals is neither accidental nor harmless. This *façon de parler* is of the essence of

the communal assessment approach. It is a manifestation of the orienting mistake (about which more will be heard later) of treating *I-we* relations rather than *I-thou* relations as the fundamental social structure.⁵⁰ Assessing, endorsing, and so on are all things we individuals do and attribute to each other, thereby constituting a community, a 'we'. But this insight is distorted by *I-we* spectacles—perhaps the same that have always been worn by political philosophers in conceiving their topic. The pretense of communal assessment is not harmless because the easy ways of cashing out the metaphor of community approval and so forth present familiar dilemmas. A notion of communal endorsement or repudiation might be built out of regularities of endorsement and repudiation by individual community members. But universal agreement is too much to ask, and how is it to be decided what less ought to count?

In fact, the approval of some almost always matters more to the community than that of others—though this division may be different for different issues. We recognize experts. But being an expert is having a certain *authority*, and that is a normative matter. One might go on to give an account of the status of having the authority to speak for the community on some matters in terms of the attitudes of attributing or recognizing that authority on the part of other community members. (A story along these lines is endorsed in what follows.) One might go on to offer sufficient conditions for the attributions of such attitudes, and so such statuses, to a community, entirely in nonnormative terms. (This move is repudiated in what follows.) But without some such story, how is one to understand talk of what the community endorses or repudiates?

Connected with this question is the problem of how community membership can be understood, in line with a communal assessment theory. This in turn is closely related to the previously mentioned problem, peculiar to regularity theories, of distinguishing those in whose practice the norms are implicit from those on whom those norms are binding. If 'norm implicit in a practice' is understood just as 'regularity truly descriptive of actual performances (or performances there are dispositions to produce under suitable circumstances)', and those performances are also the ones 'subject to' or 'governed by' the practices comprising them, then there is no possibility of irregularity, of violating a norm. Being a member of a community is rather being one who *ought* to conform to the norms implicit in the practices of the community. Community membership has this normative significance; it is a normative status.

Unless this status is understood in some way other than as being one who in fact exhibits the regularity in question, it will be impossible to violate a norm, because impossible to act irregularly. I cannot be out of step with any regularity that characterizes the behavior of each and every one of us. Extending that argument, if an account is to be offered of norms as social regularities, it is not by itself enough to identify what is correct for *all*

community members with what accords with the regularities descriptive of the practice of *some* of those community members (the experts). Some account is required of how those members are picked out. The distinction between the experts, the ones who have authority, whose actions and assessments cannot fail to be correct, and those who are subject to that authority, bound by the norms instituted by the regularities of performance of the experts, is a normative distinction, a distinction of normative status. Unless this distinction is itself understandable as a matter of regularities descriptive of the performance of community members, the social regularity account explains only some normative notions, while appealing to others.

Just as having the status of a community member is something with normative consequences—for one thereby is *subject* to a certain sort of authority, a certain standard of correctness—so having the status of an expert or official is something with normative consequences; for one thereby *exercises* a certain sort of authority. It is appropriate to ask whether the circumstances of application of the concept 'expert' (in the sense of authoritative assessor) could be specified in terms of regularities of conduct specified in nonnormative terms. Is there a distinction between actually assessing and being entitled to assess, one's assessments having authority? It is just in this sense that it is appropriate to ask more generally whether the circumstances of application of the concept 'community member' can be specified in terms of nonnormative regularities. Is there a distinction between actually being assessed and being properly assessed—being subject to the authority of assessments?

The point is that talk about the community to which a performer belongs is not obviously translatable into talk about regularities of individual performance. Belonging to the community is a concept used so as to have normative consequences of application, concerning the member's being responsible to the assessments of the community, being subject to its authority. An understanding of norms implicit in practice in terms of regularities of communal assessment requires the idea of a regularity that the performer is somehow bound by or answerable to. Exactly how must the performer be related to other performers to be appropriately taken to be responsible to their assessments?⁵¹ It is possible that appeal will have to be made to something other than regularities of performance to secure this connection. Understanding normative status, including the normative status of being a community member, in terms of some sort of liability to being rewarded or punished is one thing; understanding that liability nonnormatively is another. The claim is not that these difficulties—with specifying what it is for a community to endorse or repudiate a performance, and the related questions of how to pick out the community members, those over whom such assessments have authority, and perhaps how to pick out experts whose assessments have such authority—are insuperable. These difficulties show only that the invocation of communal assessments does not by itself provide

a recipe for construing norms implicit in practices in terms of nonnormative regularities of performance (even assessing performances).

To recapitulate: Regularity theories of norms implicit in practice, that is of practical normative status, no matter whether they invoke specifically *social* regularities, or look in addition at regularities of *attitude* or assessment, or combine those moves in a theory that relies on regularities of communal assessment, are subject to two general sorts of objection. First, the gerrymandering argument challenges them to pick out a unique specification of the regularity in question, sufficient to project it so as to determine the application of the norms to novel cases. Second, bracketing those difficulties, attempts to define normative statuses in terms of nonnormative regularities can be criticized from two directions, as either failing to reconstruct some essential features of genuine normative statuses, or as covertly appealing to normative notions.⁵² Under the first heading, they may fail to make room for the crucial distinction between *performance* and the *normative status* or significance of a performance, that is between what is done and what is correct or ought to be done, or they may fail to make room for the equally crucial distinction between the *normative status* of a performance and a *normative attitude* toward that status, that is between what is correct and what is taken to be correct. (More will be said about this second sort of insufficiency below.) The danger of regularity theories smuggling in normative notions arises both in specifying what it is for someone to be a member of a community in the sense of being governed by its norms, properly subject to assessments according to them, and again in specifying what it is for something to count as a communal attitude, an assessment by a community of conformity with its norms. This latter may take the form of specifying a subclass of community members whose assessments are imbued with communal authority and so have the status of official or expert assessments.

This mention of one sort of social practice theory is entirely preliminary. The issue of reasons for understanding the sort of practice in which norms are implicit as *social* practice, and in what sense of 'social', is addressed further along. It is one of the prime tasks of this work to elaborate a suitable notion of the social practices that institute the norms underlying explicitly propositional attitudes. For the approach eventually to be endorsed concerning the social nature of those implicit norms to be intelligible, something must be said here about the relation between normative and nonnormative vocabularies. Regularity theories, of whatever stripe, are (to adapt a phrase of Dretske's) attempts to bake a normative cake with nonnormative ingredients. Gerrymandering aside, the objections just mentioned represent the two ways such an enterprise can go wrong: by failing to produce a genuinely normative product or by employing some already normative raw materials. The discussion here of accounts that identify norms implicit in social practice with regularities of communal assessment arose from consideration of ways of pursuing the strategy that starts with Kant's distinction between

acting according to rules and acting according to conceptions of rules, and adapts it to accommodate Wittgenstein's pragmatic point about norms by rendering the normative attitude of taking or treating as correct (which according to the Kantian line is essential to the characteristic sort of normative status we exhibit) in terms of sanctions, of reward for what is (thereby) taken to be correct and punishment for what is (thereby) taken to be incorrect. The sanctions approach, the broadly retributive rendering of attitudes of assessment, offers aid and comfort to those concerned to tell normative stories with nonnormative vocabulary.

The animating idea is that the classification of performances needed to get this two-stage scheme off the ground, their being positive or negative sanctions, is something that can be made available in perfectly naturalistic terms. One might try to define the two sorts of sanctions in terms of the production of benefit versus harm to the one whose status or performance is being assessed. Or one might try to define them in terms of the preferences and aversions of the one assessed—respecting, as it were, the views of the sanctioned one rather than the sanctioner, as to what counts as benefiting and harming the sanctioned one. Less subjectively, reward and punishment might be understood in terms of giving pleasure and inflicting pain. Most austere, one might define rewarding or punishing a certain kind of behavior or performance in functional terms, as positively or negatively reinforcing the reliable dispositions to respond differentially to stimuli that are being realized by the performances responded to. This version takes it that to treat a response to a certain stimulus as incorrect is just to punish it, in the sense of responding to it in a way that in fact decreases the probability that the one being assessed will respond in that way to that sort of stimulus in the future.

