CHAPTER 1

Introduction

1.1 Breakdown and Reconstruction

There was a time in my life when I wanted to learn how to fix cars. Reflecting back, it is striking that I learned the most from breakdowns. A properly working automobile teaches only the basic system. My understanding was deepened when something went wrong, when I tried to fix it and failed, when I ended up making things worse. Perhaps this lesson can be applied to our understanding of the sciences. In the philosophy of science, we are drawn to the ideal cases. We want to know what science would be like when all the evidence is collected or collectable, not when it is influenced by personal bias. We look at practicing scientists when they are happily engaged in normal science (or staging their bloodless revolutions), not when their projects are faltering and tenure is on the line. My faint attempt to learn auto mechanics suggests that our normal approach might be backwards. To understand science, it might be wise to look at cases where it breaks down.

A particularly interesting form of breakdown occurs when the understanding of human behavior reaches a dead end. This is not the puzzlement of someone entirely new to a culture. The informative breakdowns occur when one has enough understanding to permit a genuine conflict between the working interpretation and the new events. Then there is a breach of understanding and one is struck by something that seems completely crazy. A favorite[1]—albeit fictional—example comes from Wittgenstein's *Remarks on the Foundations of Mathematics:*

> 142. . . . People pile up logs and sell them, the piles are measured with a ruler, the measurements of length, breadth, and height multiplied together,

1. This is also a favorite of Richard Zaffron, to whom I owe much more than this reference.

Copyrighted Material

> and what comes out is the number of pence which have to be asked and given. They do not know 'why' it happens like this; they simply do it like this: that is how it is done.—Do these people not calculate? . . .
>
> 148. Very well; but what if they piled the timber in heaps of arbitrary, varying height and then sold it at a price proportionate to the area covered by the piles?
>
> And what if they even justified this with the words: "Of course, if you buy more you have to pay more"?
>
> 149. How could I shew them that—as I should say—you don't really buy more wood if you buy a pile covering a bigger area?—I should, for instance, take a pile which was small by their ideas and, by laying the logs around, change it into a 'big' one. This *might* convince them—but perhaps they would say: "Yes, now it's a *lot* of wood and costs more"—and that would be the end of the matter.—We should presumably say in this case: they simply do not mean the same by "a lot of wood" and "a little wood" as we do; and they have a quite different system of payment from us. (Wittgenstein 1956: 43–44)

Wittgenstein's example has all the elements of what we will call a problem of apparent irrationality. An outsider is struck by unintelligible behavior. It is not merely foreign and strange. The actions are different from what rational and intelligent people *ought* to be doing. The divergence between what *is* happening and what *ought* to be is generated by some prior understanding. To see Wittgenstein's woodcutters as problematic, one must suppose that these people are buying and selling wood. One must know the language well enough to provide translational glosses like "big," "little," "more," and "less." Failure to make sense of these people is not just the failure to parse a sentence or follow a performance. The behavior seems irrational. Given the interpreter's current understanding of these people, they would have to be irrational to do what they are doing. The interpreter's smooth progress has broken down in a fundamental way.

Problems of apparent irrationality, then, arise when interpretation falters. Local action or speech seems irrational in the light of a background understanding. Of course, we know exactly what anthropologists, sociologists, and historians do when they come upon apparently irrational behavior. They rub their hands with glee and set about writing a monograph. When understood from the proper perspective, they explain, the bizarre activity makes sense. What was a breakdown in interpretation has been reconstructed into a deeper understanding. Wittgenstein himself suggests as much:

> We should presumably say in this case: they simply do not mean the same by "a lot of wood" and "a little wood" as we do; and they have a quite different system of payment from us.

Since the breakdown runs deep, the changes may be radical. If we rethink the meaning of "a lot of wood" and "a little wood," we are likely to shift the

Copyrighted Material

interpretation of many other words too. Their "quite different system of payment" may have implications for their social structure, religious life, and who knows what else. We hope that some such interpretive change will make sense of the events. If we find it, the behavior will (presumably) no longer seem irrational; the irrationality was merely apparent.

