Rationality as Methodology, Aim, and Explanation in Philosophy and Psychology

by

Carole J. Lee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in The University of Michigan
2006

Doctoral Committee:

Professor Elizabeth S. Anderson, Co-Chair
Professor James M. Joyce, Co-Chair
Professor Peter A. Railton
Professor Norbert W. Schwarz

To Conrad Ziesler

Acknowledgements

This thesis would not have been possible without thoughtful feedback from my committee members – Elizabeth Anderson, James Joyce, Peter Railton, and Norbert Schwarz – whose comments guided and sharpened my ideas and arguments. In addition, I am indebted to Frank Yates, Baruch Fischhoff, Jennifer Amsterlaw, and especially Norbert for help with the psychological literature.

The participants at the Seventh Annual Philosophy of Social Science Roundtable provided feedback on an earlier version of chapter 3; and, with their input, this chapter was improved and accepted for publication in a 2006 volume of *Philosophy of the Social Sciences*.

I have been blessed with mentors whose professional support and encouragement fostered my confidence in writing this dissertation and continuing on in academia. For this I am indebted to Elizabeth Anderson, Paul Roth, Alison Wylie, and Phillip Akutsu.

For financial support, I am grateful to Rackham Graduate School, the Inter-University Consortium for Political and Social Research (ICPSR), Marshall Weinberg, and The Sweetland Writing Institute.

Last but not least, a big cheer to all the people on the ground – my family and friends – for their loving patience and support. Special thanks go to: Conrad Ziesler for the domestic engineering that afforded me time and space for day to day thesis writing; Julie Lawrence and Brad Darling for graciously hosting my visits to Ann Arbor; Christie Hartley for indispensable advice on matters professional and practical; and, the Lee and Ziesler clans for endless enthusiasm and supplemental financial support.

Preface

My dissertation is an exploration of the relationship between psychological research and philosophical accounts of interpretation, rationality, and justification. Traditional philosophical views tie together interpretation and norms of rationality. For example, under the accounts of interpretation provided by Donald Davidson and Daniel Dennett, the norms of rationality inherent in the interpretive perspective guarantee that we can only discover that humans have generally or mostly rational beliefs. For them, interpretation is impossible unless others' beliefs can be construed as being generally or mostly rational.

If we put these philosophical claims into the context of psychological research on human judgment, we get the claim that psychological theories cannot provide intentional explanations unless the bulk of the beliefs attributed to subjects are rational in relation to each other. The relationship between psychology and philosophy here is one in which our philosophical account of interpretation precedes and limits what psychologists are capable of discovering and empirically testing with respect to human rationality and irrationality.

My dissertation aims to provide a more dynamic picture of the relationship between philosophical views on interpretation and rationality and psychological findings and methodology. I begin my dissertation by examining studies from the heuristics and biases research program, which uses experimental tasks to demonstrate that human judgment violates a priori rules of probability and rational choice theory. Studies by this prominent research program suggest that philosophers' claims about the special role that rationality should or must play in interpretation do not successfully cope when faced with the interpretation of irrational belief.

Although I take seriously the empirical findings by psychologists, I do not wish to suggest that we jettison all philosophical perspectives on interpretation in psychological experiments. I think an important, lasting contribution of Davidson's account of interpretation is his observation that any psychological theory on human judgment "must *include* a theory of interpretation" on subjects' beliefs about the experimental task. In my third chapter, I focus on psychologists' research on conversational pragmatics, which takes very seriously the task of interpreting subjects' beliefs about experimental tasks. These psychologists have looked to Paul Grice's account of cooperative communication to guide their interpretations of subject task construal. I argue that this Gricean turn in psychological research brings with it a more situational, reflexive perspective on interpretation. And, in doing so, it imposes important evidential standards in the interpretation of experimental results. These evidential standards require empirical information about the conditions of successful versus unsuccessful communication for specific experimental contexts.

Like philosophers, psychologists have taken a special interest in the question of whether human judgment is rational or irrational. There has been a recent trend to discover the conditions that promote rational rather than irrational judgment. And, there has also been a trend, especially among evolutionary psychologists, to impute cognitive mechanisms that can explain rational judgment and context effects. In chapters 2 and 4, I argue that these trends have interesting connections to philosophically minded projects.

Cognitive psychology's disciplinary trend towards studying rational rather than irrational judgment was in some ways a reaction to the heuristics and biases research program. Psychologists actively sought to limit the scope of Kahneman and Tversky's claims about human irrationality by modifying the original experimental tasks to decrease or eliminate judgment biases. This research did three important things: it mobilized a disciplinary return to studying rational judgment; it underscored the methodological point that experimental evidence can only properly support claims about the *particular* ways in which we are rational or irrational in *specific* contexts of reasoning; and, it demonstrated the practical implications of discovering the conditions promoting rational rather than irrational judgment (in particular, such research provided better grounds for

recommendations about how to change contexts, educational strategies, and institutions to improve human judgment).

In the second chapter I argue that these features of contemporary psychological theorizing suggest a normative account of applied cognitive psychology I call *ecological rationalism*. I argue that our social and moral interest in promoting good judgment motivates the distinction between rational and irrational judgment – just as, in medicine, the moral interest in human health motivates the distinction between health and disease. Because discovering conditions that promote rational rather than irrational judgment better grounds recommendations about how to improve human judgment, this justifies a disciplinary preference for discovering conditions that promote rational rather than irrational judgment.

Not all research in cognitive psychology is aimed at the social engineering of good judgment. In contrast to applied cognitive psychology, research in basic cognitive psychology has been interested in imputing cognitive mechanisms that can explain both rational and irrational judgment and to explain the effects of context on judgment. Speaking to these points, psychologist Gerd Gigerenzer has suggested two methodological points on how cognitive processes should be specified for the sake of legitimate or good psychological explanation. He suggests that a cognitive process should be specified so as to capture the conditions of its own validity and accounts for how the process relates to specific contents, contexts, and information formats.

In the fourth chapter, I show that a cognitive psychology that embraces Gigerenzer's methodological suggestions has intimate connections with the explanatory goals of naturalized epistemology – of reliabilism in particular. Both cognitive psychology and reliabilism invoke cognitive processes to explain the psychological transformation of inputs to output-beliefs; and both seek to explain the epistemic status of output-beliefs by reference to the same cognitive process invoked to explain its production. I also make a few observations on how the intimate connection between reliabilism and cognitive psychology recasts traditional challenges facing reliabilism.

The most prominent, recurring theme in my dissertation is the emphasis on how methodology in psychology has significant implications for philosophy. In my chapter on the Gricean turn in psychology, I focus on how a more situational, reflexive

perspective brings important evidential standards to the task of interpretation more generally. In my chapter on ecological rationalism, I argue that the lesson of context-specificity in psychological research informs a normative account of applied cognitive psychology. And, in my final chapter, I argue that methodological claims in psychology can have intimate ties with naturalized epistemology's deeper explanatory interests.

Table of Contents

Abstract

This dissertation is a study of how methodological issues in psychology can have significant implications for philosophical accounts of interpretation, justification, and psychological explanation. In the first chapter, I analyze traditional philosophical accounts of interpretation with an eye to identifying the ways in which philosophers have used rationality as a methodological tool. I argue that these forms of methodological rationalism do not successfully cope with the challenge from the heuristics and biases research program which generally argues that human judgment is irrational.

In the second chapter, I trace cognitive psychology's disciplinary trend to study conditions that facilitate rational rather than irrational judgment. This trend suggests we should seek to make rational judgment an object of study rather than a default methodology for the process of studying psychological judgment. I argue that social and moral interests in promoting cognitive health motivate and justify the interest in discovering conditions that promote rational rather than irrational judgment. I call this normative account of applied cognitive psychology ecological rationalism.

In the third chapter, I argue that psychology's disciplinary interest in creating valid questionnaires motivates discovering the conditions of successful communication. I discuss the methodological lessons that the Gricean turn in psychological research brings to questionnaire design: in particular, the Gricean turn imposes evidential requirements on psychological research about the conditions of successful versus unsuccessful communication for specific contexts and the conversational norms governing communication in experimental conditions.

In the fourth chapter, I argue that some methodological critiques of the heuristics and biases research program have intimate connections to naturalized epistemology: in particular, the ways in which Gerd Gigerenzer thinks cognitive processes should be specified for the sake of explaining human judgment suggest that cognitive psychology

and naturalized epistemology are disciplines with shared explanatory goals. I argue that both invoke cognitive processes to explain the psychological transformation of inputs to output-beliefs; and both seek to explain the epistemic status of output-beliefs by reference to the same cognitive process invoked to explain its production. To close, I make a few observations on how the shared explanatory goals between cognitive psychology and naturalized epistemology recasts traditional challenges facing reliabilism.

Chapter 1

Rationality as Method

Philosophers have long debated over whether human beings are intrinsically rational, and over what the proper norms of rationality ought to be. Aristotle conceived of humans as essentially rational animals whose actions were properly guided by reason in accordance with virtues such as courage, temperance, and generosity. Kant conceived practical reason as a capacity serving the normative function of legislating universally valid laws of morality, where the capacity for practical reason *could not help* but recognize the normative force of those laws. In the latter half of the twentieth century, prominent philosophers such as Donald Davidson and Daniel Dennett have argued that that human beliefs and/or actions are generally or mostly rational.

Social scientists have entered the debate over the nature and extent of human rationality. Economists, for example, have traditionally adopted *Homo economicus* a model of human rationality that explains individual and group behavior as if individuals are utility-optimizing and self-interested. Economists, political scientists, and psychologists sometimes adopt *methodological rationalism*, the claim that, in explaining human behavior, we should initially seek to represent what people do as rational: we should impute desires, beliefs, and other mental states so their observed behavior is rational in relation to those mental states, and so their mental states are rational in relation to each other. If we cannot arrive at rationalizing explanations that comport with the evidence, methodological rationalism allows explanations construing others as irrational. However, methodological rationalism prefers rationalizing intentional explanations over those that construe others as irrational.

Prominent research programs in psychology have cast doubt on economists' claim to model actual choice behavior. Herbert Simon's model of bounded rationality "replace[d] the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist."[1] Daniel Kahneman and Amos Tversky see their work on Prospect Theory as demonstrating that we do not maximize expected utility: we tend to have inconsistent preferences, our preferences sometimes reverse, and we have trouble making accurate probability judgments.[2]

Philosophical arguments for general and methodological rationalism have also been affected by Kahneman and Tversky's research. Philosophers such as Stephen Stich and Paul Thagard point to Kahneman and Tversky's work as contradicting general rationalism.[3] They have invoked these empirical findings in philosophical arguments discrediting the methodological rationalism that often underwrites general rationalism.[4]

However, in the debate over general rationalism, I think philosophers have overlooked important questions about the *methodological* roles rationality can legitimately play in interpretation. As such, this introductory chapter aims to provide a positive critique of traditional accounts of methodological rationalism, with an eye to identify the *methodological roles* that rationality served. My critiques of these traditional accounts suggest, not jettisoning the methodological use of rationality altogether, but reconceiving the methodological roles rationality can legitimately serve in interpretation.

In this chapter, I do not wish to dwell on the question of whether man is "mostly" rational. Rather, my interest is in identifying the methodological roles rationality has

---

[1] Herbert A. Simon, "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics* 69, no. 1 (1955): 99-100. See also, Herbert A. Simon, "Rational Choice and the Structure of the Environment," *Psychological Review* 63, no. 2 (1956).

[2] Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, no. 2 (1979). Amos Tversky and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science* 211, no. 4481 (1981). Daniel Kahneman and Amos Tversky, "On the Psychology of Prediction (1973)," in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (New York, NY: Cambridge University Press, 1982).

[3] Stephen P. Stich, "Could Man Be an Irrational Animal? Some Notes on the Epistemology of Rationality," *Synthese* 64 (1985). Paul Thagard and Richard E. Nisbett, "Rationality and Charity," *Philosophy of Science* 50 (1983).

[4] For a survey of arguments for general rationality, see Edward Stein, *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science* (Oxford, U.K.: Oxford University Press, 1996).

historically played in interpretation for the purpose of retooling them for interpretation in contemporary psychological research. I will begin by considering Max Weber's account of methodological rationalism. Weber adopts rationality as a *simplifying assumption*: he conceives irrational action as a deviation from the rational course as a way to simplify the variables considered relevant to explanation. In this account, Weber makes an important distinction between *rational* and *intelligible* action that opens up the possibility of an important class of interpretive cases: namely, cases in which others' intentional states are irrational yet intelligible. I will focus on this class of cases in my critiques of Davidson's, Dennett's, and Cohen's accounts of methodological rationalism.

Davidson adopts an account that takes the rationality of others' beliefs and meanings to be constitutive of the very notion of belief and meaning. He argues that certain principles of charity are required in belief-desire attribution. By making norms of rationality inherent to the interpretive perspective, Davidson uses rationality as a *framework* within which we build interpretations of others' beliefs, meanings, and desires. Unfortunately, the norms of rationality Davidson takes to be inherent to the interpretive perspective are norms that we do in fact violate; and, his account does not happily accommodate these cases of irrational belief and desire. An important, lasting contribution of Davidson's theory is his observation that any psychological theory on human judgment "must *include* a theory of interpretation" about subjects' beliefs about the experimental task.[5] I pick up on this Davidsonian lesson in chapter 3 which takes very seriously the task of interpreting subjects' beliefs and desires with respect to experimental tasks.