5. Normative Sanctions

In this connection it is important to realize that it is one thing to understand practical assessment as sanctioning, and quite another to understand sanctioning in nonnormative terms such as reinforcement. A retributive approach to the normative need not be given a naturalistic turn at all. Defining normative attitudes in terms of dispositions to apply sanctions does not by itself reduce the normative to the nonnormative—it just trades off one sort of norm for another. At the most basic level, to reward someone is to offer some good (either objectively or subjectively), and to punish them is conversely to inflict something bad. Benefit and harm, desirable and undesirable, are concepts that also have normative senses. Indeed, these senses would seem to be primary, so that some sort of reductive hypothesis would be needed to naturalize them. To turn the retributive story about normative attitudes into a naturalistic one, an account might for instance understand what is good (and so rewarding) in terms of what is desirable, what is desir-

able in terms of what is desired, and what is desired ultimately in terms of what is pursued.

Commitment to such a reduction is optional. Positive and negative sanctions may consist in acclaim and censure that itself has only a normative significance. A correct action might be rewarded by the grant of some extraordinary privilege or by release from some onerous obligation—and the status of such a response as reward need not depend on whether the one rewarded would in fact have been disposed to refrain from acting without the boon of entitlement or would in fact have been disposed to act so as to fulfill the obligation had it not been lifted. An incorrect action might be punished by withholding a license to act in certain other ways or by imposing an extraordinary obligation—and the status of such a response as punishment need not depend on whether the one punished is in fact disposed to refrain from acting even without the boon of entitlement, or is in fact disposed to act so as to fulfill the obligation imposed. In such cases one is rewarded or punished for what one does “in another world”—by a change in normative status rather than natural state.⁵³

Consider once again the case of discerning a practical norm in force in a community, according to which to enter a particular hut one is obliged to display a leaf from a certain sort of tree. As pointed out above, the assessing response constituting the community’s acknowledgment of such a norm (the attitude corresponding to the status) might in some cases be describable in nonnormative terms—one who violates the norm is beaten with sticks, the norm-violating behavior is negatively reinforced. But other cases are possible, for instance ones in which the assessing response is to punish by making other actions inappropriate—one who violates the norm is not permitted to attend the weekly festival. In such a case, the normative significance of transgression is itself specified in normative terms (of what is *appropriate*, of the transgressor is *entitled* to do). The punishment for violating one norm is an alteration in other normative statuses. Acting incorrectly alters what other performances are correct and incorrect.

Once again, it need not be assumed that the alteration of status according to which it becomes inappropriate to attend the tribal festival has the actual effect of disposing the transgressor not to attend. The alteration of status need itself have no reinforcing function. This could be so even if the assessing attitude corresponding to the consequential norm forbidding attendance at the festival is itself enforced by actual reinforcing responses—that is, even if it is the case that an attempt to attend the festival by one who is not entitled will be punished by beating the offending community member with sticks. In such a case, the norm regarding entitlement to attend the festival is intelligible in terms of attitudes expressed by sanctions specifiable in non-normative terms, while the norm regarding entitlement to enter the hut is intelligible in terms of attitudes expressed by sanctions specifiable only in

normative terms of consequences for entitlement to attend the festival. Since the norms governing festival attendance are directly nonnormatively intelligible, the norms governing hut entrance are also nonnormatively intelligible—but indirectly, at one remove. In this way one norm can depend on another, as the sanctions expressing assessments of the normative significance of performances according to the first norm consist in alterations of normative status with respect to the second norm.

If what qualifies some response to a performance as a sanction—and therefore, according to the retributive line being considered, as an assessment—is specifiable only in normative terms, that is in terms of the correctness or incorrectness, (the normative status) of further performances according to other norms, that kind of sanction can be thought of as being *internal* to the system of norms being discerned.⁵⁴ If, by contrast, what qualifies a response as a sanction is specifiable in wholly nonnormative terms of what various community members do or are disposed to do, without reference to the specifically normative status of their performances, that kind of sanction can be thought of as being *external* to the system of norms being discerned. In this terminology, the simple sort of dependence of one norm on another just considered occurs when the attitudes corresponding to one kind of normative status (e.g. propriety of hut entering) are expressed by normatively internal sanctions, defined in terms of another sort of normative status (e.g. propriety of festival attending). In the case described, that second sort of status is itself made intelligible by normatively external sanctions, ones that can be specified in terms of the movement of sticks and consequent alterations in dispositions to attempt festival attendance. But internal sanctions can be defined in terms of normative statuses that themselves are defined by internal sanctions referring one to still further norms. Clearly this sort of dependence of one norm on another according to the retributive paradigm can be extended and ramified, making sense of complex webs of interdependent normative statuses.

In the cases so far imagined, these webs of norms linked by internal sanctions are anchored, as each chain of definitional dependence terminates in some normative status that is definable independently, by external sanctions specified in nonnormative terms. Even this restriction can be relaxed. The consequences of an assessment of a performance as correct or incorrect with respect to one norm may extend no further than other assessments of correctness, with respect to other norms. It is possible to interpret a community as instituting normative statuses by their attitudes of assessment, even though each such status that is discerned is responded to by sanctions that involve only other normative statuses. It is compatible with the sanctions paradigm of assessment, and so of normative attitude, that it should be “norms all the way down.” Such an interpretation would not support any reduction of normative status to nonnormatively specifiable dispositions, whether to perform or to assess, whether individual or communal.

Norms acknowledged by external sanctions can be attributed to a community one by one, in an atomistic way. Attribution of any norm whose acknowledgment by the community takes the form of assessments expressed by internal sanctions, however, commits the interpreter to attributing also the norms on which it depends. Such dependences introduce a holistic element into the attribution of normative significances to the performances of a community. Using the retributive paradigm of normative attitudes of assessment to structure an interpretation of a community as exhibiting practices in which interdependent norms are implicit does not require that there be some nonnormatively specifiable behavior associated with the acknowledgment of each sort of normative status or significance discerned by that interpretation. An example of a system of practices in which the normative significances of performances must be attributed in a holistic way and are not translatable into nonnormatively specifiable dispositions is offered in Chapter 3, which presents sufficient conditions for such a system of practices to confer *propositional* contents on the statuses and attitudes it institutes.

A strategic divide looms here. Wittgenstein argued that an unproductive regress results from conceiving explicit rules as the only form of the normative. The lesson drawn from the regress of rules interpreting rules is the pragmatic one, that there must be "a way of grasping a rule which is not an interpretation, but which is exhibited in what we call 'obeying the rule' and 'going against it' in actual cases."⁵⁵ That is, there must be such a thing as norms that are implicit in practice. But what is the relation between such norm-laden practices and nonnormatively describable regularities of performance? The division of explanatory strategies arises over the question of whether the practices invoked to halt the regress⁵⁶ can be analyzed in terms of regularities and dispositions characterized without the use of normative vocabulary.

In line with Kant's insight that normative *attitudes*—the sort of uptake of or sensitivity to norms that he talked about in terms of *conceptions* of rules—are essential to the way in which our conduct is governed by norms, two suggestions have been put on the table.⁵⁷ The first is the idea of construing the normative attitude of taking or treating something as correct or incorrect in practice in terms of the application of positive and negative *sanctions*. The second is the idea that these assessing attitudes have a fundamentally social structure, so that the practices in which norms are implicit ought to be understood as essentially *social* practices. Each of these ideas could be given a reading in naturalistic or nonnormative terms, as part of a reductive explanatory strategy. Putting them together would then yield an approach that understands norms as implicit in regularities or dispositions regarding communal assessments of performances as correct or incorrect, and that understands such assessments in turn as behaviorally reinforcing rewards and punishments.