Breakdowns in understanding are interesting to philosophers for several reasons. In some cases, the behavior is unintelligible because it seems irrational. This kind of breakdown thus puts the traditional questions about rationality and interpretation into a new light. If we judge Wittgenstein's woodcutters to not only *seem* irrational, but to *be* irrational, then our questions about them become psychological and biological. On the other hand, if the interpreter tries to explain away the appearance of irrationality, there seems to be little to constrain her account. By hypothesis, nothing in her current interpretation makes sense of the behavior. Yet, some ways of reconstructing the interpretation are better than others. The problems of apparent irrationality thus exhibit the role that rationality plays in interpretive change. Interpretive change itself is an issue that has seen little direct discussion. What are the criteria by which we should judge one interpretive change to be better than another? While this question is a concrete version of a general epistemological problem, it is fundamentally dynamic.[2] The question is not how an existing interpretation might be (or fail to be) justified. Rather, the problem is to specify criteria for epistemically better or worse changes in an interpretation. There is no reason to be cagey about the answer. This essay will argue that the better interpretation is the one with more explanatory coherence.[3] This answer seems straightforward enough and it has plenty of intuitive appeal. One might even think it obvious. Yet articulating "explanatory coherence" in the context of the problem of apparent irrationality is far from trivial. Moreover, the study of interpretive breakdown and reconstruction should shed light on the structure of an interpretation. How do the various parts of an interpretation hang together? When do different interpretations or explanations conflict? As will become clear soon enough, the conceptions of explanation and coherence ground the answers to these questions.

Finally, two remarks about method and rhetorical style are appropriate. First, this work is not critically structured. It does not exhaustively describe the

2. This way of approaching the question was inspired by Gilbert Harman's approach to epistemological issues in *Change in View* (Harman 1986).

3. Explanatory coherence is no epistemological novelty. Coherence has been gaining popularity in epistemology (BonJour 1985; Lehrer 1990; Quine 1969b; Thagard 1978; Ziff 1984). Several philosophers have argued that explanation is central to justification (Harman 1973; Harman 1986; Lipton 1991b; Lycan 1988). Lycan and Harman, in particular, have pioneered the idea that coherence is to be understood in terms of explanation.

Copyrighted Material

range of alternatives in the field, demolish them, and build anew from their remains. While chapter 3 comes close to this philosophical style, the general approach is constructive. Chapter 2 presents a problem—the problem of apparent irrationality—and sections 2.5 and 4.1 make the requirements for its solution clear. The body of this work presents a systematic resolution of the problem of apparent irrationality. Other philosophical views are considered in detail only when they contain arguments that directly conflict with some part of the proposed solution. Some readers may find this approach unsatisfying, since there is no attempt to argue that this solution is the only or best one. In particular, there is no comparative discussion of Foucaultian or Habermasian approaches to rationality, nor of popular alternative conceptions of explanation like the unification approach or the statistical-relevance model. Comparative discussion is essential for the ultimate evaluation of this essay. But to engage in the systematic demolition of alternatives before constructing my own would be unconvincing. Only now that the work is complete can one begin to evaluate its strengths and weaknesses as compared with other views.

Second, in this age, no work can be unselfconscious about the gender and power implications of its own language. This project faces a double problem. There is the usual challenge of finding readable, gender-neutral constructions for the personal pronouns. In addition, there is a power asymmetry between the interpreter and the subjects of investigation. The latter have often been colonized, objectified, and exploited. Social scientists have become terribly self-conscious about the language they use to describe the humans they study. "Natives" brings forward connotations of naked, stone-age folks. Even a word as apparently innocent as "subject" has become unacceptable in certain circles. Unfortunately, the proposed alternatives like "research partner" only mask the real asymmetry of power. In the philosophy of social science, it is better, I suggest, to leave the power relation explicit in the language, lest we forget it. This essay therefore deploys a number of terms to describe the persons studied: "native," "native speaker," "subject," "local," "member of the local group," and so on. To fight the negative connotations, bear in mind that *every* human is the native of some cultural milieu. In the paradigm cases, the interpreter is an outsider trying to understand persons native to an unfamiliar locale. This contrast between interpreter and interpretee is important, and thus remains in the language of this essay. Playfully disrupting the asymmetry of power, however, I have chosen to use the pronoun "she" for the interpreter and "he" for the subject of interpretation. This resolves the problem of gender-neutral language by sprinkling both pronouns evenly throughout the text. It has the additional advantage of marking the pronouns for the reader. Feminine pronouns used anaphorically will always refer to the interpreter, and masculine pronouns will refer to the persons interpreted. I fear that the trick is too clever by half, and that some readers will be annoyed. If so, my only defense can be a plea for your tolerance of a well-intentioned peccadillo.