Dennett's methodological rationalism uses rationality as a *predictive tool.* Instrumental rationalism claims that we should describe others *as-if* they are rational agents, because doing so helps us reliably and accurately predict their future behavior. In the face of Kahneman and Tversky's research, Dennett revises his account of rationality. This move seems to suggest a more general strategy that instrumental rationalists might use: if it turns out that one's theory of rationality does not provide for predictive, rationalizing interpretation, one should rethink one's theory of rationality so that it turns

[5] Donald Davidson, "Belief and the Basis of Meaning (1974)," in *Inquiries into Truth and Interpretation* (New York, NY: Oxford University Press, 2001), 147. The italics are mine.

out to be predictive. Dennett's flexibility with respect to the norms of rationality allows the theorist to revise her theory of rationality in light of empirical evidence. Although using rationality as a predictive tool in this way leads to rather ad hoc theories of rationality that lack normative authority, Dennett does touch on an intimate connection that predictive, rationalizing psychological theories can have with naturalized accounts of justification or rationality. I will discuss this connection in greater length in chapter 4.

Cohen argues that rationality should be *descriptively determined*: when we say that most human judgments are rational, all we are saying is that most people have the ability to reason in accordance with rules that describe the deductive and inductive practices of most people. Cohen's account rests on a confused analogy to linguistics and on a misinterpretation of Nelson Goodman's account of reflexive equilibrium. However, I think his account has the potential to capture a notion of "rationality" and "rationalized interpretation" traditionally overlooked by philosophical accounts of interpretation: namely, explanation by social norms. I will argue that interpretation that "rationalizes" human judgment and action in terms of social norms provides an entrée into minimally rationalizing interpretation that is best understood as being motivated by interests in intelligible interpretation.

In what follows, I will review the accounts of interpretation provided by Weber, Davidson, Dennett, and Cohen with an eye to analyze the methodological roles rationality was designated to serve in interpretation. I will argue that these accounts of methodological rationalism fail to meet challenges posed by the heuristics and biases research program, which generally argues that human judgment is rationally defective in certain systematic ways.

1.1 Rationality as a Simplifying Assumption

Historically, methodological rationalism has its roots in Max Weber's account of verstehen. In German, "verstehen" is simply the present participle of the verb "to understand." Weber's theory conceives of this understanding as being from the subjective point of view. Weber's verstehen refers to the process of observing and

interpreting the intentional states of others for the purposes of explaining their actions. To gain an understanding into rational and irrational actions, Weber suggests that we adopt the following interpretive strategy: we should initially seek to interpret others as acting rationally; and, when their actions cannot be construed as being rational, we should invoke factors that explain the act as a deviation from the rational course. Although this interpretive strategy is motivated by an interest in simplifying the factors and motivations we might look to in explaining irrational action, it fails to provide satisfying interpretations for an important class of cases: namely, cases in which individuals are best understood not as *deviating* from the rational course of action due to perturbing factors, but as failing to be disposed towards following the rational course of action at all. Studies in cognitive psychology suggest that this class of cases is significant in understanding human judgment and decision making under uncertainty. Weber's account of verstehen provides a lasting conceptual contribution to methodological rationalism by providing a distinction between rational and intelligible interpretation. This distinction allows for the possibility of interpreting others as being irrational yet intelligible or understandable nonetheless.

1.11 Max Weber's Verstehen

For Weber, what distinguishes sociology from the natural sciences is a difference in interest and perspective. Unlike biologists, sociologists aim to explain and give meaning to social action by reference to the meaningful, subjective states of mind or intentions leading to an individual's act.[6] An action is distinguished from mere behavior insofar as we can impute the acting individual as attaching a subjective meaning – or intention – to it. That action is social insofar as the individual, in acting, takes into account the behavior of others.[7]

The basic object of generalization and explanation in Weber's account of verstehen is his notion of the "pure type." A "pure type" provides a process-model of how – for an average, typical, pure, or ideal kind of case – a type of act follows from a set

[6] Max Weber, *The Theory of Social and Economic Organization*, trans. A. M. Henderson and Talcott Parsons (New York, NY: Oxford University Press, 1947), 101.
[7] Ibid., 88.

of imputed intentional states. Pure types play an epistemic and methodological role akin to natural kind terms. In constructing natural kind terms, we create concepts and classificatory schemes to pick out causal factors responsible for the causally sustained generalizations underwriting our explanations and theories.[8] By identifying projectible terms and theories, our background theories identify which natural kind terms are relevant, and which hypotheses worth testing. In order to achieve generalizations and explanations appropriate to our disciplinary aims and practices, natural kinds are formulated to accommodate the conceptual resources, and the inductive and explanatory practices of the discipline.

Analogously, in constructing a pure type, we seek a vocabulary and conceptual scheme to pick out causally relevant factors and events responsible for an observed type of act. Our background theories suggest relevant factors or events to observe; the same background theories suggest testable generalizations and hypotheses. These factors are described at a level of abstraction appropriate to the conceptual resources and inductive and explanatory practices of the discipline.[9] For Weber, a single act may fall under multiple "pure types" depending on the theorist's background theory or interest.

Weber makes a distinction between the notion of *causal adequacy* and *intelligibility*. Whether the factors and events picked out by a type get to figure in explanation depends on whether the intention-action generalizations picked out by the type are causally adequate in Weber's technical sense: for any purported intentional explanation for an action, we must be able to determine that there is a probability" that "a given observable event (overt or subjective) will be followed or accompanied by another event."[10] This probability, which is on principle "always in some sense calculable," is measured by "the determinable frequency" for an average, typical, pure, or ideal type of

---

[8] Richard Boyd, "Kinds as the 'Workmanship of Men': Realism, Constructivism, and Natural Kinds," in *Rationalitat, Realismus, Revision: Proceedings of the Third International Congress, Gesellschaft Fur Analytische Philosophie*, ed. Julian Nida-Rumelin (Berlin, Germany: de Gruyter, 1999).

[9] Under Weber's account, pure types do not have the explanatory or inductive *power* associated with natural kind terms. Weber admits that the cost of seeking explanations in interpretive terms is "the more hypothetical and fragmentary character of its results" in comparison to the natural sciences. However, this does not diminish sociology's status as a science: as a science, sociology's causal-intentional claim and explanations defer to empirical evidence. See Weber, *The Theory of Social and Economic Organization*, 103-4.

[10] Ibid., 98-9.

case.[11] So, although an interpretation may be quite intelligible, it may fall short of our standards of causal adequacy, and thus, our standards of good interpretation.

Weber's methodological rationalism directs us to define a pure type by the intentional states that would rationalize the act, or by those intentional states that account for the act's deviation from the idealized, rationalized type. Weber suggests that because pure types are abstract and ideal, "it is probably seldom if ever that a real phenomenon can be found which corresponds exactly to one of these ideally constructed pure types."[12] For cases where an act is fully rational, the rationalizing explanation picks out factors normatively and causally relevant to the act.[13] In the majority of cases where the act is less than rational, we seek to identify factors accountable for deviations from the hypothesized ideal type of action. Once we have identified these factors, we can formulate pure types to account for them. For example, Weber recognizes that "irrational" actions sometimes result from strong emotions; and, he suggests that the strong emotions can explain why an agent's act deviated from the ideal type of action.[14]

Weber's argument for methodological rationalism suggests the following ceteris paribus generalizations: other things being equal, we should interpret others as acting rationally; and, when their actions cannot be construed as being rational, we should invoke factors that explain the act as a deviation from the rational course.

1.12 Rationality as a Simplifying Assumption

Weber does not provide empirical arguments for the claims that (i) all things being equal, we would be rational in the absence of interference, and that (ii) irrational behavior is caused by irrational factors responsible for the behavior's deviation from the

---

[11] Notice that this formulation of the causal adequacy condition does not commit Weber to the claim that the probability must be above a certain threshold (for example, better than chance) in order for the intentional explanation to be causally adequate. Ibid., 99-100.

[12] Ibid., 110-1.

[13] Ibid., 92.

[14] "Empathetic or appreciative accuracy is attained when, through sympathetic participation, we can adequately grasp the emotional context in which the action took place. . .The more we ourselves are susceptible to them the more readily can we imaginatively participate in such emotional reactions as anxiety, anger, ambition, envy, jealousy, love, enthusiasm, pride, vengefulness, loyalty, devotion, and appetites of all sorts, and thereby understand the irrational conduct which grows out of them. Such conduct is 'irrational,' that is, from the point of view of the rational pursuit of a given end." Ibid.

rational course. Rather, his arguments are methodological: he adopts these ceteris paribus generalizations as a way to simplify the domain of possible factors and interpretations available to the interpreter.

Under Weber's account of verstehen, the sociologist's methodological burden is to solve the question of how to go about picking out pure types that figure in predictions and causal explanations. But, for any single act, we may tell many complex intentionally-laden stories. There are a seemingly endless number of variables that may be combined in different ways: sensory perceptions, desires, presuppositions, beliefs, plans, commitments, idiosyncratic habits of thought or feeling, prejudices, impressions, and memories.

In response to the problem of numerous, "qualitatively heterogeneous" intentional states, Weber suggests we adopt an expedient, simplifying conceit:[15] "[f]or the purposes of a typological scientific analysis it is convenient to treat all irrational, effectually determined elements of behavior as factors of deviation from a conceptually pure type of rational action."[16] For example, a "panic on the stock exchange can be most conveniently analyzed by attempting to determine first what the course of action would have been if it had not been influenced by irrational affects."[17] For Weber, it is only by describing irrational action as "deviations from this hypothetical course" that we are able to render irrational action intelligible.[18]

Weber clearly states that this use of ideal types does not implicate a "rationalistic bias" in interpretation – it only serves as a necessary simplifying methodological device.[19] But why is it that Weber thinks that irrational action is most intelligible when described as a deviation from a rational, ideal type of action? For Weber, the value of looking to ideal types in interpreting irrational action is that they provide needed classificatory systems in analyzing human motivations: "[t]he more sharply and precisely the ideal type has been constructed, thus the more abstract and unrealistic in this sense it is, the better it is able to perform its methodological functions in formulating the clarification of terminology, and in the formulation of classifications, and of

[15] Ibid., 111.
[16] Ibid., 92.
[17] Ibid.
[18] Ibid.
[19] Ibid.

hypotheses."[20] These classificatory schemes are "better" because they have "the merit of clear understandability and lack of ambiguity."[21]

Weber's idea seems to be that, by invoking pure type classification schemes in the interpretation of irrational action, we have a sharper understanding of the factors involved in irrational action. However, the heuristics and biases literature provides examples of interpretation in which it is not simplest or more accurate to describe subjects as deviating from some ideal, rational course because of some perturbing factor. For example, consider the framing effects literature – a literature brimming with studies that demonstrate systematic preference reversals with changes in decision frames. The classic study demonstrating this is the *Asian disease problem* which asks subjects the following question:

> Problem 1 [$N = 152$]: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
>
> If Program A is adopted, 200 people will be saved. [72 percent]
>
> If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28 percent]
>
> Which of the two programs would you favor?[22]

The wording of this problem suggests that the outcomes of the two alternative outcomes are gains rather than losses: the program *saves* lives. The majority choice in this problem was to choose to save 200 lives rather than the riskier choice: 72 percent preferred the certain over the uncertain gain.[23]

A second group of respondents was given the cover story provided in problem 1, with a negative shift in decision frame: here, the outcomes were described in terms the number of people who *die*:

> Problem 2 [$N = 155$]:
>
> If Program C is adopted 400 people will die. [22 percent]
>
> If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78 percent]

---

[20] Ibid., 111.
[21] Ibid., 92.
[22] Tversky and Kahneman, "The Framing of Decisions and the Psychology of Choice," 453.
[23] Ibid.

Which of the two programs would you favor?

When the logically identical choice problem was described in negative terms, the majority choice reversed: the certain death of 400 people was less acceptable than the riskier choice: 78 percent preferred the uncertain over the certain loss.[24] Changing to a within-subject design did not vary subject responses.[25]

It seems that changing the decision frame from positive to negative induced a reversal of preference in violation of the principle of invariance.[26] The *invariance principle* tells us that, no matter how it is that we describe a decision situation, these varying linguistic representations should yield the same preference: in other words, preferences should be invariant under logically equivalent descriptions of the acts, outcomes, and states of nature constituting a decision situation.

I think that Kahneman and Tversky's framing effects studies do not fit with Weber's use of rationality as a simplifying tool. Weber's use of rationality as a simplifying assumption might suggest that we interpret subject violations of the invariance principle as a *deviation* from the more rational course because of some perturbing factor. Kahneman and Tversky discovered that it is not that subjects' judgments *deviate* from an otherwise hypothetical, rational course: Kahneman and Tversky's point is that subjects simply do not seem disposed to follow the rational course. Expected utility theory demands that subjects seek to maximize *total* states of wealth or welfare (i.e., the total expected states of living/dead). Kahneman and Tversky's work suggests that subject responses are not sensitive to total or absolute states of wealth or welfare. Rather, it seems that subjects are sensitive to *changes* in total states of wealth or welfare: in the Asian Disease problem, subjects are sensitive to the perceived loss or gain of life, rather than to the total numbers of expected living/dead. Because subjects are sensitive to gains and losses rather than to total or absolute states of wealth or

---

[24] Ibid.