Wittgenstein certainly emphasizes the social nature of the practices un-

derlying the norms involved in discursive intentional states, and the importance to those practices of similarity or agreement of dispositions to respond to performances as correct or incorrect. There is accordingly a temptation to understand him as responding to the regress-of-rules argument by putting forward a reductive social regularity account of the practices in which norms are implicit.⁵⁸ But he need not be understood this way. His insistence that unless the responsive dispositions of a community are consilient, there can be no proprieties of practice is a point concerning presupposition, not reduction. Wittgenstein's is the somewhat delicate position first, that the usefulness of normative attributions, the viability of this stratum of discourse, presupposes a variety of regularities of performance and disposition; second, that those regularities obtain is *not* part of what is *asserted* by such attributions.

An analogy he comes back to again and again is the measurement of length, in which the possibility of practices of measurement presupposes features of the world such as the rigidity, spatial invariance under transportation, and temporal constancy of measuring rods, interpersonal comparability of measurements, the functional equivalence of various means of measuring the same length, the irrelevance to the result of such contextual features as whether the object measured is sacred or profane, to be used in sport or commerce, and so on. That we can be trained so as almost always to respond in the same way when applying concepts to novel cases, for instance, is a necessary condition of there being a practice determining what response is correct in such cases. But this is not to say that what it is for it to be correct consists in this agreement, as the reductive social regularity account of those norms would have it. There are three levels at which performances can be discussed: a level of norms explicit in rules and reasons, a level of norms implicit in practice, and a level of matter-of-factual regularities, individual and communal. To say that various claims made at the third level state necessary conditions for the applicability of vocabulary of the sort employed at the first two is not to make a reductive claim. The social regularity view conflates the second and the third levels, and so misunderstands Wittgenstein's remarks about the significance of matter-of-factual regularities, by taking them to involve commitment to the possibility of a reduction of the normative to the dispositional.⁵⁹

V. FROM ASSESSMENT TO THE SOCIAL INSTITUTION OF NORMS

1. *Pufendorf on the Institution of Norms by Attitudes*

As discursive beings whose characteristic activities are applying concepts, giving and asking for reasons, taking-true and making-true, we live and move and have our being in a space structured by norms. Yet we can describe, and largely successfully cope with, the not-us around us, while

restricting ourselves to a resolutely nonnormative vocabulary. In thinking about the relation between acting according to conceptions of rules (on a suitably pragmatic reading of what that consists in), as *we* do, and merely acting according to rules, as the rest of *it* does, it is important to distinguish two ways in which the normative significances we assign to things might be thought to be unnaturalized second-class citizens in an intrinsically insignificant natural world. These correspond to two different sorts of domestication to which normative statuses might be subjected. Couched in terms of supervenience, they are the claim that settling all the facts specifiable in *nonnormative* vocabulary settles all the facts specifiable in *normative* vocabulary, on the one hand, and the claim that settling all the facts concerning normative *attitudes* settles all the facts concerning normative *statuses*, on the other.

These are intimately related claims; the difference between them is subtle, and they are often run together. Each is the heir to a line of thought central to and characteristic of the Enlightenment project of disenchanting the natural world and humanizing values. The first can trace its origins to atoms-in-the-void physicalism—the conviction that a specification of the values of an appropriate range of dynamic variables for all the fundamental particles provides a *complete* description of everything that deserves to be called *real*. The second is animated by the humanistic thought that the merely natural world is devoid of values, that the worth of things and the fitness of actions is a product of our activity—that unlike natural properties, normative proprieties are in the eye of the human beholder. These ideas are of course at work in the thought of many Enlightenment philosophers. For present purposes it will suffice to consider briefly representative statements by one of the earliest.

The second line of thought emerges most clearly in the thought of the pioneering philosopher of law Samuel (Freiherr von) Pufendorf (1632–1694). Although not much read by philosophers today, his magnum opus *De Jure Naturae et Gentium* stands at the beginning of a tradition of Enlightenment thought about norms that culminates in Kant's practical philosophy (which was greatly influenced by Pufendorf). The relative unfamiliarity of these seminal views perhaps excuses quotation at greater than usual length. These passages all come from the opening of the work, in Chapter 1, entitled "On the Origin and Variety of Moral Entities": "It is for us to observe, how, chiefly for the direction of the acts of the will, a specific kind of attribute has been given to things and their natural motions, from which there has arisen a certain propriety in the actions of man, and that outstanding seemliness and order which adorn the life of men. Now these attributes are called Moral Entities, because by them the morals and actions of men are judged and tempered, so that they may attain a character and appearance different from the rude simplicity of dumb animals."⁶⁰ We are distinguished from the brutes by the fact that our actions are subject to assessment according to their

propriety, a special kind of attribute over and above the natural motions of things, an attribute that has somehow been *given* to them.

What is the source of those special normative attributes? "We seem able, accordingly, to define moral ideas most conveniently as certain modes [qualities], added to physical things or motions, by intelligent beings, primarily to direct and temper the freedom of the voluntary acts of man, and thereby to secure a certain orderliness and decorum in civilized life."⁶¹ Intelligent beings add these properties to things by their activities. They are called 'moral' entities in virtue of their practical function as guides to action.

Now as the original way of producing physical entities is creation, so the way in which moral entities are produced can scarcely be better expressed than by the word *imposition*. For they do not arise out of the intrinsic nature of the physical properties of things, but they are super-added, at the will of intelligent entities, to things already existent and physically complete, and to their natural effects, and, indeed, come into existence only by the determination of their authors. And these authors give them also certain effects, which they can also remove at their own pleasure without any accompanying change in the object to which they had been added. Hence the active force which lies in them does not consist in their ability directly to produce any physical motion or change in any thing, but only in this, that it is made clear to men along what line they should govern their liberty of action.⁶²

These norms are not part of the intrinsic nature of things, which is entirely indifferent to them. They are imposed by the will of intelligent beings and can affect things only through their effect on the will of such beings—beings who can act according to a conception of them. "Since, therefore, moral entities have been instituted to bring order into the lives of men, for which purpose it is required that they also, who must live according to their rule, should adopt a set standard in their relations toward one another, in determining their actions, and finally in fixing their attitude toward those things which are used in the lives of men; for this reason they are understood to be inherent primarily in men, but also in their actions, and even, to some extent, in things."⁶³ Our activity *institutes* norms, *imposes* normative significances on a natural world that is intrinsically without significance for the guidance or assessment of action. A normative significance is imposed on a nonnormative world, like a cloak thrown over its nakedness, by agents forming preferences, issuing orders, entering into agreements, praising and blaming, esteeming and assessing.⁶⁴

One of the defining characteristics of early science is its disenchantment (*Entzauberung*, in the word we owe to Weber) of the world. The meanings and values that had previously been discerned in things are stripped off along with the supernatural and are understood as projections of human interests, concerns, and activities onto an essentially indifferent and insignificant mat-

ter. The Enlightenment disenchantment of the world and its assignment to us of responsibility for the norms, values, and significance we nonetheless find in the world are two sides of one coin. Meaningless objects and meaning-generating subjects are two aspects of one picture.⁶⁵ On this view, valuing is the source of values—a tradition carried on by figures as disparate as Mill and Nietzsche. Contractarian theories, and those that invoke positive law to explain various rights and obligations, are species of this genus. Each explains these deontic statuses in terms of what agents are doing in instituting or constitutively recognizing such entitlements and commitments.