Copyrighted Material

1.2 Synopsis

Chapter 2: Apparent Irrationality

To illuminate the problem with which this work is concerned, we will begin by examining four cases of apparent irrationality drawn from the social scientific literature. In each case, the interpreter was presented with prima facie irrational speech or behavior, and in each case the interpreter went on to devise a more adequate interpretation. The aim of this work is to uncover the methodological principles that govern this sort of interpretive change. Each case exposes one or more facets of the problem, and the final section, section 2.5, analyzes the problem of apparent irrationality and poses four questions that are central to its resolution:

1. What are the criteria for preferring one interpretation to another?
2. Is it possible for an interpreter to prefer an interpretation that attributes to the interpretees standards of rationality different from her own?
3. What are the epistemic grounds for claiming that a person or group adheres to a norm, rule, or standard?
4. Under what conditions are interpretations that include descriptions and evaluations from the subjects' point of view to be preferred over interpretations that employ ideas or evidence to which the subjects do not have access?

Chapter 3: Interpretive Change

Given the examples from which the problem of apparent irrationality arises, the first and second questions of section 2.5 are closely related. The four examples of chapter 2 show that whether an action appears irrational depends heavily on the interpretation from which it is understood. The content of the interpretation is at least part of the basis for evaluating the rationality of the subjects. Any methodology of interpretive change, then, must be able to adjudicate among interpretations that entail different assessments of the rationality of the agents. This means that the discussion of interpretive change involves the question of whether and to what extent social scientists need to presuppose that their subjects are rational. Most of the literature takes Peter Winch's "Understanding a Primitive Society" (1964) as the proximate origin of the position that an interpreter must identify and deploy local criteria of rationality. In response, others argued that an interpreter must presuppose her interlocutors to be rational, and thus at least some criteria of rationality are not discovered (Davidson 1984; Hollis 1967a; Hollis 1967b; Lukes 1967; Lukes 1982; Macdonald and Pettit 1981; Root 1986). Any criterion of interpretive

Copyrighted Material

choice that can address the problems of apparent irrationality will thus stake out a position on this issue and thereby answer question 2.

Section 3.1 addresses the principles advanced under the name "the principle of charity." This section identifies and rejects three different versions of the idea: a principle of logical charity, a principle of semantic charity, and a principle of confirmation charity. The first and third require agreement between the interpreter and the locals. The principle of semantic charity is the most satisfactory of the three, but it does little to guide interpretive choice. Explanatory grounds for choice of interpretation are usually discussed under the guise of the "principle of humanity." These principles require that the best interpretation minimize unexplained disagreement. Most defenders of this idea maintain that explanation is asymmetric. False or irrational belief is explained, while true and rational belief are interpreted. Section 3.2 criticizes the arguments that true and rational belief need to be treated differently from false or irrational belief. Section 3.3 argues that a symmetrical principle is the best candidate for a principle of interpretive change. The principle of explanatory coherence holds that the best interpretation is the one with the most explanatory coherence. According to this approach, criteria of rationality are explanatory posits. The interpreter hypothesizes that the locals adhere to one standard or another, and this hypothesis is successful insofar as it can be embedded within the interpretation that has the most coherence. The explanatory coherence position on the rationality issue, then, is neo-Winchean insofar as it gives an affirmative answer to question 2.