[25] For example, see Ibid.: 455. Amos Tversky and Daniel Kahneman, "Rational Choice and the Framing of Decisions," *The Journal of Business* 59, no. 4, Part 2: The Behavioral Foundations of Economic Theory (1986): S255, S68-9. However, others have found that subjects show risk aversion in both gain and loss decision frames. For example, see Jerwen Jou, James Shanteau, and Richard Jackson Harris, "An Information Processing View of Framing Effects: The Role of Causal Schemas in Decision Making," *Memory & Cognition* 24, no. 1 (1996).

[26] Tversky and Kahneman, "The Framing of Decisions and the Psychology of Choice," 453.

welfare, Kahneman and Tversky describe subjects as being sensitive to factors that it is irrational or illegitimate for them to be sensitive to from the perspective of expected utility theory. In this case, thinking of subjects as being disposed to follow expected utility theory's hypothetical, rational course does no crucial work in simplifying our interpretations of subjects' irrational choices in this kind of case.

1.13 Rational versus Intelligible Interpretation

Even though rationality does not always usefully serve to simplify our interpretations of irrational action, Weber makes an important distinction that is crucial for the interpretation of irrational belief and action. On the one hand, Weber's account allows for the interpretation of rational action, which he conceptualizes as action that follows the rules of logic, mathematics, and means-end reasoning.[27] On the other hand, he allows for the possibility of interpreting irrational action, so long as that irrational action is *intelligible*. What Weber ultimately seeks in verstehen is "subjectively understandable" interpretation, where an action is "subjectively understandable" so long as we can understand "the subjective 'states of mind' of actors."[28]

Weber rightly points out that we can understand the subjective states of minds of others even when they make irrational judgments and choices. For example, he suggests we use empathy when it comes to interpreting irrational action:

> we are able to understand errors, including confusion of problems of the sort that we ourselves are liable to, or the origin of which we can detect by sympathetic self-analysis. . . Even when such emotions are found in a degree of intensity of which the observer himself is completely incapable, he can still have a significant degree of emotional understanding of their meaning and can interpret intellectually their influence on the course of action and the selection of means.[29]

Weber seems to think that the intelligibility of irrational actions is somehow less than the intelligibility of rational actions.[30] However, irrational action is intelligible from a subjective point of view nonetheless.

---

[27] Weber, *The Theory of Social and Economic Organization*, 90-1.
[28] Ibid., 88.
[29] Ibid., 92-4.
[30] Weber says: "The highest degree of rational understanding [of subjective meaning] is attained in cases involving the meanings of logically or mathematically related propositions; their meaning may be immediately and unambiguously intelligible. We have a perfectly clear understanding of what it means

Weber's account of verstehen suggests using rationality as a kind of simplifying assumption for the interpretation of irrational action. In particular, he suggests that we interpret irrational action as a deviation from the rational course. However, this interpretive strategy fails to capture psychologically significant cases in which there are no perturbing factors responsible for the "deviation" from the rational course because the actors were not disposed to follow the rational course in the first place. Yet, Weber's fundamental interest in intelligible or subjectively understandable interpretation opens up an important class of interpretive cases: namely, cases in which others' intentional states are irrational yet intelligible from a subjectively sympathetic point of view. I will focus on this class of cases in my critiques of Davidson's, Dennett's, and Cohen's accounts of methodological rationalism.

1.2 Rationality as Framework

For Davidson, the norms of rationality are constitutive norms of belief and preference: that is, they are a priori conditions on the applicability of the concepts of belief and preference.[31] This is so because the principles of charity that Davidson takes to be required in belief-desire attribution imply that norms of rationality are inherent to the interpretive perspective. Rationality here provides a kind of framework within which we build interpretations of others' beliefs, meanings, and desires. Without the framework, there is no basis on which to build our interpretations: we simply lose the ability to gain evidence of others' intentional states. Unfortunately, the norms of

---

when somebody employs the proposition 2 x 2 =4 or the Pythagorean theorem in reasoning or argument, or when someone correctly carries out a logical train of reasoning according to our accepted modes of thinking. In the same way we also understand what a person is doing when he tries to achieve certain ends by choosing appropriate means on the basis of the facts of the situation as experiences has accustomed us to interpret them. Such an interpretation of this type of rationally purposeful action possesses, for the understanding of the choice of means, the highest degree of verifiable certainty. With a lower degree of certainty, which is, however, adequate for most purposes of explanation, we are able to understand errors, including confusion of problems of the sort that we ourselves are liable to, or the origin of which we can detect by sympathetic self-analysis." Ibid., 91.

[31] Many thanks to Peter Railton for substantial guidance on interpreting Davidson's methodological use of rationality.

rationality Davidson takes to be inherent to the interpretive perspective are norms that we do in fact violate; and, his account does not happily accommodate these cases of irrational belief and desire. What Davidson's theory contributes towards interpretation in experimental contexts is the lesson that any psychological theory on human judgment "must *include* a theory of interpretation" about subjects' beliefs about the experimental task.[32] I pick up on this Davidsonian lesson in chapter 3 which takes very seriously the task of interpreting subjects' beliefs and desires with respect to experimental tasks.

1.21 Davidson's Principle of Charity

Davidson stages his account of interpretation and methodological rationalism in the context of radical translation. Radical translation is the problem of translating a foreign language without having either a speaker's beliefs or the meaning of her sentences in advance. The problem is that belief and meaning are epistemologically interdependent: we can infer what someone believes if we understand the meaning of her utterances; and we can infer what she means by an utterance, if we know what she believes. Thus, the task of radical translation is to simultaneously deliver a theory of belief and a theory of meaning.

Davidson takes the epistemic interdependence of belief and meaning as a deep fact about interpretation. In light of this, he believes that the only viable solution for getting a toehold in radical interpretation is to adopt two basic principles of charity. The first principle of charity, Davidson's *charitable principle of truth*, is the more famous of the two. It "directs the interpreter to translate or interpret so as to read some of his own standards of truth into the pattern of sentences held true by the speaker."[33] The interpreter uses external causes of assent as evidence for the speaker's beliefs and meanings: that is, the content of the speaker's beliefs and utterances are given by the external states of affairs that prompt him to hold sentences true.

Since there are a limited number of sentences that the interpreter recognizes as true under the conditions of assent, the principle of charity acts to restrain "the degrees of

---

[32] Davidson, "Belief and the Basis of Meaning (1974)," 147. The italics are mine.
[33] Donald Davidson, "A Coherence Theory of Truth and Knowledge," in *Epistemology: An Anthology*, ed. Ernest Sosa and Jaegwon Kim (2000: Blackwell Publishers Ltd., 2000), 160.

freedom allowed belief while determining how to interpret words."[34] This turns "the methodological problem of interpretation" into one of determining "how, given some sentence a man accepts as true under a given circumstance, to work out what his beliefs are and what his words mean."[35] Davidson suggests that the charitable principle of truth be invoked to discriminate between competing theories of interpretation: charity's "basic methodological precept" identifies "a good theory of interpretation" as one which "maximizes agreement" about what is true.[36]

This allows Davidson to employ a theory of meaning in which the interpreter formulates a Tarski-like truth theory *T* for an object language *L*.[37] For each sentence *s* of *L* the truth theory *T* (which is expressed in the metalanguage) entails a *T*-sentence that gives *s*'s truth conditions. The idea is that "the definition works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence."[38] Under Davidson's account of interpretation, truth conditions, stated in one's own language, form the basis of meaning for a foreign language.

Davidson recognizes that, even though the charitable principle of truth allows us to use truth conditions as the basis of belief and meaning, this is not sufficient for adequate translation: *T*-sentences alone are not sufficient for adequate translation. The problem is that there are too many true things – too many possible *T*-sentences – that hold at the time and context of a given utterance. These possible *T*-sentences may all be extensionally adequate: that is, their empirical implications may be true. Yet, they can still fail to capture the necessary and sufficient conditions relevant to the truth of the utterance in question – and thus, fail to capture the meaning of the utterance. Davidson considers the following extensionally adequate *T*-sentence: "'Snow is white' is true-in-English iff grass is green." This case is supposed to work because (Davidson imagines) both are always true, and so have the same extension. Given only the conditions of

---

[34] Ibid.
[35] Donald Davidson, "Thought and Talk (1975)," in *Inquiries into Truth and Interpretation* (New York, NY: Oxford University Press, 2001), 162.
[36] Ibid., 169.
[37] Donald Davidson, "Truth and Meaning," in *Inquiries into Truth and Interpretation* (New York, NY: Oxford University Press, 2001).
[38] Ibid., 24.

utterance, "there are many things people do believe, and many more" true things "that they could" believe.[39]

To reduce the indeterminacy of meaning, Davidson adopts a holist constraint on meaning. Davidson's holist constraint requires that, in order to identify or translate any belief, we must refer to the total "pattern of belief." The idea is supposed to be that, by identifying the conditions of utterance for "Snow is white," and figuring out how this belief fits intelligibly with other beliefs imputed to the speaker, we can distinguish legitimate translations from extensionally adequate *T*-sentences. A speaker's expressed beliefs about "snow" and "white" things, Davidson imagines, will prevent the interpreter from translating "Snow is white" as "Grass is green."

Davidson recognizes, however, that holism only resolves the problem of extensionality if the interpreter can safely assume that the speaker's expressed beliefs about "snow," "white" things, and "Snow is white" are connected to one another in consistent and otherwise intelligible ways. To bridge this gap, Davidson adopts the second, lesser known principle of charity: the charitable principle of rationality. The *charitable principle of rationality* urges interpreters to construe others' beliefs and desires as being maximally coherent,[40] where relations of coherence are defined by the theorist's principles of grammar, evidence,[41] logic, and set theory.[42] Ultimately, the coherence theory of belief is meant to pick up the slack between what a speaker could possibly mean/believe and what she actually means/believes.[43] The methodological advantage of the charitable principle of rationality is that it simplifies recommendations about how to cope with irrational or mistaken beliefs: for Davidson, the "best explanation" for erroneous belief is "nothing but epistemology seen in the mirror of meaning:" we should minimize the epistemic sins committed by the speaker.[44]

Davidson's argumentative goal is to demonstrate that both the charitable principles of truth and rationality are necessary for overcoming the interdependence of belief and meaning in radical translation: "[c]harity is forced on us" in order "to

---

[39] Davidson, "A Coherence Theory of Truth and Knowledge," 155.
[40] Davidson, "Thought and Talk (1975)," 169.
[41] Davidson, "A Coherence Theory of Truth and Knowledge," 159.
[42] Davidson, "Thought and Talk (1975)," 169.
[43] Davidson, "A Coherence Theory of Truth and Knowledge," 159.
[44] Davidson, "Thought and Talk (1975)," 169.

understand others."[45] And, because these principles of charity are necessary for the very possibility of interpretation, it follows that beliefs and meanings are generally or mostly rational: that is, the norms inherent to the interpretive stance make it impossible not to discover anything but the general rationality of others' beliefs and utterances.

1.22 Irrational Belief and Desire

Some of the studies from the heuristics and biases research program seem to demonstrate that human judgments systematically violate a priori rules of probability and rational choice: for many of Kahneman and Tversky's classic studies, the most plausible interpretation of subjects' responses – given the evidence gained by those particular studies at that point in time – was the one provided by the uncharitable researchers.[46] For example, Kahneman and Tversky discovered that subjects are more swayed by anecdotal evidence than base rate information in making conditional probability judgments. Consider their famous lawyer-engineer question:

> A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.
>
> The same task has been performed by a panel of experts, who were highly accurate in assigning probabilities to the various descriptions. You will be paid a bonus to the extent that your estimates come close to those of the expert panel.[47]

In Kahneman and Tversky's original study, the *low-engineer group* was told that there were 30 engineers and 70 lawyers. The *high-engineer group* was told that there were 70 engineers and 30 lawyers. Both groups were provided the same five personality descriptions, most of which were stereotypical of an engineer or lawyer. Kahneman and Tversky offered the following as an example of one of the personality descriptions:

---

[45] Donald Davidson, "On the Very Idea of a Conceptual Scheme (1974)," in *Inquiries into Truth and Interpretation* (New York, NY: Oxford University Press, 2001), 197.

[46] Attempts to rationalize subject responses required more complex and less plausible assumptions than the uncharitable interpretation originally adopted by Kahneman and Tversky – at least until more evidence was discovered in subsequent years by social psychologists. I will say more about further evidence that psychologists subsequently discovered – evidence suggesting that the best interpretation is not one that construes human judgment as systematically irrational – in chapters 2 and 3.

[47] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 53.

> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.
> The probability that Jack is one of the 30 engineers in the sample of 100 is ------%.

Kahneman and Tversky found that subjects' predictions about how probable it was that a given person was an engineer or lawyer were independent of the base rates of engineers/lawyers, in violation of Bayes' Rule. In light of this evidence, it seems that the simplest and best interpretation is that, given conditional probability questions of this form, subjects systematically violate Bayes' Rule. This example provides at least a *prima facie* case in which the best interpretation construes subjects as violating a priori rules that Davidson identifies as constitutive norms of rationality.

In the face of such cases, Davidson has three possible options. First, he can reject the possibility that we systematically violate certain a priori rules. Second, Davidson can reject certain a priori rules as requirements of rationality, and argue that less strict standards of rationality are inherent in the interpretive perspective. Or, third, Davidson can admit that we are systematically irrational in the sense that we regularly violate specific a priori rules but provide some way of accommodating error without undermining his commitment to using rationality as a framework in interpretation. In considering which of these responses Davidson would take, I will look to his account of interpretation of choice behavior because this work focuses explicitly on problems of interpretation in psychological research.