Pufendorf does not suggest, and he does not believe, that the activity by which we institute norms is itself describable in the purely physical terms that suffice to describe the antics of merely natural objects. His claim is that the normative statuses of things, the normative significances we take them to have, are products of our practical normative attitudes, as expressed in our activity of imposing those significances and acknowledging them in assessments.⁶⁶ He does not conjoin this thesis with any sort of physicalism about the mechanism by which these moral secondary qualities arise from our practical activity. It is clearly possible to agree with the dictum of another Enlightenment thinker, Hamlet, that "There is nothing either good or bad, but thinking makes it so," without holding in addition a reductionist view about such thinking.

One who does conjoin these commitments is Hobbes. He explains, not good and evil, but *calling* things good and evil. He understands the use of the words 'good' and 'evil' as expressing appetites, desires, or aversions. He expresses his commitment to fundamental normative statuses being instituted by our attitudes this way: "But whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth *Good*: And the object of his Hate, and Aversion, *Evill* . . . For these words of Good, Evill . . . are ever used with relation to the person that useth them: There being nothing simply and absolutely so; nor any common Rule of Good and Evill, to be taken from the nature of the objects themselves; but from the Person of the man."⁶⁷ Gauthier comments on one difference between the sort of view Hobbes endorses concerning the relevant norm-instituting practical attitudes and the kind of view Hamlet endorses on this point: "If things considered in themselves are neither good nor bad, if there is no realm of value existing independently of animate beings and their activities, then thought is not the activity that summons value into being . . . Desire, not thought, and volition, not cognition, are the springs of good and evil."⁶⁸

Pufendorf would not disagree. Where he does disagree is with Hobbes's subsequent endorsement of materialism about the will. The claim that normative statuses are instituted by our normative attitudes entails the claim that the normative proprieties so instituted are just natural properties of a special kind only in the context of a collateral claim that the norm-instituting practical attitudes can themselves be specified in nonnormative terms.

One can hold, as Pufendorf does, that there are no values apart from our acknowledgment or recognition of them, or more generally our attitudes toward them, without being thereby obliged in addition to understand those attitudes in terms of desires or preferences that can be characterized independently in value-free terms.

2. *Kantian Autonomy: The Authority of Norms Derives from Their Acknowledgment*

Pufendorf's idea that normative statuses are *instituted* by our practical attitudes makes a stronger claim than the idea previously extracted from Kant's demarcation of us as beings who act not only according to rules but according to our conceptions of rules. For the latter idea requires only that the normative statuses of demarcational interest essentially *involve* the uptake or grasp of such statuses, that is, our practical attitudes toward them. But normative statuses could be taken to be unintelligible apart from normative attitudes without thereby being taken to be instituted by and therefore in some sense to supervene on those attitudes. However Kant does in fact subscribe also to a version of the stronger thesis about attitudes instituting statuses, for the case of the genuinely moral normative statuses characteristic of us as agents.⁶⁹

Kant's practical philosophy, his account of us as agents, takes its characteristic shape from his dual commitments to understanding us as *rational* and as *free*. To be rational, for him, means to be bound by rules. But Kant is concerned to reconcile our essential nature as in this way bound by norms with our radical autonomy. He combines the essential defining moment of our dependence on universals with that of our independence as particulars (as Hegel puts the point) in the thesis that the *authority* of these rules over us derives from our *acknowledgment* of them *as* binding on us. Our dignity as rational beings consists precisely in being bound only by rules we endorse, rules we have freely chosen (like Odysseus facing the Sirens) to bind ourselves with. We do not have the freedom to opt out entirely—choosing to be bound by no rules at all would be choosing to relinquish our rationality entirely. Yet if something other than our own attitudes and activity could bind us, we would not be free. Autonomy consists, as the etymology demands, in setting up laws for ourselves.

This view of Kant's inherits a venerable Enlightenment tradition.⁷⁰ It is based on a certain picture of the nature of the *authority* of rules or laws (the only form of norm considered). Pufendorf takes it that "since good repute, or moral necessity, and turpitude, are affections of human actions arising from their conformity to some norm or law, and law is the bidding of a superior, it does not appear that good repute or turpitude can be conceived to exist before law, and without the imposition of a superior."⁷¹ More generally: "A

law may most conveniently be defined as a decree by which a superior obligates a subject to adapt his actions to the former's command."⁷²

The key notion is that of a *superior*, someone who has *right* to command.

An obligation is properly laid on the mind of man by a superior.⁷³

The power of obligating, that is, the faculty of laying an intrinsic necessity on persons to do something, properly lies in him who has authority or sovereignty.⁷⁴

To see the question of authority in terms of who can command, make a rule binding, or lay down the law is another bit of fallout from the origin of thought about norms in thought about the institution of explicit positive laws. The issue of sovereignty is just the issue of "who's to be master, that's all," as the linguistic Leninist, Humpty Dumpty, says. The authority of norms depends on the nature of the author of the commands that make them explicit; their bindingness derives from the interpersonal relation of superior to subordinate.

The *consequences* of such a relation of authority being in force, what follows from a rule or law inheriting the authority of a superior lawgiver, Pufendorf conceives in terms of sanctions.⁷⁵ The *antecedents* or grounds on which such a relation is based, what makes one individual superior to a subordinate other in this normative sense, he conceives disjunctively: "Mere strength [to sanction] is not enough to lay an obligation on me at the desire of another, but that he should in addition have done me some special service, or I should of my own accord consent to his direction . . . But when a man of his own accord consents to the rule of another, he acknowledges by his own act that he must follow what he himself has decided."⁷⁶ The "special service" clause is put in to allow our creator a special claim on our obedience.

That special pleading aside, the standard Enlightenment thought, common as well to Hobbes and Rousseau among the progenitors of Kant's theory, is that our own *acknowledgment* or *endorsement* of a rule is the source of its authority over us—in short that our normative statuses such as obligation are instituted by our normative attitudes. Authority is not found in nature. The laws of nature do not bind us by obligation, but only by compulsion. The institution of authority is human work; we bind ourselves with norms. Contract theories are the result of combining a conflation of norms with their explicit expression in rules or laws, an understanding of their authority or bindingness on the hierarchical model of superior/subordinate (each given aid and comfort by the tradition of legalism), and an insistence on rational dignity as demanding autonomy.

Kant's reconciliation of us as free in virtue of being rational, with us as bound by norms in virtue of being rational—and so of freedom as constraint by a special kind of norm, the norms of rationality⁷⁷—accordingly involves treating the normative status of moral obligation as instituted by normative

attitudes. It is our attitude toward a rule, our acknowledgment or recognition of moral necessity alone, that gives it a grip on us—not just in terms of its effect on our actual behavior, but in terms of our liability to assessment according to the rule that expresses that necessity. In this sense the norms that bind us rational creatures are instituted by our practical attitudes and activity. They are what we bring to the party. But while Kant in this way endorses the supervenience of moral normative status on moral normative attitude, he does not endorse any sort of naturalism or reductionism about those attitudes. He does not take it that specifications of those normative attitudes supervene on specifications of the movements of particles, described exclusively in the vocabulary of natural science. Grounding normative status in normative attitude does not entail relinquishing the distinction between normative proprieties and natural properties.

3. *Objectivity and the Social Institution of Conceptual Norms*

Kant also takes it that we are genuinely bound by the rules we endorse. This requires that once we endorse one, it is not up to us what it demands—there is some fact of the matter as to what we have thereby obliged ourselves to do. Although the status of being obliged to follow a particular rule is instituted by our attitudes, what is correct according to that rule is not simply determined by what we *take* to be correct according to it. The status of correctness of a performance according to a rule does not collapse into the attitude of assessing that performance as correct. Endorsing a rule gives it a grip on us. Part of that grip is that the rule does not mean just whatever we later might take it to mean. So Kant underwrites not only the possibility of *mistakes of performance*, which was already claimed to be essential to there being norms in play, but also the possibility of *mistakes of assessment*.