This book will speak at length about "interpretation." While it would beg our central questions to begin with a strict definition, something must be said about this notion. The paradigm cases of interpretation for this work are the monographs of ethnographers and historians. This kind of inquiry is distinctive because it depends on claims about word meaning and intentional action. As a first pass, interpretation is different from other sorts of theorizing because it aims at understanding meaning. Many kinds of inquiry outside of ethnography and history rely on claims about meaning, and this work will speak broadly about the "interpretive disciplines" with the intention of capturing a broad cross-section of social scientific practice. There are, however, two associations of the word "interpretation" that need to be disengaged from its use here. Artistic works are interpreted too. Whether aesthetic interpretation can be encompassed by the methodology presented here is an open question. From the perspective of this work, nothing is lost if the two sorts of interpretation turn out to be very different. Also, self-interpretation is arguably an activity in which all humans engage. At the outset, it will be an open question whether the interpretive activities of an ethnographer or historian are the same or different from the interpretive activities of her subjects. As the essay progresses, we will see that these two kinds of interpretation do not need to coincide, and that in fact they will often diverge. This book will argue for an epistemology

Copyrighted Material

of interpretation whereby an interpreter can represent the interpretive activities of others without necessarily engaging in the same kind of activity.

Chapter 4: Explanation

Many philosophers have argued that some aspects of human life cannot be reached by explanation. Disputants over the relative roles of explanation and understanding typically approach the problem with a conception of explanation already formed. This seems backwards. The question should be What concept of explanation is adequate to the demands of the problem of apparent irrationality and the question of interpretive change? Section 4.1 uses the explanatory coherence principle of interpretive change and the problems of apparent irrationality to establish criteria of adequacy for any conception of explanation. In particular, any adequate conception of explanation will have to give an explanatory role to claims about meaning, intentions, and norms. Chapters 5, 6, and 7 try to show how the model of explanation presented in chapter 4 satisfies the criteria of adequacy articulated in section 4.1.

Section 4.2 articulates the most promising candidate for satisfying the criteria of section 4.1. The erotetic, or "why-question," model of explanation takes an explanation to be an answer to a why-question (Garfinkel 1981; Lipton 1991b; van Fraassen 1980). This section presents the formal features of why-questions and their answers. Sections 4.2, 4.3, and 4.4 develop the model and defend it against its critics, making adjustments where necessary. The erotetic model is notorious for giving a prominent role to interest in choice of explanation. While most writers on this topic have tried to find ways of limiting the role of interest, section 4.4 argues for *laissez-faire* contextualism. There are no restrictions on the topics, foils, or relevance criteria beyond those made necessary by the presuppositions of the question and the interests of the investigator. The final section of the chapter (section 4.5) explores the way in which the erotetic model of explanation and the principle of explanatory coherence combine to provide a detailed account of interpretive change. At the end of the book, the account provided in section 4.5 is revisited in the light of the intervening chapters.

Chapter 5: Intentional Action and Social Explanation

The primary problem addressed in this chapter is whether intentional action explanations and social explanations of the same action necessarily conflict. Humans think of themselves as agents. Some explanations, however, treat actions as instances of social processes or as having a social function. The examples of chapter 2 show that such group-level explanations are sometimes useful when dissolving apparent irrationalities. An adequate account of interpretive change must provide a plausible way of resolving the apparent conflict

Copyrighted Material

between intentional action explanation and group-level explanations of the same action. This issue of how individual and group-level explanations are to be reconciled has been one of the pillars of discussion in the philosophy of social science. One position on this question has been that all so-called group-level explanations, insofar as they are legitimate, are ultimately reducible to intentional action explanations. The alternative view maintains that group-level explanations are sui generis. This chapter will argue for the latter position with respect to several specific forms of group-level explanation. Indeed, appeal to group-level explanations turns out to be essential to the explanatory coherence account as developed here. This leaves us with an acute form of the "invisible hand problem." That is, individuals are treated as if their intentions or motivations were irrelevant to the action at hand (Macdonald and Pettit 1981; Rosenberg 1988). Chapter 5 approaches the problem by analyzing the form of intentional action and social functional explanations, and by providing a general account of the conditions under which two kinds of explanation (or two explanations of a single event) conflict. Section 5.2 analyzes intentional action explanations and explanations of belief. Since functional explanations are the most likely to conflict with intentional action explanations, section 5.3 provides a detailed analysis of them. The final section (section 5.4) argues that there is no necessary conflict between functional explanations and reason-giving explanations, even when these concern the same event.