Davidson's first possible response is to reject the possibility of discovering that humans systematically violate certain a priori rules. It seems that Davidson hews to this option in the case of human choice and the rational choice axioms. When it comes to the attribution of desires, Davidson's coherence constraint directs the interpreter to construe those desires (understood as preferences) as conforming to rational choice axioms.[48] In the 1950's, he and Patrick Suppes worked on *Subjectively Expected Utility* (SEU) theory, a psychological model that described people as maximizing the product of their subjective utility and subjective probability, and that accordingly assigned subjective probabilities (beliefs) and utilities (desires) to subjects in ways that preserved the axiom

[48] Donald Davidson, "Psychology as Philosophy (1974)," in *Essays on Actions and Events* (Oxford, UK: Oxford University Press, 1980), 236-7.

of transitivity.[49] His commitment to the consistency of preferences indicates that he takes this to be fundamental to the very possibility of arriving at an interpretative theory or psychological theory. For him, preferences are defined in terms of their transitivity with respect to one another, just as mass is defined in terms of the transitivity of the relation "heavier than."

> Just as the satisfaction of the conditions for measuring length or mass may be viewed as constitutive of the range of application of the sciences that employ these measures, so the satisfaction of conditions of consistency and rational coherence may be viewed as constitutive of the range of applications of such concepts as those of belief, desire, intention and action.[50]

For Davidson, transitivity is a constitutive norm of preferences, and – as such – are a priori conditions on applicability of the concept of preference. So, if we made measurements that violated the transitivity of preferences, we would deny that we have measured "preferences" or chalk the inconsistency up to experimental error: if the norm of transitivity does not apply, then neither does the concept of preference. Davidson does not accept the possibility that we are fundamentally irrational in certain systematic ways. Indeed, he thinks it would be impossible to design an experiment to test whether beliefs or desires violated those norms:

> It is not easy to describe in convincing detail an experiment that would persuade us that the transitivity of the relation of *heavier than* failed. Though the case is not as extreme, I do not think we can clearly say what should convince us that a man at a given time (without any change of mind) preferred *a* to *b*, *b* to *c*, and *c* to *a*. The reason for our difficulty is that we cannot make good sense of an attribution of preference except against a background of coherent attitudes.[51]

If we abandon transitivity of preferences, we would abandon the framework in which desire attributions are built. This is why Davidson claims that "[t]o see too much unreason on the part of others is simply to undermine our ability to understand what it is they are so unreasonable about."[52]

Rationality is the only framework within which we can build interpretations of others' beliefs, meanings, and desires. Without the framework,

---

[49] Ward Edwards, "Behavioral Decision Theory," *Annual Review of Psychology* 12 (1961): 474.
[50] Davidson, "Psychology as Philosophy (1974)," 236-7.
[51] Ibid.
[52] Davidson, "Belief and the Basis of Meaning (1974)," 153.

there is no basis on which to attribute intentional states: we simply lose the ability to gain evidence of others' beliefs, meanings, and desires.[53]

> charity is not an option, but a condition of having a workable theory. . . If we can produce a theory that reconciles charity and the formal conditions for a theory, we have done all that could be done to ensure communication. Nothing more is possible, and nothing more is needed.[54]

Such a strong account of the rationality of beliefs and desires is methodologically undesirable insofar as it precludes the intentional understanding of a psychologically interesting class of interpretive cases: cases in which people hold systematically irrational beliefs and desires.

In the face of these cases, Davidson's second option is to conclude that certain a priori rules – such as the axioms of rational choice theory – fail to capture the requirements of rationality. Davidson's commitment to SEU theory suggests that he would reject this option. Davidson's explicit definitions of the coherence relations for belief, which implicate a priori rules of logic, probability, and set theory[55] suggests Davidson is likewise committed to canonizing these a priori formal rules in a theory of rationality.

Davidson's third option is to admit that we are fundamentally irrational in the sense that we systematically violate a priori rules identified as norms of rational belief and choice. Sometimes Davidson seems to adopt this approach. In accounting for how it is that the interpreter should go about diagnosing which of a speaker's beliefs are irrational, Davidson explicitly turns to psychological theory.[56] He suggests that, in

---

[53] Thanks to Railton for a heavy dose of interpretive guidance in these points.
[54] Davidson, "On the Very Idea of a Conceptual Scheme (1974)," 197.
[55] Davidson, "Thought and Talk (1975)," 169.
[56] If we look more closely at Davidson's explication of radical translation, we find that he implicitly relies on psychological theory for translation. Recall that, in describing how the conditions of utterance inform translation, Davidson modestly and plausibly assumes that it will generally be salient to the interpreter *which* aspect(s) of the conditions are *relevant* to the correct translation of the utterance. This is because Davidson allows the interpreter a modest psychological theory about what conditions of utterance are *causally* relevant to the utterance: "Communication begins where causes converge: your utterance means what mine does if belief in its truth is systematically *caused by* the same events and objects." By having this kind of modest psychological theory, an interpreter can determine that, by "Snow is white," the speaker means to say something about snow rather than grass, or about whiteness rather than greenness, by tracking the speaker's eye movements – his focus on the white snow falling around them. For the class of ostensible objects and for other claims, we might turn to psychology to tell us about the relative likelihood of what someone believes: for example, he says we must "[t]ake the objects of a belief to be the causes of that belief. And what we, as interpreters, must take them to be is what they in fact are. . . We can't in general first identify beliefs and meanings and then ask what caused them. The causality plays an indispensable

attributing irrational and false beliefs, the interpreter is to rely on "common-sense or scientific knowledge of explicable error."[57] In particular, diagnosing which among a speaker's beliefs are irrational should be guided by knowledge about "the relative likelihood of various kinds of mistakes."[58]

Davidson's reference to psychological theory is not meant to threaten the charitable principle of rationality's prescription to construe others as being maximally rational given the parameters of actual judgment and choice. This is because, under his account, even psychological theory must include theories of interpretation; and, these included theories of interpretation necessarily rely on the charitable principle of rationality. Davidson has pointed out that psychologists studying human performance on rational choice tasks can only interpret subjects' beliefs, desires, and meanings with reference to a theory of interpretation: "the attribution of desires and beliefs (and other thoughts) must go hand in hand with the interpretation of speech," and "neither the theory of decision nor of interpretation can be successfully developed without the other."[59] Any psychological theory on human judgment "must *include* a theory of interpretation."[60] And, since all cases of interpretation are a species of radical interpretation for Davidson, he thinks that psychological theories are themselves constrained by the charitable principle of rationality. However, if Davidson takes it to be impossible to interpret subjects as violating a priori rules of logic, probability, and preference, then this overly limits what it is that psychologists can be said to discover.

1.23 The Davidsonian Lesson

I agree with Davidson that psychologists should concern themselves with good interpretation; however, good interpretation does not require rationalizing others' beliefs

---

role in determining the content of what we say and believe." Davidson, "A Coherence Theory of Truth and Knowledge," 161.

[57] Davidson, "On the Very Idea of a Conceptual Scheme (1974)," 196.

[58] Donald Davidson, "Radical Interpretation (1973)," in *Inquiries into Truth and Interpretation* (New York, NY: Oxford University Press, 2001), 136.

[59] Davidson, "Thought and Talk (1975)," 163.

[60] Davidson, "Belief and the Basis of Meaning (1974)," 147. Italics mine. This idea resonates throughout Davidson's work. For example, he says "the attribution of desires and beliefs (and other thoughts) must go hand in hand with the interpretation of speech." Davidson, "Thought and Talk (1975)," 163.

in the way Davidson imagines. What we want from interpretation is intelligibility: we want to gain insight into the intentional states of others and how they are or are not related to their observed behavior. Intelligible interpretation does not require conformance to rules of rationality, though I concede that interpretation might require some *minimal* rationality: in order to interpret a speaker's beliefs and desires, we need to be able to assume that her meanings and beliefs do not vary wildly: we assume that she is generally consistent in the meaning/use of an utterance-*p*, where a speaker is said to be consistent in the meaning/use of utterance-*p* so long as the meaning of her utterance-*p* does not change radically on each occasion of utterance.[61] We assume that her utterance-*p* means the same thing, because we assume that her belief-*p* has not inexplicably changed. The need for the speaker to have some consistency in her meaning/use in utterance-*p*, and some degree of constancy in her belief-*p* helps to explain why Davidson is right in saying that "disagreement about the truth of attributions of certain attitudes to a speaker by that same speaker may not be tolerable at all, or barely."[62] It seems that the intelligibility of interpretation requires some minimal consistency and constancy of belief.

As Weber's account of verstehen suggests, intelligibility also seems to implicate our capacity for empathy: that is, our ability to "adequately grasp the emotional context in which the action took place" through "sympathetic participation."[63] The more susceptible we are to the emotions or emotional dispositions of others, the more equipped we are to imaginatively participate "in such emotional reactions as anxiety, anger, ambition, envy, jealousy, love, enthusiasm, pride, vengefulness, loyalty, devotion, and appetites of all sorts," including especially "irrational conduct."[64] We can gain access to others' intentional states without this kind of empathetic understanding: it is just one method by which we can gain an understanding of others' irrational beliefs and desires.

---

[61] This is not to say that the same utterance-*p* necessarily has an identical meaning on all occasions of use. However, it is in value of the speaker's having been generally consistent in her different usage that enables the interpreter to distinguish the different meanings of and expressed beliefs conveyed by utterance-*p*.
[62] Davidson, "Thought and Talk (1975)," 169.
[63] Weber, *The Theory of Social and Economic Organization*, 91-2.
[64] Ibid., 92. Thanks to Elizabeth Anderson for this point.

For Davidson, the norms of rationality are inherent in the interpretive perspective and serve as a framework within which interpretations of others' beliefs, meanings, and desires are built. Without the framework, there is no basis on which to build our interpretations: we simply lose the ability to gain evidence of others' intentional states. This is why Davidson claims that it is impossible to create empirical studies that demonstrate certain kinds of irrational judgment and preference. Unfortunately, this position precludes the very possibility of interpreting people as having beliefs and desires that violate those norms of rationality – a significant class of interpretive cases in psychological research. However, Davidson does make an observation of lasting significance to research on heuristics and biases: namely that all psychological theories on human judgment "must *include* a theory of interpretation" about subjects' beliefs about the experimental task.[65] I pick up on this Davidsonian lesson in chapter 3: I argue that psychologists working on conversational pragmatics take very seriously the task of interpreting subjects' beliefs and desires with respect to experimental tasks. Indeed, this basic Davidsonian insight brings with it important evidential and methodological standards for psychological experiments on human judgment.

1.3 Rationality as Predictive Tool

Dennett suggests that we use our theories of rational belief, desire, and action as tools to predict future behavior. When pressured to predict and account for cases of irrational belief and choice, Dennett moved to adopt a less strict account of rationality. This move suggests a general strategy instrumental rationalism can employ to preserve the claim that describing others *as-if* they are rational agents is instrumental towards predicting future behavior: if it turns out that a theory of rationality does not provide for predictive, rationalizing interpretation, then modify it until it is predictive. However, this somewhat ad hoc method of arriving at a theory of rationality does not necessarily lead to a theory of rationality we would recognize as normative: just because we tend to believe and behave in certain ways does not necessarily imply that we condone such patterns of

[65] Davidson, "Belief and the Basis of Meaning (1974)," 147. The italics are mine.

belief or behavior. However, Dennett does pick up on an intimate tie that rationalizing psychological theories can have with theories of naturalized epistemology.

1.31 Dennett's Intentional Stance

Instrumental rationalism claims that we should describe others *as-if* they are rational agents because doing so is useful in reliably and accurately predicting future behavior. Instrumental rationalism is compatible with instrumentalism about belief and intentional states more generally. This is because instrumental rationalism is not concerned to identify deep-lying motivations as real causes for behavior. Indeed, instrumental rationalism holds there is no fact of the matter about what intentional states others *really* do or do not have. Rather, their goal is to assign rationalized, intentional states for the sake of arriving at predictive theories.

Dennett's *intentional stance* provides a classic account of instrumental rationalism. His account defines a system as intentional if and only if we can "reliably and voluminously" predict its behavior by assuming that it will behave rationally.[66] Dennett's intentional stance applies not just to humans, but to any system for which this interpretive strategy bears a predictive theory: Dennett suggests the intentional stance works on other living beings (such as birds, fish, reptiles, insects, and clams) as well as artifacts (such as thermostats and lightening).[67]

Dennett adopts an instrumental account of belief: if we fail to describe or predict the system's behavior as-if it were rational, then the intentional stance cannot describe the system as having beliefs or as being intentional at all. For Dennett, "*all there is* to really and truly believing that p (for any proposition p) is being an intentional system for which p occurs as a belief in the best (most predictive) interpretation."[68] Claims like these serve to underscore the instrumental rationalist's primary interest – not in *true* interpretations – but in predictive interpretations.

---

[66] Daniel C. Dennett, "True Believers: The Intentional Strategy and Why It Works," in *The Intentional Stance* (Cambridge, MA: The MIT Press, 1987), 15.
[67] Ibid., 22.
[68] Ibid., 29.

Dennett argues that his instrumental account of intentionality and rationality makes his account of interpretation no less objective. He claims that what is objective about interpretation are "the *patterns* in human behavior that are describable from the intentional stance."[69] It is the observable patterns in human behavior that are objective. And, whether or not our rationalizing interpretations provide for predictive models of those patterns of behavior is itself an objective matter.