Wittgenstein appeals to this possibility as a criterion of adequacy for an account of norms being in force; talk of norms being implicit in a practice requires that there be room for a distinction between what is correct according to the norm and what the one whose performances are being assessed *takes* to be correct. He brings this consideration to bear against the possibility of setting up rules for oneself whose meaning is determined *only* by one's own actual dispositions to make assessments. This is the line of thought that concludes: "One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'."⁷⁸ The thought here is that the distinction between status and assessment (the attitude of taking or treating a performance *as* correct, appropriate, or in order) is essential to the notion of genuinely normative status. It is motivated by the idea that assessing is itself something that can be done correctly or incorrectly, and furthermore that it is the norm according to which perform-

ances are being assessed that determines which assessments are correct or incorrect. If there is no distinction to be made between correct and incorrect assessments, then there is no sense in which the performances being assessed are governed by a norm *according to which* they are being assessed.

One of Wittgenstein's most important claims is that the practices in which the norms that articulate meanings and their uptake in understanding are implicit must be *social* practices. It is clear that the emphasis on their social character emerges for him somehow from the need to keep the notion of what one is committed to by the application of a concept distinct in principle from what one *takes* oneself to be committed to thereby. One natural way of understanding how such considerations can lead to the conclusion that discursive practice must be social practice is elaborated by Crispin Wright.⁷⁹ He combines an understanding of conceptual norms as instituted by practical normative attitudes—taking or treating various uses as correct or incorrect—with a way of maintaining a distinction between conceptual commitments and individual assessments of those commitments. He does so by identifying the normative status of being a correct application of a concept with being taken to be such a correct application, not by an individual, but by the whole community. According to this story, although individual performances can be correct or incorrect, and assessments of correctness by individuals can be correct or incorrect, no such difference applies to communal assessments. The community is incorrigible about what is a proper application of a concept and what is not. Communally endorsed applications of a rule, or, in the idiom employed here, acknowledgments of a norm implicit in the practice of the community, cannot be mistaken. "For the community itself there is no authority, so no standard to meet."⁸⁰

Thus Wright secures a distinction between the commitments one undertakes in employing a particular concept and any individual's attitudes toward or assessments of those commitments, but at the cost of obliterating any such distinction between normative status and the attitudes of the whole community. There clearly are socially instituted norms of this sort. Whatever the Kwakiutl treat as an appropriate greeting gesture for their tribe, or a correctly constructed ceremonial hut, is one; it makes no sense to suppose that they could collectively be wrong about this sort of thing. The question is whether *conceptual* norms ought to be understood as being of this type. There is good reason to think they ought not. It is a fundamental feature of our understanding of our concepts that they incorporate *objective* commitments. Thus, our use of the term 'mass' is such that the facts settle whether the mass of the universe is large enough that it will eventually suffer gravitational collapse, independently of what we, even all of us and forever, take those facts to be. We could all be wrong in our assessment of this claim, could all be treating as a correct application of the concepts involved what is objectively an incorrect application of them.

On Wright's view the objectivity we take our conceptual norms to have is

an illusion that must be surrendered if they are to be properly understood. The normative attitudes discussed here under the heading of assessment—taking or treating applications of concepts as correct or incorrect—he talks about in terms of “ratification.” Understanding conceptual norms as objective, in the sense that the whole community can coherently be conceived to be wrong in its assessment of the commitments involved in some applications of its concepts, is taking them to be ratification-independent, in his terminology. McDowell, whose insistence on the objectivity of conceptual norms was discussed in this connection in Section IV, summarizes his disagreement with Wright on this point in this way: “In Wright’s reading . . . Wittgenstein’s point is that the natural contractual conception of understanding should not be discarded, but purged of the idea—which it must incorporate if the intuitive notion of objectivity is to have application—that the patterns to which our conceptions oblige us are ratification-independent. I expressed a suspicion above that this purging would not leave a residue recognizable as a conception of meaning and understanding at all.”⁸¹ Wright takes it that understanding how the normative statuses involved in concept use are instituted by practical normative attitudes of assessing or ratifying the propriety of particular applications of concepts, while keeping normative statuses from collapsing into normative attitudes in a way that obliterates the norms entirely, at once requires understanding the practices of concept use and its assessment as *social* practices and relinquishing the idea that conceptual norms are *objective*.

A central aim of the present study is to show, by contrast, how these criteria of adequacy can be satisfied without giving up the objectivity of conceptual norms. Indeed the primary explanatory challenge to a social practice theory of discursive commitments is to show how, starting with the sort of norms for which Wright’s analysis is correct—normative statuses about which the community’s all-inclusive practical assessment cannot be mistaken, such as who is really married or what obligations are incurred by spitting in front of the chief—genuine, and therefore *objective*, conceptual norms can be elaborated. These bind the community of concept-users in such a way that it is possible not only for individuals but for the whole community to be mistaken in its assessments of what they require in particular cases.

How does objectivity precipitate out of the social soup of norms that are whatever the community takes them to be? According to the answer elaborated in Chapter 8, it is precisely the objectivity of conceptual norms, when properly understood, that leads to the requirement that the practices in which such norms are implicit be *social* practices. The objective representational dimension of conceptual content—the kind of correctness of claiming or concept application that answers not to individual or communal attitudes or assessments but to the properties of the things represented—turns out to depend on the *social* articulation of the inferential practice of giving and asking for reasons. Focusing on the distinction of social perspec-

tive between *acknowledging* (and thereby undertaking) a commitment oneself and *attributing* a commitment to another makes it possible to understand the objectivity of conceptual norms that consists in maintaining the distinction between the normative *statuses* they incorporate and the normative *attitudes* even of the whole community—while nonetheless understanding those statuses as instituted by the practical normative attitudes and assessments of community members. Far from precluding the possibility of conceptual objectivity, understanding the essentially social character of the discursive practice in which conceptual norms are implicit is just what makes such objectivity intelligible.

VI. FROM INTENTIONAL INTERPRETATION TO ORIGINAL INTENTIONALITY

1. *The Stance Stance*

The normative house has many mansions. The particular norms of concern in this work are *discursive* normative statuses, the sort of commitment and entitlement that the use of concepts involves. These norms, it will be claimed, are instituted by *social* practices. These are practices that incorporate the distinction of social perspective between two kinds of practical attitude one can adopt toward a commitment: *acknowledging* it (oneself) and *attributing* it (to another). Elaborating an account along these lines is pursuing three of Wittgenstein's grand themes: the insistence on the *normative* character of language and intentionality, the *pragmatist* commitment to understanding these norms in terms of practices rather than exclusively in terms of rules, and the recognition of the essentially *social* character of such norms. One way in which the significance of the social character of the attitudes that institute intentional norms can begin to be approached is by considering the relation between the practical activity of intentional interpretation and the intentional states that are attributed by such interpretations.

Dennett's original account of intentional systems and intentional explanations provides a useful place to start.⁸² One characteristic feature of that account is the idea that intentionality ought to be understood in terms of ascriptions of intentionality. Explanatory pride of place is granted to a certain sort of *attitude*, what Dennett calls a "stance." To adopt the intentional stance toward some system is to offer an intentional explanation of its behavior, by attributing intentional states to it. Adopting the intentional stance toward something is *taking* or treating it in practice *as* an intentional system. The status of being an intentional system, of exhibiting intentional states, is instituted by this attitude or stance: "A particular thing is an intentional system only in relation to the strategies of someone who is trying to explain and predict its behavior."⁸³ Dennett's explanatory strategy is first

to define what it is to adopt the intentional stance, that is to offer an intentional explanation, then to explain when it is appropriate to adopt that stance, and finally to define an intentional system as whatever is appropriately treated as one by adopting the intentional stance toward it. A few words are in order about each of these moves.