This chapter has two important consequences. First, the chapter shows how the erotetic model can be applied to intentional action, thus satisfying one of the criteria of section 4.1. Second, it gives part of the answer to question (4) of chapter 2: Under what conditions are interpretations that include descriptions and evaluations from the subjects' point of view to be preferred over interpretations that employ ideas or evidence to which the subjects do not have access? An interpreter may discover ways of explaining actions or social structures that draw on conceptions to which the locals have no access. The local and the interpreter's characterizations might be in prima facie conflict. Under what conditions should we prefer the interpreter's explanations or descriptions to the locals'? An adequate criterion of interpretive change must provide some plausible way of resolving the conflict between "internal" and "external" (or "emic" and "etic") characterizations. Both the political and epistemic aspects of this issue have received attention in philosophy (Bohman 1991; Feleppa 1988; Giddens 1976; Habermas 1984; Schutz 1967; Winch 1958). It is also important, if somewhat passé, in anthropology (Headland, Pike, and Harris 1990). According to the arguments of this chapter, the principle of explanatory coherence entails that many social phenomena will be best understood by interpretations that include both emic and etic perspectives. Intentional action explanations, as analyzed by section 5.2, must present the action to be explained, the alternatives, and the agent's reasons in terms familiar to the agent. This requirement anchors emic characterizations

Copyrighted Material

of action. At the same time, explanatory coherence may be increased by historical, functional, or structural explanations. The account of how intentional action and functional explanation can coexist in a single, coherent interpretation, then, gives us part of the account of how internal and external perspectives on human action can be reconciled. Chapters 6 and 7 also address this issue.

Chapter 6: Meaning

The importance of meaning for social life is a recurring theme among those who have taken the interpretive side of the explanation/understanding debate. The problem of meaning is whether an epistemology that sees no fundamental distinction between the interpretive and noninterpretive disciplines can capture meaning (including the self-understanding of the agents). The interpreter confronts subjects who are already engaged in the process of interpreting their own speech and action. Any methodology that grounds interpretation in simple observation of utterance and behavior can never reach the social reality at which it aims. Therefore, many have argued that the study of meaning requires a unique methodology. Section 6.1 argues that, as it has traditionally been articulated, the problem arises from an empiricist notion of observation and a foundationalist conception of justification. An explanatory coherence view of interpretive change countenances neither of these ideas. Such a view does require, however, that claims about linguistic meaning must be either explained or explanatory. Section 6.2 sketches the topics, foils, and relevance criteria of explanations that involve meaning, and thereby shows how the erotetic model of explanation satisfies yet another criterion from section 4.1. The chapter ends with a discussion of the so-called "double hermeneutic" of interpretation. Rejecting a strict theory/observation distinction and a foundationalist conception of justification shows that both natural and social inquiry are involved in something like a hermeneutic circle. Social inquiry, however, is "interpretive in a double sense" (Makkreel 1985: 239). Section 6.3 articulates the broad agreement between the explanatory coherence view and proponents of a hermeneutic methodology for interpretation. In the light of this consilience, this section criticizes the arguments that there is an additional level of interpretation unreachable by an explanatory methodology.

Chapter 7: Normativity

Rationality is normative, and it is only one particularly important aspect of the norms, rules, and values that saturate human life. The examples of chapter 2 show that hypotheses about local norms, rules, and values are crucial to any resolution of problems of apparent irrationality. An adequate account of interpretive change must therefore provide grounds for choosing between interpretations that attribute

Copyrighted Material

different norms, values, or rules. The examples of chapter 2 also show that there are some important wrinkles to this issue. Norms and rules are not always explicit. Where they are explicit, there may be local conflict over their authority or justification. Many have thought that, like meaning, norms cannot figure in explanations. The challenge in this case is that when norms become the subject of explanation or when they are used as explanantia, they collapse into mere descriptions. According to these arguments, the lacuna remaining in our account can never be filled by explanations.