1.32 Rationality as a Predictive Tool

Notice that instrumental rationalism, as I have stated it, is agnostic with respect to which theory of rationality we ought to adopt in interpreting others. Instrumental rationalism directs us to describe others *as-if* they are rational agents because doing so is instrumental towards reliably and accurately predicting their future behavior. *Which* account of rationality we ought to adopt is left open. All that instrumental rationalism demands is that – whatever account of rationality we adopt – it is such that it makes for predictive, rationalizing, intentional interpretation.

Yet, Dennett's intentional stance seems to bring with it a particular account of rational belief, desire, and choice. In 1981, Dennett suggested that intentional-system theory is a "close kin of – and overlapping with – such already existing disciplines as epistemic logic, decision theory and game theory, which are all similarly abstract, normative and couched in intentional language."[70] In 1987, Dennett suggests that the "ideal of perfect rationality" – a priori rules of logic, decision theory, and game theory – should be where an interpreter *begins* interpretation: we begin prediction with the strong "assumption that people believe all the implications of their beliefs and believe no contradictory pairs."[71]

He provides more substantive principles of belief that allow interpreters to infer a system's belief based on what it rationally ought to believe given its goals and circumstances: he suggests that "[e]xposure to x, that is, sensory confrontation with x

---

[69] Ibid., 25.
[70] Daniel C. Dennett, "Three Kinds of Intentional Psychology," in *Reduction, Time, and Reality*, ed. R. A. Healey (New York, NY: Cambridge University Press, 1981), 19.
[71] Dennett, "True Believers: The Intentional Strategy and Why It Works," 21.

over some suitable period of time, is the *normally sufficient* condition for knowing (or having true beliefs) about *x*."[72] However, exposure to *x* is not a sufficient condition for knowing about *x* if *x* is wholly irrelevant to one's interests. For this kind of problem, Dennett suggests we "attribute as beliefs all the truths *relevant* to the system's interests (or desires) that the system's experience to date has made available."[73] And, we should attribute "desires for those things a system believes to be good for it" and attribute "desires for those things a system believes to be best means to other ends it desires."[74]

In 1990, Stephen Stich observed that Dennett's idealized account of rationality – what I will call the *idealized intentional stance* – does not serve as a predictive tool. The problem is that an ideal account of rationality does not have the theoretic resources to predict familiar cases of mistaken and irrational belief.[75] In every day life, we readily recognize that we do not strictly conform to the rules of formal logic: we do not believe all the implications of our beliefs; and we are vulnerable to harboring the occasional inconsistent belief. The heuristics and biases research program seems to demonstrate that, not only do our beliefs violate rules of formal logic, they also violate a priori rules of probability and rational choice theory. The intentional stance – when coupled with an ideal account of rationality – has no way of predicting systematic and mundane violations of the rules of formal logic, probability, and rational choice theory.

In light of these psychologically significant cases, it seems there is a more predictive theory available to us: folk psychology. As Stich observes, our ordinary, folk psychological intentional explanations easily predict and describe our various cognitive failings (such as mistakes in calculation or wishful thinking) and our moral failings (including akratic acts). Folk psychology also has a capacious conceptual apparatus to explain systematic failures in terms of vices, distractions, personality defects, etc. Ironically, Dennett's idealized intentional stance account suffers from the same flaw as Davidson's account of interpretation: their commitment to describing others as conforming to idealized rules of reasoning leave no room for intentional explanations for systematic violations of those rules.

---

[72] Ibid., 18.
[73] Ibid. Italics mine.
[74] Ibid., 20.
[75] Stephen P. Stich, "Dennett on Intentional Systems," in *Mind and Cognition: A Reader*, ed. William Lycan (Oxford, UK: Basil Blackwell Ltd, 1990), 173.

If Dennett's account of instrumental rationalism is really interested in predictive interpretation, then Dennett must amend his position or provide further argument for the idealized intentional stance. Dennett has two major responses to the problem of predicting, describing, and explaining irrational belief. His first response is to toe the line. If "there is no *saving* interpretation – if the person in question is irrational – no interpretation at all will be settled on" that attributes beliefs and desires to the system.[76] Instead, we are to "descend from the level of beliefs and desires to some other level of theory" afforded by the design or physical stance (where the design stance predicts behavior by looking at how the system was designed to behave, and the physical stance predicts its behavior by looking at its physical constitution, environment, and the physical laws governing its behavior).[77] It is hard to take this response seriously: just because we cannot describe others as conforming to a priori rules of rationality does not imply that we cannot render their beliefs and desires *intelligible* in interpretation and explanation.

Indeed, it seems Dennett is not willing to adopt this position. When it comes to human behavior, Dennett claims there is an

> *unavoidability of the intentional stance with regard to oneself and one's fellow intelligent beings.* This unavoidability is itself interest relative; it is perfectly possible to adopt a physical stance, for instance, with regard to an intelligent being, oneself included, but not to the exclusion of maintaining at the same time an intentional stance with regard to oneself at a minimum, and one's fellows *if* one intends, for instance, to learn what they know.[78]

I take Dennett's statements about the unavoidability of the intentional stance to support my claim that he would not choose to descend to the physical or design stances in interpreting human cognitive foibles.

We can concede to Dennett that *intentional interpretation* is valuable/desirable; however, why should we think that the idealized intentional stance is required for intentional interpretation? In some passages, Dennett wants to provide evolutionary grounds for thinking that the idealized intentional stance makes for reliable, predictive interpretation. He claims that evolution "guarantees" that the intentional stance will always turn out to provide the most predictive theory because "[t]he fact that we are

---

[76] Daniel C. Dennett, "Making Sense of Ourselves," in *Mind and Cognition: A Reader*, ed. William Lycan (Oxford, UK: Basil Blackwell Ltd, 1990), 188.
[77] Ibid., 187.
[78] Dennett, "True Believers: The Intentional Strategy and Why It Works," 27.

products of a long and demanding evolutionary process *guarantees* that using the intentional strategy on us is a safe bet."[79]

Unfortunately for Dennett, evolution provides no such guarantee. We have Darwinian reasons for thinking that humans have evolved to harbor systematically unreliable mechanisms of inference, desire, and choice. Evolutionary processes do not fully optimize for their environments: for example, our visual capacities are not optimized for seeing colors in dim lighting. In addition, irrational belief, desire, and choice processes may have had decisive survival value: for example, overgeneralizing about what foods might be poisonous based on one's experiences of feeling ill may lead a forager to have mostly false beliefs about what is edible; however, this conservative strategy may have had decisive survival value.[80] By 1990, Dennett seems to concede that "obvious counterexamples" of "evolved manifest irrationality" undermine his account of the idealized intentional stance.[81]

Dennett's evolutionary argument for methodological rationalism might be construed as arguing for a different kind of claim. He might not be claiming that evolution guarantees the reliability of *all* of our cognitive processes – a result that would obviate the possibility that others are irrational. Rather, he might be claiming that evolution guarantees the reliability of our evolved capacity for using the idealized intentional stance in predicting others. It is not that others are guaranteed to be ideally rational. It is that we can't help but *see* them as being ideally rational; and this evolved general strategy of interpretation is a reliable way of predicting their behavior.

The evolutionary story behind the reliability of the intentional stance would go something like this: the intentional stance – this evolved capacity for a rationalizing theory of mind – allowed our ancestors to predict the behavior and thoughts of others in order to confer the evolutionary advantage of coordinating with and protecting themselves against other people.[82] Evolution guarantees a co-evolution of conduct and theory of mind. We might, for example, see our ability to recognize emotions from facial

---

[79] Ibid., 33. Italics mine.
[80] Stich, "Dennett on Intentional Systems," 177.
[81] Dennett, "Making Sense of Ourselves," 192.
[82] R. Axelrod and W. Hamilton, "The Evolution of Cooperation," *Science* 211 (1981), A. Leslie, "The Theory of Mind Impairment in Autism: Eidence for a Modular Mechanism of Development?," in *Natural Theories of Mind*, ed. A. Whiten (Oxford, U.K.: Blackwell Publishers, Ltd., 1991).

expressions as an evolved, adaptive capacity for dealing with problems of cooperation and trust.[83]

This is an interesting twist in the argument. However, our folk psychological ways of explaining others' irrational and mistaken beliefs and behavior belies a crucial premise: namely, that our inescapable theories of mind construe others as being ideally rational. Dennett still has not provided an adequate argument for why it is that our evolved capacities for theories of mind rely on the *ideal* intentional stance rather than folk psychology.

I think Dennett's final response to the problem of irrationality is the most interesting for the fate of instrumental rationalism. Under pressure from Stich's critiques, Dennett concedes that identifying "rationality with *logical consistency and deductive closure* (and other dictates of the formal normative systems such as game theory and the calculus of probability) I am embarrassed by absurdities."[84] Rather, he suggests that the a priori rules of logic, decision theory, and game theory serve as "the final *benchmark* of rationality" that the theorist invokes them "in the course of criticizing" and "reformulating" accounts of rational "strategies, designs, interpretations."[85]

What is going on with Dennett's shifting account of rationality? Recall that instrumental rationalism is committed to the claim that – whatever account of rationality we adopt – it is such that it makes for rationalizing interpretation that reliably and accurately predict someone's future behavior. It seems that Dennett decided to revise his theory of rationality in order to improve its predictive power. And, if the intentional stance's theory of rationality is revised in a way that matches or bests the predictions put forth by folk psychology, then Stich's objection – namely, that folk psychology is more predictive than the intentional stance – no longer holds.

Dennett's move to adopt different notions of rationality seems to suggest a more general strategy that instrumental rationalists might use in preserving the claim that describing others *as-if* they are rational agents is instrumental towards predicting future behavior: if it turns out that a theory of rationality does not provide for predictive,

---

[83] Robert H. Frank, *Passions within Reason: The Strategic Role of the Emotions* (New York, NY: W. W. Norton & Company, 1988). Thanks to Railton for this point.
[84] Dennett, "Making Sense of Ourselves," 192.
[85] However, in the same breath, Dennett, in his slippery way, wants to say that we can also reject these normative standards as being wrong. I will ignore this for now. Ibid., 193.

rationalizing interpretation, then modify it until it is predictive. It might be best to understand instrumental rationalists as committing themselves to a theory of rationality *conditional upon* its ability to provide for predictive, rationalizing interpretation.

Dennett wants to be able to maintain this kind of flexibility: he ultimately wants to adopt "the *flexible line*" which "insists" that rationality "is not what it appears to be to some theorists – so the idealization will require some "fiddling."[86] The ability to revise one's theory of rationality for the sake of improved predictive power is one of the features of the *flexible intentional stance* that is supposed to make it akin to a "scientific" theory.

Dennett would like for "rationality" to be both normative and a powerful predictive tool. However, I don't think he can have it both ways. Dennett wants to conceive of the term "rational"

> as a general-purpose term of cognitive approval – which requires maintaining only conditional and revisable allegiances between rationality, so considered, and the proposed (or even universally acclaimed) methods of getting ahead, cognitively, in the world. I take this usage of the term to be quite standard, and I take *appeals* to rationality by proponents of cognitive disciplines or practices to require this understanding of the notion."[87]

However, we are left wondering why we should think that predictive theories have normative authority: just because we tend to believe or behave in certain ways does not mean we endorse those patterns of believing and behaving. From a normative point of view, the flexible line suggests a descriptive account of rationality that is not necessarily of normative interest. I will say more about this in my examination of Cohen's account of interpretation.

1.33 Explaining Rational Patterns of Belief and Behavior

In his normative mode, Dennett claims that a theory of rationality ought to be formulated in light of broader considerations having to do with an agent's interests, resources, and goals. For example, he wants to agree with Simon's account of bounded rationality that "*it is rational* in many instances to *satisfice* – e.g., to leap to possibly

---

[86] Ibid., 192.
[87] Ibid., 195.

"invalid" conclusions when the costs of further calculation probably outweigh the costs of getting the wrong answer."[88] I think Dennett is right to seek an account of rationality that takes into consideration specific contexts and goals of reasoning. And, as I will argue in chapter 4, I think he is right that there is an intimate connection to be found in some kinds of predictive psychological theorizing and our canons of justification or rationality. I will argue that naturalized epistemologists and some (not all) research programs in psychology share explanatory goals: in particular, they seek to explain why people make particular judgments and seek to explain why those judgments are "justified." However, just because some predictive psychological theories mesh with our theories of justification or rationality does not support the stronger generalization that *all* predictive psychological theories do.

Dennett suggests that we use our theories of rational belief, desire, and action as tools to predict future behavior. When pressured to predict and account for cases of irrational belief and choice, Dennett modifies his account of rationality. This move suggests a general strategy instrumental rationalism employs to preserve the claim that describing others *as-if* they are rational agents is instrumental towards predicting future behavior: if it turns out that a theory of rationality does not provide for predictive, rationalizing interpretation, then modify it until it is predictive. Dennett does seem to point to an intimate connection that predictive psychological theorizing and our theories of justification or rationality can *sometimes* have, as I will discuss in chapter 4. However, instrumental rationalism's ad hoc method of arriving at a theory of rationality does not necessarily lead to a theory of rationality we would recognize as normative: in the next section, I will argue that just because we tend to believe and behave in certain ways does not necessarily imply that we condone such patterns of belief or behavior.