The beginning of wisdom about intentional explanation lies in appreciating the *normative* significance of attributing intentional states (mentioned in Section II of this chapter). Attributing suitably related beliefs and desires is attributing a certain sort of *reason* for action. Taking someone (1) to believe that it is raining and that the only way to stay dry is to open an umbrella and (2) to desire to stay dry is taking that individual to have a reason to open an umbrella.⁸⁴ To say this is not yet to say that the one who has such a reason *will* act according to it, even in the absence of competing reasons for incompatible courses of action. What follows immediately from the attribution of intentional states that amount to a reason for action is just that (*ceteris paribus*) the individual who has that reason *ought* to act in a certain way. This 'ought' is a *rational* ought—someone with those beliefs and those desires is rationally obliged or committed to act in a certain way. The significance of the states attributed is in the first instance a matter of the force of the better reason, rational force. That, as previously remarked, is a normative affair. Intentional interpretations attribute normative statuses, whose significance concerns practical proprieties. This is not to deny that reasons can be causes. It is just to unpack slightly what is meant by saying that they are reasons. The relation of such normative attributions of status and propriety to attributions of natural states and properties is a further issue.

Dennett acknowledges the normative core of intentional attribution and the corresponding distinction between physical and intentional explanation: "Deciding on the basis of available empirical evidence that something is a piece of copper or a lichen permits one to make predictions based on the empirical theories dealing with copper and lichens, but deciding on the basis of available evidence that something is (may be treated as) an intentional system permits predictions having a normative or a logical basis rather than an empirical one."⁸⁵ Attributing a natural state or property such as being copper supports *descriptive* conclusions about how the subject of those attributions *will* (in fact) behave. Attributing a normative status or propriety such as having beliefs and desires that amount to a reason for opening one's umbrella supports *prescriptive* conclusions about how the subject of those attributions *ought* (rationally) to behave. Within Dennett's project, however, the ultimate interest of intentional explanation lies in its use in deriving predictions concerning actual behavior. Some additional premise is required to get from the prescriptive conclusions that intentional attributions immediately supply to the descriptive predictions Dennett is concerned with.

He supplies the additional premise, in the form of a *substantive rationality assumption*, to the effect that agents generally do what one ought (ration-

ally) to do, what one is committed by one's intentional states to do. To be rational in Dennett's sense is to act as one rationally ought, to act as one's intentional states commit or oblige one to act. In order to derive predictions of actual behavior from attributions of intentional states, it is necessary to add the assumption that the subject to which those states are attributed is rational in this sense. In other words, intentional interpretation supplies a *primary* intentional explanation of the *normative* status of the one interpreted, an account of what performances are *appropriate* in the light of the beliefs and preferences attributed. Supplemented by a substantive rationality assumption, these normative characterizations can be used to ground predictions about actual performances, yielding a *secondary* intentional explanation of behavior described in nonnormative terms. The substantive rationality assumption provides the bridge that connects the normative significance of intentional attribution with the actual dispositions of the subject of such attribution.

Dennett's most controversial claim is his stance stance—his claim that there is no room for a distinction between actually being an intentional system and being appropriately treated as one. Intentional systems, things that have intentional states, just are whatever things it is predictively useful to adopt the intentional stance toward. The point of the stance idiom is that the notion of someone viewing or treating something as an intentional system is to be prior, in the order of explanation, to that of being an intentional system. The only notion of intentional system Dennett permits himself is "what one is treating something as when one offers intentional explanations of its behavior." Intentionally interpreting, adopting an intentional interpretive stance, is a practical attitude, and proprieties governing that practical attitude institute intentional states and hence normative statuses. Intentional states and intentional systems are, if not in the eye of the beholder, in the successful explanatory strategies of the theorist.

In the same way, Dennett distinguishes the significance of *ascribing* intentional states to a system from that of *describing* the system. Intentional ascriptions are appropriate according to their predictive utility, not their descriptive accuracy. The appeal to stances or attitudes as prior in the order of explanation to intentional states or normative statuses need not be read this way, however. For the contrast between talking about something in intentional vocabulary and talking about it in physical vocabulary is not for Dennett a distinction between adopting a stance and doing something else. The physical stance is also a stance. What appears in the instrumentalist reading as a distinction between what is really out there and what it is convenient from the point of view of prediction to attribute (the naturalized version of the proprieties of takings as intentional attributions), between describing and ascribing, between representing and adopting a stance or attitude, is, for Dennett, not itself a factual matter, a matter of what is out there, a matter of whether our representings do or do not correspond as they ought

to what is represented. It is rather a distinction between two stances that one may adopt. Taking there to be a physical fact of the matter determining the proprieties of our takings is adopting one stance; taking it that there are only predictive conveniences determining those proprieties is another. It is stances all the way down.

For Kant the difference between the realm of Nature and the realm of Freedom, and hence in the ordinary sense the distinction between facts and norms, is itself not a factual but a normative difference (the difference between acting according to rules and acting according to conceptions of rules). So one might say that for Dennett the difference between physical systems and intentional systems is itself a normative difference, a matter of the *propriety* of adopting different explanatory-predictive stances to the system in question. After all, for something to be a sample of copper is just for it to be proper or correct to treat it as one, in one sense of 'proper or correct' (the objective representational sense discussed at the end of the previous section), just as for something to be an intentional system is for it to be proper or correct to treat it as one, in another sense of 'proper or correct'. The question is how to understand the relation between the kinds of norms that govern the adoption of these different sorts of stance or attitude. It follows that Dennett's strategy of treating the normative significances of intentional states as instituted by the attitudes of interpreters does not by itself involve a commitment to reducing the normative to the nonnormative, insofar as it is *proprieties* of attitudes that are invoked. That reductive commitment comes in later, in explaining those proprieties. Understanding those proprieties in terms of predictive success, as Dennett does (a strategy different from that to be pursued here) gives an objective basis to the norms governing the adoption of the intentional stance. It puts Dennett in a position to say that talk of the predictive utility of adopting that stance is just a way—indeed, the only one available to us—of specifying an important kind of objective pattern of behavior. Thus the normative status of being an intentional system does not collapse into the adoption of normative attitudes of intentional interpretation.

2. *Different Stances and Kinds of Intentionality*

Understanding being an intentional system in terms of being appropriately taken or treated as an intentional system by being intentionally interpreted is not as such a reductive strategy for understanding intentionality in nonintentional terms. Offering intentional explanations of the behavior of others is something that only intentional systems can do. What is the relation between the intentionality that an intentional interpreter and attributor attributes, and that which the interpreter exhibits or possesses? To attribute beliefs (and desires and intentions), to adopt the intentional stance, one must have the concept of belief (desire, intention) and the capacity to

acquire beliefs concerning the appropriateness of applying that concept in understanding the behavior of candidate intentional systems. According to Dennett, intentional systems that can take up the intentional stance toward other systems have a special kind of intentionality. Intentional interpreters belong to "the subclass of intentional systems that have language, that can communicate": "Just as not all intentional systems currently known to us can fly or swim, so not all intentional systems can talk, but those which can do this raise special problems and opportunities when we come to ascribe beliefs and desires to them. That is a massive understatement, since without the talking intentional systems, of course, there would be no ascribing beliefs, no theorizing, no assuming rationality, no predicting."⁸⁶ Clearly, then, it is not possible to understand the second-class sort of intentionality attributed by creatures who offer intentional explanations of others, without understanding the first-class sort of intentionality those attributors themselves display. Dennett's assumption that possessing intentional concepts and attributing intentional states such as belief—that theorizing, predicting, assuming, and explaining—all presuppose specifically linguistic capacities is not universally shared, although good reasons for it will emerge. For present purposes, what matters is the distinction between first- and second-class intentionality—the kind possessed by attributors of intentionality, and the kind possessed by those to whom intentionality is attributed, rather than the specific characterization of the former. For it now appears that the intentionality of relatively simple systems such as animals and chess-playing computers, toward which Dennett takes it to be appropriate to adopt the intentional stance, can be understood only against the background of an understanding of the more complex systems capable of adopting that explanatory stance.