The initial problem of chapter 7, then, is to show why claims about norms do not collapse into claims about the beliefs or dispositions of individual agents. The first step is to recognize the parallel between the problem of the loss of normativity and the long-standing issue about reductionism in the social sciences. Section 7.1 uses features of the erotetic model to argue that, in general, such reductions fail because the individual-level and the group-level explanations answer different questions. Along the way, this section characterizes the relevance relations and contrast classes of explanations that employ norms. Claims about norms answer questions about the group, not about individual action or belief. Section 7.2 unpacks the relationship between an agent's intentional action and the rules, norms, standards, or criteria that apply to it. This permits us to address a problem left over from section 2.5: articulate the difference between a common mistake and an implicit norm. The final section pulls together the remaining loose ends of the project. The answer to question (4) of chapter 2 is supplemented with an account of whether, and to what extent, the local agents need to be able to articulate the norms under which they act. This account opens a space to show how an interpretation might portray a group as systematically mistaken about their own actions or the norms under which they act. Also, it allows us to account for the fact that interpretations are often contested at the local level. Norms, rules, and values can be embedded into an interpretation without effacing local disputes about them.

A caveat concerning the notions of "norms," "rules," "standards," and "criteria" is in order here. This chapter will not ground its arguments in any specific conception of rules and rule-following. In particular, the use of these terms should not be taken as indicating a commitment to realism about norms, rules, or standards. Brandom (1994) has articulated a very powerful conception of the social character of norms. This chapter is consistent with his view, but also to the realistic alternatives that are antithetical to it. Similarly, this chapter does not presuppose that all rules, standards, or norms can be made explicit as beliefs of the person who conforms to them. Perhaps, as Dreyfus argues (1980), some aspect of rule-following must remain implicit as the agent's practical know-how. This book will argue, against Dreyfus, that rules can be represented by the interpreter (section 6.3), but this will not require that the person interpreted represent them in the same way as the interpreter does. The focus

Copyrighted Material

of this book is epistemological, not metaphysical. The character and ontological status of rules and rule-following are interesting philosophical issues, but they are beyond the scope of this work.

Chapter 8: Conclusion

As has already been remarked, the problem of apparent irrationality is fundamentally dynamic. In response to its demands, this essay constructs a dynamics of interpretation—an account of the forces and principles that determine how interpretations ought to change. The first section of the conclusion reiterates the essay's main points about interpretive change. The criteria for choice of interpretation articulated in section 4.5 are deepened in the light of the subsequent development of the model. The result is an understanding of interpretive change with several distinctive features: There is no unalterable core or "bridgehead" to interpretation. Claims about rationality are explanatory. Multiple explanatory perspectives are possible, even necessary, within a single interpretation. There is no knowledge of language without knowledge of culture. An interpretation does not need to homogenize local opinion. On the contrary, the most coherent interpretation will capture the real differences among the subjects. And finally, a coherent interpretation can represent some *genuine* irrationality, so there is no threat of artificial over-rationalization.

The resolution of the problem of apparent irrationality presented in this essay has ramifications for the long-standing debate over interpretation and explanation. Philosophers of social science have divided themselves into (roughly) two camps. On one side are methodological assimilationists who believe that there is a single methodology for all inquiry. On the other side are methodological separatists, who think that there are deep epistemological differences among the sciences. There are at least four issues that divide philosophers in this domain: the importance of causes or laws in explanation, the importance of understanding meaning, the relevance of normativity, and the relevance of the agents' self-interpretations. As it turns out, the explanatory coherence view does not fit easily into either camp. On each of the four issues, the explanatory coherence view articulated here either denies a presupposition of the debate or admits the significance of the phenomenon without agreeing to its purported ramifications. The position motivated in this book is globally assimilationist and locally separatist. All inquiry is explanatory, and all explanations are answers to why-questions. Therefore, there is no deep methodological divide between the social and natural sciences. At the same time, the very flexibility of the erotetic model of explanation means that there will be substantive differences among particular disciplines. The particular structure of an inquiry—the character of its explanations and their interrelations—is determined by the interests of the investigators and the facts about the domains they are investigating.

Copyrighted Material