1.4 Rationality Descriptively Determined

[88] Ibid., 194. Note that Dennett actually describes a case of maximizing expected utility, and not a case of true satisficing. In a true case of satisficing, one does not have any basis for believing that the costs of further calculation outweigh the costs of getting the wrong answer – one simply withholds judgment on this matter. Thanks to Anderson for this comment. See Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge, U.K.: Cambridge University Press, 1983).

Cohen adopts an account of rationality that systematizes the "untutored intuitions" of most "ordinary people." This account of rationality implies that, as an empirical fact, most people draw rational inferences. On the one hand, Cohen looks to the competence/performance distinction in linguistics to support this account. On the other hand, he looks to Nelson Goodman's account of reflective equilibrium. I will argue that both arguments are flawed in important ways. However, the ultimate problem with his account of rationality is that is canonizes patterns of reasoning that have no normative merit: just because most people draw certain kinds of inferences does not mean that those inferences have normative authority. What we might preserve from Cohen's account is an account of "rationalizing" interpretation involving social norms: if we understand Cohen's account as an account of social norms, where social norms are defined in terms of the actual practice of most individuals in a society, then we can use these norms to make light of people's actions in ways that render them intelligible.

1.41 L. J. Cohen's Cognitive Competence

Cohen argues that our norms of rationality should be defined as whatever norms systematize the inferences drawn by "the untutored intuition" of "ordinary people."[89] By designing our canons of rationality to correspond "point by point" with the inductive practices of most people, Cohen's account is committed "to the acceptance of human rationality as a [general] matter of fact."[90] In his account, to say that people's judgments are rational is somewhat tautologous: when we say that most human judgments are *rational*, all we are saying is that most people have the ability to reason in accordance with rules that describe the deductive and inductive practices of most people.

Cohen draws support for his descriptively derived account of rationality from two unrelated sources. On the one hand, Cohen draws support for his account by borrowing the competence-performance distinction in linguistics. In linguistics, linguistic knowledge is constituted by the abstract, unconscious, and complex rules underlying our

[89] L. Jonathan Cohen, "Can Human Irrationality Be Experimentally Demonstrated?," *The Behavioral and Brain Sciences* 4 (1981): 317-8.
[90] Ibid.: 321.

linguistic abilities.[91] Noam Chomsky argues that such linguistic knowledge is innate. And, it is by studying our linguistic performance (behavior) that we can learn more about our linguistic competence (innate knowledge).

The distinction between linguistic performance and competence is important in accounting for grammatical error. On the occasion when native speakers violate linguistic rules, the error is generally explained away most simply as performance error resulting from "grammatically irrelevant conditions" such as "memory limitations, distractions, shifts of attention and interest, and errors" in "applying his knowledge of language in actual performance."[92] Describing random errors as performance errors preserves the assumption of linguistic competence. When systematic mistakes are observed among native speakers, linguists can simply designate that speaker or group of speakers as speaking a different dialect – a move which, again, preserves the assumption that the speaker(s) have an underlying linguistic competence.

Analogously, Cohen imputes to epistemic agents an innate, underlying cognitive competence. He claims that on the occasion that we observe errors in judgment, we should blame these cognitive mishaps, not on deep flaws in human intuition (cognitive competence), but on the suboptimal nature of particular contexts for cognition (cognitive performance). Following Chomsky, Cohen suggests interfering psychological factors in cognitive performance include limitations in memory, disabilities, and motivational factors, and insufficient education to encourage "the maturation of innate ability."[93] Cohen claims we can expect regular lapses in cognitive performance since conditions are rarely ideal for the exercise of our underlying cognitive competence.[94]

Unfortunately, Cohen's analogy between linguistics and cognition is not sufficiently close to legitimize his conception of cognitive competence and performance. Paul Thagard and Richard Nisbett point out many of the important ways in which linguistics and cognition differ from one another. Unlike the case of syntactical structures, we do not have evidence to suggest that inferential competence is innate. One

---

[91] Stein, *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*, 40.
[92] Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, MA: The MIT Press, 1965), 3.
[93] Cohen, "Can Human Irrationality Be Experimentally Demonstrated?," 322. To be consistent with his claims about the epistemic authority of untutored intuition, he specifies "education (that is, education in subjects other than logic and probability theory)."
[94] Ibid.: 321-2.

important reason for thinking linguistic competence is innate is that children can recognize, understand, and produce novel sentences in a language without having had sufficient exposure to the linguistic rules to have learned them from experience. However, there does not seem to be an analogous phenomenon in human cognition. Our inferences do not seem to rely on a set of abstract, unconscious, complex, innate rules of inference: this is illustrated by the great diversity of ways in which we reason about problems of inference.[95]

The most striking disanalogy between linguistics and cognition is that we have witnessed innovations and drastic improvement in standards of inductive reasoning – such as the development of probability and statistical theory – while we have not witnessed commensurate innovation or improvements in grammar.[96] This is an important point: the very idea that our principles of reasoning can *improve* suggests that we take our previous patterns of inference to be inadequate, illegitimate, invalid, or incorrect. This negative evaluation of our cognitive practices does not presume that we are blessed with an innate cognitive competence. Rather, it suggests we rely on developments in normative theory to articulate and diagnose the ways in which we are cognitively flawed; and, we promote education as a means to *improving* our ability to reason and draw inferences. Cohen cannot justify his claim to general cognitive competence with this analogy.

Perhaps this final disanalogy between linguistics and cognition would not trouble Cohen. Because his account of rationality defines our canons of rationality in terms of the untutored intuitions of most individuals, most people – by default – get to count as having a cognitive competence meeting this descriptively determined standard of rationality. If specialists discover new standards of inductive inference, and it turns out that most individuals fail to draw inferences in conformance to that new standard, this need not undermine Cohen's claim that individuals are cognitively competent with respect to his theory of rationality. Indeed, Cohen critiques the heuristics and biases studies for holding subjects' judgments to standards of reasoning that he does not take to constitute the proper norms of rationality. Unfortunately, the norms of rationality he

[95] Thagard and Nisbett, "Rationality and Charity," 259. Richard E. Nisbett et al., "The Use of Statistical Heuristics in Everyday Inductive Reasoning," *Psychological Review* 90, no. 4 (1983).
[96] Thagard and Nisbett, "Rationality and Charity," 259.

proposes – the norms with respect to which we are said to be "cognitively competent" – violate our basic intuitions about the nature of evidential support. I will say more about this shortly.

1.42 Cohen's Reflective Equilibrium

In drawing support for his account of cognitive competence, Cohen turns to Nelson Goodman's solution to Hume's problem of induction. Hume observes that the only way we can justify our rules of induction, is with reference to those very same rules. Goodman embraces this circularity as a "virtuous" one, since "the problem of justifying induction" is nothing "above and beyond" the problem of making mutual adjustments between inductive rules and accepted inferences. "[I]n the agreement achieved lies the only justification needed for either."[97] From Goodman's account of reflective equilibrium, Cohen jumps to the conclusion that the inferences most individuals make under ideal conditions of reasoning count as instances of accepted inductive practices and inferences.

However, Cohen's conclusion is misguided. According to Goodman, "*[a] rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.*"[98] The *acceptability* of inductive practices and inferences is a normative notion. For Goodman, it is true that in most cases, "inductive practice informs definition of valid induction, which in turn guides inductive practice." However, Goodman warns that "we *can* sometimes deny that common inferences count as instances of "valid induction.""[99] That most people have certain intuitions about inductive reasoning does not necessarily mean that we condone or ascribe normative authority to those intuitions.

From a psychological point of view, it is precisely those cases in which we deny the validity of common inferences that psychologists of the heuristics and biases research tradition have been interested to discover and explain. However, Cohen's account has no resources for accounting for the systematic, robust gap between actual inferential practice

---

[97] Nelson Goodman, *Ways of Worldmaking* (Indianapolis, IN: Hackett Publishing Company, 1978), 63-4.
[98] Ibid., 64.
[99] Ibid., 66-7. Italics mine.

and normative theory. By defining our norms of rationality as those that describe human inferential practice, his account has the consequence of obviating the very possibility of discovering that human intuition systematically violates cannons of rationality. Insofar as we are interested in keeping open the possibility of interpreting others so that the majority of subjects are systematically irrational, Cohen's account strikes one as unintuitive – especially in light of the findings by the heuristics and biases research program.

The distinction between actual inference and normative inference is also important from a normative point of view. If we were to define our canons of rationality so as to describe inferential patterns, we would arrive at an account of "rationality" that is not of normative interest. To illustrate, let's consider one of the psychological studies Cohen takes issues with – Kahneman and Tversky's *Linda Problem*:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
>
> Linda is a teacher in elementary school.
> Linda works in a bookstore and takes Yoga classes.
> Linda is active in the feminist movement.
> Linda is a psychiatric social worker.
> Linda is a member of the league of Women Voters.
> Linda is a bank teller.
> Linda is an insurance salesperson.
> Linda is a bank teller and is active in the feminist movement.[100]

According to probability theory, the probability of two events A and B is equal to or less than the probability of each of its conjuncts: $p(A \,\&\, B) \leq p(A)$ and $p(A \,\&\, B) \leq p(B)$. If we identify Linda's being active in the feminist movement as event A, and Linda's being a bank teller as event B, subjects should rank the probability that Linda's both a bank teller and active in the feminist movement ($A \,\&\, B$) as equal or lower than the ranking for A or B considered alone. They found that the vast majority of statistically naïve *and* statistically sophisticated subjects rated the conjunction of events as more probable than the conjunct, in violation of the conjunction rule.

---

[100] Amos Tversky and Daniel Kahneman, "Judgments of and by Representativeness," in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (New York, NY: Cambridge University Press, 1982).

In critiquing the heuristics and biases conclusion that subjects systematically violate the conjunction rule in this kind of context, Cohen suggests that we interpret subject responses as conforming to a Baconian theory of probability. Francis Bacon's *Novum Organum* sought to formulate a procedure to ascertain the empirical basis for inductive generalizations: in particular, he was interested in the procedure involved in establishing causal claims about the necessary and sufficient cause for an event type.[101] Cohen formulated a Baconian theory of probability that tries to capture some of Bacon's key ideas.[102] The Baconian theory of probability formulated by Cohen violates commonly accepted axioms of probability defined by Pascal and others. In particular, Cohen's Baconian account of probability does not require that the conjunction of two events has to be equal to or less than the probability of either its conjuncts.[103]

Although Cohen's Baconian account of probability would allow us to describe subjects' judgments as being "rational" insofar as they conform to a Baconian account of probability, it turns out that Cohen's account of probability runs counter to our basic notions of evidential weight. Cohen's account redefines the conditional probability $p(A \mid B)$ to equal $p(\text{not-}B \mid \text{not-}A)$. This axiom implies that the inductive probability that "a bird which has just been sighted is white if it is a raven must be equal to the probability that the bird is not a raven if it is not white" – a clearly untenable result.[104] In addition, Cohen's account cannot assign a non-zero probability to more than one member of a set of mutually exclusive hypotheses: so, in a murder investigation in which there are several suspects and the murderer is known to have acted alone, the probability of guilt can be

---

[101] Francis Bacon, *The New Organon and Related Writings*, ed. Fulton H. Anderson (New York, N.Y.: The Liberal Arts Press, 1960).

[102] Cohen writes: "The theory has four key ideas. (i) The traditionally distinct methods of agreement and difference are generalized into a single 'method of relevant variables' for grading the inductive reliability of generalizations about natural phenomena in any domain that is assumed to obey causal laws. (ii) The (Baconian) probability of an *A*'s being a *B* is identified with the inductive reliability of the generalization that all *A*'s are *B*'s. (iii) Judgments of Baconian probability are seen to constrain one another in accordance with principles that are derivable within a certain modal-logical axiom-system but not within the classical calculus of chance. (iv) Baconian probability-functions are seen to deserve a place alongside Pascalian ones in any comprehensive theory of non-demonstrative inference, since Pascalian functions grade probabilification *on the assumption that* all relevant facts are specified in the evidence, while Baconian ones grade it *by the extent to which* all relevant facts are specified in the evidence." L. Jonathan Cohen, "On the Psychology of Prediction: Whose Is the Fallacy?," *Cognition* 7 (1979): 389.

[103] Ibid.: 391.

[104] Daniel Kahneman and Amos Tversky, "On the Interpretation of Intuitive Probability: A Reply to Jonathan Cohen," *Cognition* 7 (1979): 409.

non-zero for only one suspect – another counterintuitive result.[105] Because the Baconian account of probability violates our basic notions of evidential weight, it does not meet a normative standard we easily condone. Cohen observes that "[n]ormal Baconian probabilities are not merely not equivalent to Pascalian ones, but are not even any kind of function of the latter."[106] If human judgment really does conform to Baconian rules of probability, then human judgment is systematically irrational in the sense that it violates intuitions of evidential support that we endorse.

1.43 Cohen and Cultural Norms

Cohen's interpretation of reflective equilibrium leads to an unacceptable account of epistemic norms. However, I think he points to a pattern of theorizing familiar to certain kinds of interpretation. In particular, we might think that some cultural norms – such as norms of politeness, authority, or communication – should be defined with reference to general cultural practice. Once systematized, these cultural norms can make behavior intelligible by providing a glimpse into the patterns of intentional and motivational processes lying behind observed behavior. Interpretations guided by an understanding of cultural norms need only be minimally rationalizing since people may have long forgotten the original point of the custom, which no longer applies in the current context.[107] Notice that interpreting others in terms of social behavior and norms allows interpreters to learn new concepts and systems of meaning which enable them to create a new discourse capable of putting the interpreter's and the interpreted's meanings/interpretations of an event in clear contrast. Davidson and Dennett, in contrast, provide more ethnocentric methods of interpretation which presuppose the interpreter's norms of rationality in the interpretive perspective.[108]

---

[105] Ibid.
[106] Cohen, "On the Psychology of Prediction: Whose Is the Fallacy?," 392.
[107] Thanks to Anderson for pointing this out.
[108] See Charles Taylor, "Interpretation and the Sciences of Man," in *Philosophy and the Human Scienecs* (Cambridge, UK: Cambridge University Press, 1985). Charles Taylor, "Understanding and Ethnocentricity," in *Philosophy and the Human Sciences* (Cambridge, UK: Cambridge University Press, 1985).