Thus it must be asked whether the fact that something is an intentional system in the first-class sense of attributing intentionality is a fact of the same general sort as the fact that something is an intentional system in the second-class sense of attributed intentionality. Dennett says of the second, attributed variety of intentionality, that the only facts in the vicinity are facts about the propriety of adopting a certain kind of stance toward it. Is the same thing true of the ascriber? Are the only facts about whether what one is doing is ascribing or attributing intentionality facts about the practical propriety of adopting a certain stance toward the interpreter, treating it in a certain way? Is adopting the intentional stance something one really does, or is the taking of a stance merely something that is sometimes appropriately attributed, so that it can be correct to adopt the stance that someone is adopting the intentional stance? Is it in this sense stances all the way down?

Distinguishing *simple intentional systems*, which are merely intentionally interpretable, from *interpreting intentional systems*, systems toward which the intentional stance can be adopted from systems that can adopt that stance toward others, is distinguishing instituting intentionality from instituted intentionality. Simple intentionality, which on this line is in the

eye of the beholder, is for that reason *dependent* on and in an important sense *derivative* from the intentionality exhibited by interpreters. The clearest examples of the derivative character of some intentionality or conceptual content are those in which interpreters explicitly assign some meaning to an intrinsically meaningless event by deciding to understand it in a certain way. They *make* an event mean something ("One if by land, and two if by sea . . .") by *taking* it to mean that, by understanding it that way. The meaning is conferred on the occurrence by the response to it that becomes appropriate, by the conclusions that are drawn from it (" . . . And I on the opposite shore will be"). The intentional content of the signal derives from the intentional content of the beliefs it makes appropriate for its audience. Noises and marks on paper do not mean anything all by themselves. Meaning is correlative with understanding, and they understand nothing. It is the possibility of *our* understanding them as expressing a content involving the application of concepts that makes *them* mean anything. Our understanding, our practices of interpretation institute that meaning, which derives from them.

The intentionality, the conceptual content, of noises and marks is borrowed from and dependent on that of the thoughts and beliefs that interpret them, the takings, or practical attitudes that attribute such content. On pain of an infinite regress, it seems necessary to distinguish the *derivative intentionality* such merely interpretable items display from the *original intentionality* their interpreters display. Clearly the simple intentionality of systems that can be interpreted as having and acting according to beliefs and desires is not derivative in the same sense in which that of inanimate marks and noises is. Nonetheless, on a view such as Dennett's the intentional contentfulness of the states of such systems depends on their interpretability by other, more capable systems. In the case both of inanimate and animate interpretable, the attempt to understand the sort of intentionality they display drives one back to the practical attitude or activity of interpreting. They exhibit no intentionality intelligible in its own right, apart from the practical attitudes of the interpreting systems.

Here, then, is a challenge: to maintain the stance toward both simple and interpreting intentional systems—that is, to acknowledge that the normative status of being such intentional systems is intelligible only by reference to the normative attitude of taking or treating something as such a system, that is interpreting it as one—while at the same time securing the distinction between original and derivative intentionality—and so not allowing the notion of intentional normative status to collapse into that of the normative attitude of intentional interpretation. This ought to seem hard to do. Indeed, Searle claims in effect that it is impossible⁸⁷—that if derivative intentionality is to be intelligible, so must a sort of 'intrinsic' intentionality possessed by intentional interpreters, which can be made sense of quite apart from any reference to anyone taking or treating the states of those interpreters as intentionally contentful. From the point of view of the present project,

the regress argument he employs to derive that conclusion is flawed by its dependence on an inappropriate model of what it is to take or treat something as intentionally contentful. For following his treatment in *Speech Acts*,⁸⁸ he understands taking or treating a mark or noise as expressing a certain claim or proposition as depending on propositionally explicit beliefs and intentions regarding it—as interpreting it in Wittgenstein’s sense. A version of the regress-of-rules argument then shows that those beliefs and intentions cannot have their meaning conferred on them in the same way. But this leaves out the possibility of conferral of such content by *implicit* practical taking or treating of states, performances, and expressions as intentionally contentful. This is the possibility pursued in Chapter 3.

The theory developed in this work can be thought of as an account of the stance of attributing original intentionality. It offers an answer to the question, What features must one’s interpretation of a community exhibit in order properly to be said to be an interpretation of them as engaging in practices sufficient to confer genuinely propositional content on the performances, statuses, attitudes, and expressions caught up in those practices? The key to the account is that an interpretation of this sort must interpret community members as taking or treating each other in practice as adopting intentionally contentful commitments and other normative statuses. If the practices attributed to the community by the theorist have the right structure, then according to that interpretation, the community members’ practical attitudes institute normative statuses and confer intentional content on them; according to the interpretation, the intentional contentfulness of their states and performances is the product of their own activity, not that of the theorist interpreting that activity. Insofar as their intentionality is derivative—because the normative significance of their states is instituted by the attitudes adopted toward them—their intentionality derives from each other, not from outside the community. On this line, only communities, not individuals, can be interpreted as having original intentionality.

For this to work, the practices that institute the sort of normative status characteristic of intentional states must be *social* practices. Those practices essentially incorporate a distinction of social perspective between the attitudes of *undertaking* a commitment, as someone who believes that a bear is approaching might be taken to be committed to believing that an animal is approaching, and *attributing* a commitment, as the one who interprets another as having such a belief might do. The first sort of attitude toward a normative status must be attributed even to simple intentional systems—the rationality that is for Dennett the mother of intention is a way of talking about the sense of ‘ought’ in which one who believes a bear is approaching ought to believe that an animal is approaching. Just for that reason, the second sort of attitude is an implicit version of adopting the intentional stance. According to the account offered in Chapter 3 of the practices that confer distinctively propositional contents (and accordingly underlie all dis-

cursive intentionality, the conceptual contentfulness of expressions, performances, attitudes, and statuses), the practical normative attitudes of undertaking and attributing commitments come as a package—neither is intelligible apart from the other. Undertaking a commitment just is doing something that makes it appropriate for that commitment to be attributed. Normative statuses of the sort whose paradigm is provided by the inferentially articulated commitments constitutive of rationality are instituted by constellations of socially perspectival normative attitudes of attributing and undertaking such commitments. This is the *I-thou* structure of norm-instituting social practices that was contrasted above with the *I-we* sociality many theorists appeal to, and which is understood here as arising out of the more primitive perspectival variety.

3. Summary

The point of this chapter is to motivate the criteria of adequacy governing the account of discursive practice presented in Chapter 3, as well as the basic raw materials deployed there to satisfy those conditions. The first major point is the *normative* significance of intentional states, meanings, and the sort of understanding that is the uptake of those meanings. The second point is that norms that are *explicit* in the form of rules, principles, or claims (Wittgenstein's "interpretations") depend for their intelligibility—their determining a distinction between performances that are correct and incorrect, appropriate and inappropriate—on a more fundamental form of norms that are *implicit* in practice—in what is *done* rather than what is *said*. Making this distinction raises the question of how to understand the practice of making propositionally explicit claims (formulating principles, promulgating rules, and so on) in terms of norms that are implicit in practices.

The third point is that the attempt to understand norms implicit in practices by identifying the correct/incorrect distinction with the regular/irregular distinction (one strategy for reducing the normative to the nonnormative) will not work, for a reason parallel in form to the one that shows the need for a notion of norms implicit in practice in the first place. The regress-of-rules or regress-of-interpretations argument against *regularism* is that if rules were the only form of norms, they would fail to sort performances into those that are correct according to the rule and those that are not. For applying the rule is itself something that can be done correctly or incorrectly, and any performance that is correct according to one interpretation is incorrect according to others. The gerrymandering argument against *regularism* is that if norms are understood as regularities, they fail to sort performances into those that are correct (regular) and those that are not. Any course of conduct embodies many regularities, and any candidate performance that is regular according to one of them is irregular according to others. The two strategies

do not provide the resources to *privilege* one of the competing interpretations or regularities.