By defining norms of rationality so as to systematize the intuitions and inferences of most people, Cohen guarantees that the inductive practices of most people are rational. Cohen's theory of rationality and interpretation fails for the same reason Dennett's flexible line fails: just because people generally draw certain kinds of inferences does not mean that we condone those patterns of inference. However, the fact that most people behave in particular ways can sometimes be sufficient for systematizing those behaviors into cultural or social norms. These cultural and social norms can help to render that behavior intelligible by providing some account of the intentions and motivations of the individuals acting in accordance with those norms. I will explore how conversational norms may be understood as social or cultural norms in chapter 3.

1.5 Conclusion

In philosophical accounts of interpretation and rationality, methodological rationalism has followed from the methodological reliance on rationality in various ways. Weber's methodological rationalism is a strategy of interpretation that uses the preferred rational course of action as a model to simplify explanations for irrational actions. Davidson's methodological rationalism results from using rationality as a framework within which interpretations are built. Dennett and Cohen both seek to define norms of rationality to systematize actual inferential practice. Dennett uses rationality as a kind of predictive tool; while, Cohen uses norms of rationality to systematize the intuitions and inferences of most people.

Weber's and Davidson's accounts of methodological rationalism do not successfully cope with the challenge posed by the heuristics and biases literature on human judgment and choice. Weber's strategy of interpreting irrational action seems to do no work in guiding interpretation in the heuristics and biases research in which the irrationality of human judgment is not best understood as a deviation from an otherwise rational course. And, Davidson's use of rationality as a kind of framework does not allow for the possibility of empirically testing whether human judgment and choice systematically violate particular a priori rules of rationality.

Dennett's and Cohen's accounts of methodological rationalism ultimately fail because, in order for a principle to count as a norm of rationality, it is not sufficient to show that most people draw inferences in accordance with it – in the past, present, or future. In addition, both accounts do not have the resources for interpreting a psychologically significant class of cases of interpretation: cases in which the inferences and intuitions of most individuals is best understood as being irrational even under the best of circumstances.

I think it is telling that in contemporary psychological research, the interest with respect to rationality is not primarily an interest in traditional methodological rationalism. Traditional methodological rationalism provides a default methodology for the *process of studying human judgment*. In contrast, contemporary psychological research seems to make rational judgment the *object of study* itself. In the next chapter, I argue that applied cognitive psychology aims to discover the conditions promoting rational rather than irrational judgment. I also argue that this shift in disciplinary aim is motivated by a social and moral interest in promoting cognitive health.

In chapter 3 I discuss psychological research which takes to heart the Davidsonian lesson that all psychological theories on human judgment should include a theory of interpretation about subjects' beliefs about the experimental task. I argue that psychological research on conversational pragmatics import important evidential and methodological standards for psychological experiments on human judgment. And, I argue for using naturalized norms of conversation which are defined with reference to general cultural practice.

In chapter 4 I take up Dennett's intuition that there is an intimate connection between predictive psychological theories and our theories of justification or rationality. I will argue that some of the methodological standards and explanatory goals of psychological theories resemble the explanatory goals naturalized epistemologists adopt in formulating theories of justification. I also begin to explore the ramifications this might have on some classic problems facing reliabilist theories of justification.

Chapter 2

Ecological Rationalism

Over the years, cognitive psychologists have changed their methodological focus with respect to human judgment and choice. In the 1950's and 1960's, the disciplinary tendency in research on decision making had been to see "man as an intuitive statistician,"[109] as a sampling of prominent research demonstrates: Ward Edwards, the founder of judgment and decision making research, theorized that the mind is a reasonably good (though conservative) Bayesian statistician;[110] Wilson Tanner and John Swets introduced the theory of signal detectability (TSD) for psychophysical judgments, which described the mind's detection of a stimulus (such as an auditory tone or light signal against a "noisy" background) as an inference following the Neyman-Pearson technique of hypothesis testing;[111] and, Jean Piaget and Barbel Inhelder took the formal laws of probability to be the laws of the adolescent and adult mind.[112]

However, Daniel Kahneman and Amos Tversky's paper "Judgment Under Uncertainty: Heuristics and Biases" published in a 1974 issue of *Science* successfully changed psychology's primary disciplinary interest. This paper summarized an

---

[109] C. R. Peterson and Lee Roy Beach, "Man as an Intuitive Statistician," *Psychological Bulletin* 68 (1967). This view fits the psychological literature on decision making in particular. Social psychological research in the 1950's and 1960's witnessed other, competing perspectives focused on cognitive consistency, wishful thinking, group dynamics, and social comparison processes. For these areas of research, see Shelley E. Taylor, "The Social Being in Social Psychology," in *The Handbook of Social Psychology*, ed. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey (Boston, MA: The McGraw Hill Companies, Inc., 1998). Thanks to Norbert Schwarz for this comment.

[110] Ward Edwards, "Nonconservative Information Processing Systems," (Ann Arbor, MI: University of Michigan, Institute of Science and Technology, 1966).

[111] Wilson P. Tanner and John A. Swets, "A Decision-Making Theory of Visual Detection," *Psychological Review* 61, no. 6 (1954).

[112] Jean Piaget and Barbel Inhelder, *The Origin of the Idea of Chance in Children* (New York, NY: Norton, 1951; reprint, 1975).

ambitious scope of studies demonstrating the systematic irrationality of human judgment. The article discussed twelve biases: insensitivity to prior probabilities, the effect of arbitrary anchors on estimates of quantities, availability biases in judgment of frequency, illusory correlation, nonregressive prediction, and misconceptions of randomness. The "accumulation of demonstrations in which intelligent people violate elementary rules of logic or statistics" raised serious doubts about "the descriptive adequacy of rational models of judgment and decision making."[113]

Kahneman and Tversky's research brought with it a disciplinary shift away from describing human judgment and decision making as conforming to standards of rationality and/or correctness. Their self-avowed "methodological focus on errors and the role of judgment biases" became an institutional norm.[114] In the decade that followed, articles reporting good and poor performance were published in comparable numbers; however, psychologists became disproportionately interested in experimental tasks demonstrating poor participant performance.[115] Studies reporting poor subject performance were cited an average of 27.8 times while studies reporting good subject performance were cited an average 4.7 times: a 6:1 ratio.[116] The disciplinary focus on irrational judgments extended to judgments traditionally studied by other social scientific domains: researchers provided work demonstrating systematically irrational judgments and choices in medical diagnosis,[117] law,[118] economics,[119] management science,[120] and political science.[121]

---

[113] Daniel Kahneman and Amos Tversky, "On the Study of Statistical Intuitions," in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (New York, NY: Cambridge University Press, 1982), 494.

[114] Daniel Kahneman and Amos Tversky, "On the Reality of Cognitive Illusions," *Psychological Review* 103, no. 3 (1996): 582.

[115] L. Lola Lopes, "The Rhetoric of Irrationality," *Theory and Psychology* 1, no. 1 (1991).

[116] Jay J. J. Christensen-Szalanski and Lee Roy Beach, "The Citation Bias: Fad and Fashion in the Judgment and Decision Literature," *American Psychologist* 39 (1984).

[117] {Casscells, 1978 #513;Eddy, 1982 #557;Elstein, 1990 #579}

[118] Michael J. Saks and Robert F. Kidd, "Human Information Processing and Adjudication: Trial by Heuristics," *Law and Society Review* 15, no. 1 (1980). C. R. Sunstein, "Behavioral Analysis of Law," *University of Chicago Law Review* 64 (1997), William C. Thompson and Edward L. Schumann, "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy," *Law and Human Behavior* 11 (1987). R. B. Korobkin and T. S. Ulen, "Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics," *California LAw Review* 88 (2000).

[119] Kahneman and Tversky, "Prospect Theory: An Analysis of Decision under Risk.", Tversky and Kahneman, "The Framing of Decisions and the Psychology of Choice.", Tversky and Kahneman, "Rational Choice and the Framing of Decisions."

The disciplinary tide is once again turning. In the 1990's, prominent psychological research has focused on the different ways in which human judgment can be said to be rational. Evolutionary psychologists have called for a return to seeing humans as good intuitive statisticians.[122] Gerd Gigerenzer in particular has been especially vociferous in arguing for a shift in the disciplinary focus towards rational cognitive processes:

> Our goal is to. . . shift the focus from human errors to human engineering. . . [and] help people reason the Bayesian way without even teaching them.[123]
>
> [W]e propose a new theoretical model for confidence in knowledge based on the more *charitable* assumption that people are good judges of the reliability of their knowledge, provided that the knowledge is representatively sampled from a specified reference class.[124]
>
> [A]fter 40 years of toying with the notion of bounded rationality, it is time to overcome the opposition between the rational and the psychological and to reunite the two."[125]

There has also been a flourishing of research identifying the kinds of educational innovations, inferential tools, and technologies that can improve the considered judgment of Joe Q. Public as well as socially recognized "experts" such as physicians, lawyers, and judges.[126] Important methodological critiques and innovations that have arisen from this turn towards focusing on tasks in which human judgment and decision making can be said to be rational – I will say more some of these in the following chapters.

In this chapter, I want to focus on an important change in aim for research on human judgment and choice. *Ecological rationalism* reconceives the preference for rational interpretations as a preference for *discovering conditions* that promote rational

---

[120] M. H. Bazerman, *Judgment in Managerial Decision Making* (New York, NY: Wiley, 1990).
[121] P. M. Sniderman, R. A. Brody, and Philip E. Tetlock, *Reasoning and Choice: Explorations in Political Psychology* (New York, NY: Cambridge University Press, 1991).
[122] Leda Cosmides and John Tooby, "Are Humans Good Intuitive Statisticians after All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty," *Cognition* 58 (1996). Gerd Gigerenzer, "From Tools to Theories: A Heuristic Discovery in Cognitive Psychology," *Psychological Review* 98, no. 2 (1991).
[123] Gerd Gigerenzer and Ulrich Hoffrage, "How to Improve Bayesian Reasoning without Instruction: Frequency Formats," *Psychological Review* 102, no. 4 (1995): 685.
[124] Gerd Gigerenzer, Ulrich Hoffrage, and Heinz Kleinbolting, "Probabilistic Mental Models: A Brunswikian Theory of Confidence," *Psychological Review* 98, no. 4 (1991): 506. Italics mine.
[125] Gerd Gigerenzer and Daniel G. Goldstein, "Reasoning the Fast and Frugal Way: Models of Bounded Rationality," *Psychological Review* 103, no. 4 (1996): 666.
[126] Ward Edwards, Harold Lindman, and Lawrence D. Phillips, "Emerging Technologies for Making Decisions," in *New Directions in Psychology 2* (New York: Holt, Rinehard and Winston, Inc., 1965), Herbert A. Simon, *The Sciences of the Artificial*, 3 ed. (Cambridge, MA: The MIT Press, 1996).

judgment. Ecological rationalism recognizes rational judgment as a *focus* of research. Unlike discoveries about the conditions of irrational judgment, discoveries about the conditions promoting rational judgment suggest positive recommendations about how we might reform environments and social institutions to promote better reasoning. And, our political and moral interests in promoting what I will call *cognitive health* justifies ecological rationalism. I will foreground my account of ecological rationalism in the first section of this chapter.

In what follows, I will trace the development of ecological rationalism in psychological research. After the heuristics and biases research program's burgeoning during the 1970's and early 80's, psychologists began to critique the overgeneralizations researchers drew from Kahneman and Tversky's studies about judgment biases and human irrationality. In response, researchers actively sought to limit the scope of Kahneman and Tversky's claims about human irrationality by modifying the original experimental tasks to decrease or eliminate judgment biases. This responsive research served three purposes. By actively seeking to identify the experimental conditions promoting rational judgment, these researchers mobilized a disciplinary return to studying rational judgment. In addition, this research underscored the methodological point that experimental evidence can only properly support claims about the *particular* ways in which we are rational or irrational in *specific* contexts of reasoning. I will call this point the *lesson of context-specificity*.

Research seeking to identify the conditions promoting rational judgment began to demonstrate the practical implications of discovering the conditions promoting rational rather than irrational judgment: such research provided better grounds for recommendations about how to change contexts, educational strategies, and institutions to improve human judgment. Psychologists sought to redirect research agendas and questions so as to improve cognitive health: prominent researchers explicitly argued that moral and political interests should determine what kinds of judgments, tasks, and subjects should constitute significant areas of research. To close, I will note key lines of research in ecological rationalism.

2.1 Ecological Rationalism

2.11 Methodological Rationalism versus Ecological Rationalism

Recall that methodological rationalism claims that we should seek to interpret the beliefs and/or choices of others as rational: other things being equal, we should impute desires, beliefs, and other mental states so their observed behavior is rational in relation to those mental states, and so their mental states are rational in relation to each other. If the evidence does not support a rationalizing interpretation, methodological rationalism allows explanations construing others as irrational. However, methodological rationalism prefers rationalizing intentional explanations over those that construe others as irrational.