The fourth point, then, is that there is another move available for understanding what it is for norms to be implicit in practices. This is to look not just at what is *done*—the performances that might or might not accord with a norm (be appropriate or inappropriate)—but also at *assessments* of propriety. These are attitudes of taking or treating performances *as* correct or incorrect. If such attitudes are themselves understood on the model of propositionally explicit beliefs or commitments—as responding to a performance as correct by *saying* of it that it is correct—then the regress objection to regulism about norms reappears. But such assessing attitudes can also be understood as implicit in practice.

One way of doing that is to look to *sanctions*—treating a performance as correct by responding in practice with a reward (or the withholding of punishment) and treating it as incorrect by responding in practice with a punishment (or the withholding of a reward). What counts as a reward or punishment might be construed naturalistically, for instance as any response that positively or negatively reinforces the behavior responded to. Or it might be construed normatively, for instance in terms of the granting of special rights or the assignment of special obligations. Again, the assessing attitudes taken to be relevant to normative statuses can be taken to be implicit in the responses of other individuals, or of responses associated in some way with the whole community. In any of these cases, if the normative status of being a correct performance were identified solely by appeal to *regularities* exhibited by assessments, then the gerrymandering objection would be reinstated.

The fifth point, then, is that one way to demystify norms is to understand them as *instituted* by the practical attitudes of those who acknowledge them in their practice. Apart from such practical acknowledgment—taking or treating performances as correct or incorrect by responding to them as such in practice—performances have natural properties, but not normative proprieties; they cannot be understood as correct or incorrect without reference to their assessment or acknowledgment as such by those in whose practice the norms are implicit. At the same time, it must be acknowledged that a cardinal criterion of adequacy of any account of the *conceptual* norms implicit in discursive practice is that it make intelligible their *objectivity*. Doing so requires that the normative status of being a correct application of a concept not collapse into normative attitudes, as construing correctness as consisting just in being *taken* to be correct conflates them. The objectivity of conceptual norms requires that any attitude of taking, treating, or assessing as correct an application of a concept in forming a belief or making a claim be coherently conceivable as mistaken, because of how things are with the objects the belief or claim is about.

The next chapter opens the discussion of the propositional contents that are conferred on expressions, performances, attitudes, and statuses by their

playing a suitable role in a system of discursive normative social practices. The idea of normative statuses as *instituted* by practical attitudes, which has been put on the table in this chapter, should be distinguished from the idea of their intentional contents as *conferred* by the social practices in which those statuses and attitudes play a role. As the terms are used here, the institution of status by attitude has to do solely with pragmatics, the study of the practices in which discursive norms are implicit. The conferral of content by practice has to do with the relation between such pragmatics and semantics, which is the study of conceptual contents.⁸⁹ The raw materials for a pragmatics that have been assembled here are employed, in Chapter 3, to contribute to both explanatory projects. The next chapter accordingly begins the investigation of concept use and intentional contentfulness.

Appendix: Wittgenstein's Use of *Regel*

It should be admitted that Wittgenstein's own terminology in some ways obscures the very point he is after in the regress-of-rules argument. For he uses "rule" extremely broadly, to cover much more than is allowed in the usage endorsed here. According to this latter usage, rules are discursively articulated and propositionally contentful; they determine what is correct by describing the correct performances, *saying* what must be true of a performance for it to be correct. By contrast, Wittgenstein uses "rule" (*Regel*) in at least three importantly distinct senses. First is the sense that coincides with the usage preferred here: rules explicitly say what one is to do and are consulted as such by those who follow them—the rule followers' performance is governed by their understanding of the concepts used to characterize what they ought, according to the rule, to do. Second, he sometimes uses "rule" to describe *whatever* guides or is consulted by those whose behavior is being assessed, whether or not it is discursively or conceptually articulated. Finally, he even sometimes talks about following a rule whenever someone's behavior is subject to normative assessment, whenever responsibility to proprieties of conduct is attributed, regardless of whether there is anything the one "following the rule" is aware of or consulting, or being guided by in determining what to do.

In one central text, Wittgenstein offers two senses in which games may be said to be played according to rules: "The rule may be an aid in teaching the game . . . Or it is an instrument of the game itself—Or a rule is employed neither in the teaching nor in the game itself; nor is it set down in a list of rules. One learns the game by watching how others play. But we say that it is played according to such-and-such rules because an observer can read these rules off from the practice of the game—like a natural law governing the play."⁹⁰ The first is following a rule in the sense in which that phrase is used

here, according to which one must understand what the rule says and then try to produce performances that the concepts it employs properly apply to. The other, corresponding to the third of the senses distinguished above, is totally external, involving norms that are only in the eye of the beholder, as the remark about natural laws indicates. These are the two senses that Kant distinguishes as acting according to a *conception* of a rule, as agents do, and merely acting according to rules, as inanimate objects do.

In the context of the regress-of-rules argument, this third sense of rule-following, in which it coincides with simple regularity, must be marginal—the question of how to understand a way of grasping a rule that is not an interpretation hardly arises for inanimate objects as they act according to the laws of physics. This is the sense that seems to be involved in the discerning of rules wherever it would be correct to apply ‘same’ or ‘agreement’.⁹¹ Wittgenstein is even willing to appeal to this sense in such outré (according to the usage preferred here) cases of “rule-following” as those involving rules relating pain to pain-expressing behavior.⁹² These would seem to be cases in which the rule is entirely in the eye of the beholder, who takes there to be a regularity. Insofar as they are not, these are cases of the second sort, where the performer is being guided by something, but not by something explicit and articulate. It is in this sense that he is willing to call tables of colors and even signposts “expressions of rules.”⁹³ He seems to call it “following a rule” wherever there is some object whose features it would be appropriate to cite in justifying one’s performance, exhibiting it as appropriate or correct. Though in one place he seems to be careful not to call a map a rule,⁹⁴ in others he is even willing to say a line can function as a rule,⁹⁵ and a line is clearly not a saying of any sort.

This multiplicity of senses cries out for the distinctions to be explicitly marked terminologically, which is the intent of the relatively more regimented uses of ‘rule’ and ‘practice’ that is employed here. The important point is that there is a way of grasping a rule that is not an interpretation. This should be talked about, as Wittgenstein at least sometimes does, in terms of *practices*—grasping a rule without interpreting it is grasping it in practice, rather than by substituting one expression of a rule for another. Most cases of understanding explicit claims and obeying explicit orders should be understood in this way. Such application of a rule is something that can be done correctly or incorrectly. Practices in this sense are the primitive sort of acknowledgment that performances are governed by norms. But according to this way of regimenting the idiom, not all practices are graspings of rules. There are practices that involve the acknowledgment of norms without involving rules at all, except in the sense that others, looking on, may be able to state rules—whose expressions are not available to the practitioners.

According to this way of using the term, rules are explicit statements that specify what is correct and incorrect by *saying* (describing) it. Obeying an

order (one of the activities Wittgenstein most often links with rule-following in the *Investigations*) thus counts as a kind of rule-following. But one ought not to say that there are rules involved at all in any practice that does not involve claiming, judging, and describing, though of course there are proprieties of practice in more primitive 'games'. That 'practice' is not to be restricted to "ways of grasping rules that are not interpretations" for Wittgenstein seems to be clear from one of the passages cited above: "To obey a rule, to make a report, to give an order, to play a game of chess, are customs (uses, institutions)."⁹⁶ Here rule-following is explicitly just one example. Making a report is not following a rule, though it is governed by proprieties of practice, nor, typically, is giving an order.