Traditionally, methodological rationalism is understood as a default method for interpretation. Weber used rationalizing interpretation as a way to simplify the interpretation and explanation of irrational action. Davidson used the rationality of an interpretation as a framework within which attributions of beliefs, meanings, and desires were built. Dennett suggests using rationality as a kind of predictive tool in interpreting others. And, Cohen suggests that, because our theories of rationality systematize most people's untutored intuitions, most people's inferences are rational. These theories are cases of *methodological* rationalism insofar as they condone the interpretive method of assuming that subjects are rational until proven otherwise (for simplifying, transcendental, or instrumental reasons).[127]

Methodological rationalism provides a default methodology for the *process of studying human judgment.* In contrast, ecological rationalism makes rational judgment the *object of study* itself.[128] Ecological rationalism, in its interest in seeking rationalizing interpretation, reconceives the methodological preference as a preference for discovering conditions that promote rational judgment and marks a changed aim in research on human judgment and choice. This change in focus need not imply methodological rationalism: that is, it does not imply that we should begin inquiry under the defeasible assumption that others are generally or mostly rational. In the next chapter, I will say

[127] Thanks to Peter Railton for clarifying this.
[128] Thanks to Railton for this insight and way of stating the distinction between methodological rationalism and ecological rationalism.

more about the conditions in which we should adopt methodological rationalism and prefer charitable interpretations in interpreting subject responses in experimental contexts. For now, I will continue to develop my account of ecological rationalism.

2.12 Contextual Values and Determinations of Significance: Cognitive Health

Why should we care to discover conditions that promote rational judgment rather than those underlying any other kind of judgment? What do we gain by talking about rationalizing interpretations *qua* rationalizing interpretations, rather than *qua* empirically supported theory?[129] What does contemporary rationalism add when the empirical evidence is sufficient on its own for legitimate psychological theorizing? Ecological rationalism is justified in relation to our moral and political interest in promoting cognitive health. Just as the moral valuation of human health gives rise to the distinction between health and disease in medical research, psychology's interest in promoting cognitive health animates the distinction between rational and irrational judgment. And, just as medical research aims to identify the causes and successful interventions for preventing, managing, and curing disease, cognitive psychology should seek to identify the conditions of rational and irrational judgment for the purposes of identifying the causes and successful interventions for preventing and managing irrational judgment, where these interventions include education and institutional restructuring. The preference for discovering conditions that promote rational judgment (rather than irrational judgment) is justified because they provide better grounds for making positive recommendations about how to implement successful intervention strategies.

The political and moral interest in promoting cognitive health serves as a *contextual value* insofar as it prefers hypotheses, questions, explanations, domains of evidence, and classification schemes that speak to the interest of improving human reasoning. Contextual values are characteristically contrasted with constitutive values:

[129] David Henderson, who has written on the connection between charitable interpretation and explanation, concludes that there is no way to cash out the preference for rationalizing interpretation that contributes towards psychological theorizing. David K. Henderson, *Interpretation and Explanation in the Human Sciences* (Albany, NY: State University of New York Press, 1993).

they include political, moral, and other values taken from the social context in which the sciences are practiced, and are not traditionally considered reliable indicators of truth.[130]

Feminist epistemologists have pointed out that contextual beliefs, values, and commitments often serve to identify what gets to count as significant concepts, explanations, questions, and phenomena. Consider accuracy – a relatively uncontested epistemic value. *Accuracy* prefers theories that can, within their domain, deduce consequences that are in demonstrated agreement with the results of previous, existing, and future observations of significant phenomena.[131] Accuracy is characteristically the most decisive of all the criteria because scientists are particularly unwilling to give up the predictive and explanatory powers that depend on it.[132] *Empirical adequacy*, often synonymous with accuracy, values theories that can account for all the relevant evidence.

Feminist epistemologists have pointed out that definitions and characterizations of accuracy and empirical adequacy are not sufficient for constraining how these constitutive values are used in theory choice. For example, a theory is said to be more or less more accurate with respect to a class of phenomena. However, an antecedent decision must be made about what the comparison class ought to be – a decision about which phenomena are the important phenomena requiring accurate explanation and prediction – as a precondition to making comparative judgments of the accuracy of competing theories. Two scientists both agreed on accuracy's trumping value can disagree over what the significant comparison classes are and ultimately arrive at different theory valuations. Analogously, two scientists both agreed on empirical adequacy's trumping value can disagree over what the significant comparison classes are and, again, arrive at conflicting valuations. It is here that contextual values have a covert role: contextual values serve to identify comparison class membership, and prefer some questions, phenomena, evidence, and explanations over others.[133]

In some domains of scientific research – especially in the social and human sciences – the role of contextual values is explicit and considered salubrious. For

---

[130] Elizabeth Anderson, "Knowledge, Human Interests, and Objectivity in Feminist Epistemology," *Philosophical Topics* 23, no. 2 (1995): 28, Helen E. Longino, *Science as Social Knowledge* (Princeton, N.J.: Princeton University Press, 1990).
[131] Thomas S. Kuhn, "Objectivity, Value Judgment, and Theory Choice," in *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), 321.
[132] Ibid., 323.
[133] Helen E. Longino, "Gender, Politics, and the Theoretical Virtues," *Synthese* 104, no. 3 (1995): 396.

example, Elizabeth Anderson has pointed out that medical research's moral valuation of human health gives rise to distinctions between health and disease and motivates research on the causes of and successful interventions in preventing, managing, and curing disease.[134] In *clinical research*, the phenomena requiring accurate explanation and prediction are health and disease; and, medical research providing competing hypotheses about the cause or best intervention for disease will have to be accurate and empirically adequate in these respects. For clinical research, it is the contextual value of improving and promoting human health that makes the inquiries of epidemiologists, doctors, and health researchers distinctly "medical" in its normative sense. In contrast, *basic research* in medicine is not connected to the production of health outcomes or the discovery of therapeutic interventions.[135]

A similar distinction can be made in psychological research. In what I will call *applied cognitive psychology*, the phenomena requiring accurate explanation and prediction are rational and irrational judgment; and, psychological research providing competing hypotheses about the cause or best intervention for irrational judgment will have to be accurate and empirically adequate in these respects. The contextual value of promoting cognitive health identifies comparison class membership, and prefers therapeutically relevant questions, phenomena, evidence, and explanations.[136] In contrast, basic research in cognitive psychology need not connect with the interest of promoting cognitive health or the discovery of successful interventions for irrational judgment. However, as I will argue in chapter 4, even basic research in cognitive psychology seems to have a special interest in the distinction between rational and irrational judgment.

If the contextual determination of significance is a *legitimate* criterion of theory choice for applied cognitive psychology, then it follows that the interest in cognitive health plays a legitimate role in putting a premium on psychological theories that capture what we take to be politically and morally significant about human judgments and choices. The contextual interest in cognitive health motivates researchers' decisions

---

[134] Anderson, "Knowledge, Human Interests, and Objectivity in Feminist Epistemology." Elizabeth Anderson, "Feminist Epistemology: An Interpretation and Defense," *Hypatia* 10 (1995).
[135] Thanks to Railton for this contrast between clinical and basic research in medicine.
[136] Longino, "Gender, Politics, and the Theoretical Virtues," 396.

about the kind of hypotheses, questions, explanations, domains of evidence, and classification schemes to seek and prefer in applied cognitive psychology. And, it is by reference to cognitive health that therapeutically-minded psychologists can determine the informational value of the claims, explanations, and theories offered in their research.[137]

In the following sections, I will argue that research in contemporary psychology appreciates how contextual values should guide research agendas. I will trace how this appreciation emerged from research critical of the heuristics and biases research program, the lesson of context-specificity, the disciplinary shift towards discovering conditions promoting rational judgment, and the emerging *rationale* for this disciplinary shift. I will also argue that focusing solely on errors does not suggest ways in which we can successfully promote better reasoning. If we are interested in interventions that take the form of education/training or institutional reform, we must learn how to launch *successful* interventions.[138] And, we can evaluate what is more likely to count as a successful intervention if we have evidence about the conditions that promote *rational* judgment.

2.2 The Critique: How Robust are Judgment Biases?

Kahneman and Tversky's heuristics and biases research program did not denounce human reasoning as *universally* fallacious. From a theoretical point of view, their work has always recognized that heuristic-driven judgment is usually rational or valid: they claim that "heuristics are highly economical and *usually* effective."[139] They freely admit to a systematic focus on tasks eliciting irrational judgment. However, they have maintained that the "main goal of this research" is more general and scientific in nature: that of understanding "the cognitive processes that produce both *valid* and invalid judgments."[140] Their recognition that heuristics are sometimes valid and that human judgment is sometimes rational embraces the more cautious, qualified conclusion that

---

[137] Anderson, "Knowledge, Human Interests, and Objectivity in Feminist Epistemology," 41.
[138] Richard E. Nisbett et al., "Teaching Reasoning," in *Rules for Reasoning*, ed. Richard E. Nisbett (Hillsdale, NJ: Larence Erlbaum Associates, 1993).
[139] Amos Tversky and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (1974).Italics mine.
[140] Kahneman and Tversky, "On the Reality of Cognitive Illusions," 582. Italics mine.

human judgment exhibits particular kinds of biases under *some* conditions or contexts of reasoning.

Rhetorically speaking, however, Kahneman and Tversky seemed to encourage their readers to draw much stronger conclusions.[141] They have said things like: "[i]n making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction."[142] This unqualified conclusion suggests the stronger claim that *under no circumstances* do people seem to conform to the rules of probability or statistics. Such unqualified, stronger claims – coupled with a nearly unwavering focus on tasks eliciting irrational judgment – presented human irrationality as a kind of universal, immutable fact, "like gravity."[143] Research in other social scientific fields certainly got this impression as did some psychologists. [144] Kahneman and Tversky did not take pains to disabuse researchers from this impression. As Baruch Fischoff remarked, the "retelling of these results has tended to accentuate the negative" about human judgment "in part because the pioneering studies showed their caution more in claims that were not made than in claims that were denied."[145]

Psychologists were quick to critique the over-generalizations drawn from Kahneman and Tversky's studies. The year after Kahneman, Slovic, and Tversky's canonical book *Judgment Under Uncertainty: Heuristics and Biases* was published, Ward Edwards, the founder of research on human judgment, objected to this genre of research for having failed "to heed the urging of Egon Brunswik (1955) that generalizations from laboratory tasks should consider the degree to which the task (and

---

[141] On the rhetoric on the rationality versus irrational of human judgment, see Richard Samuels, Stephen P. Stich, and Michael Bishop, "Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear," in *Common Sense: Reasoning and Rationlity*, ed. Renee Elio (Oxford: Oxford University Press, 2002).

[142] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 237.

[143] Lopes, "The Rhetoric of Irrationality," 67.

[144] Between 1975 and 1980, Kahneman and Tversky's *Science* article was cited 227 times in 127 different journals. Of these, about 20 percent of the citations were in sources outside of psychology; and, of these, all these used the citation to support the over-generalization that people are poor decision-makers. Diana Berkeley and Patrick Humphreys, "Structuring Decision Problems and the 'Bias Heuristic'," *Acta Psychologica* 50, no. 3 (1982).

[145] Fischoff suggested that psychologists "should monitor the way that those results are used, for cases where the hedges are either trimmed or magnified, either by those who fail to appreciate the niceties of experimental design or by those who choose to ignore them, in order to achieve some rhetorical purpose. Baruch Fischhoff, "Reconstructive Criticism," in *Analysing and Aiding Decision Processes*, ed. Patrick Humphreys, Ola Svenson, and Anna Vari (Amsterdam, Holland: North-Holland Publishing Company, 1983), 521-2.

the person performing it) resemble or represent the context to which the generalization is made."[146] He criticized the heuristics and biases research program for not making explicit the fact that "both their methods and their selection of subjects encourage the occurrence of error."[147] Edwards disagreed so strongly with the overgeneralizations drawn from this research that he felt "ashamed about my own role in starting it off."[148]

Other psychologists agreed with Edwards that the unqualified "rejection of human capability to perform probabilistic tasks is extremely premature."[149] Robin Hogarth pointed out that "the conditions under which such heuristics can be valid have not been specified and that research had only covered a narrow spectrum of judgment and decision behavior.[150] The year after *Judgment Under Uncertainty* was published, Fischoff also questioned the robustness of the heuristics and biases studies and suggested that the "reanalysis of existing studies" should "acknowledge that all faithfully collected and replicated data have some range of validity. The "trick" is to clarify what that range is."[151]

2.3 Discovering Conditions that Promote Rational Judgment

Researchers in psychology began work to focus on identifying the range of validity for judgment biases. Fischoff proposed to do this by identifying the conditions in which judgment biases disappear. He imagined a project of destructive testing – a tool in engineering – where "a proposed design is subjected to conditions intended to push it to and beyond its limits of viability" with the goal of identifying "where it is to be trusted

---

[146] Ward Edwards, "Human Cognitive Capabilities, Representativeness, and Ground Rules for Research," in *Analysing and Aiding Decision Processes*, ed. Patrick Humphreys, Ola Svenson, and Anna Vari (Amsterdam, Holland: North-Holland Publishing Company, 1983), 509. He cites Egon Brunswik, "Symposium on the Probability Approach in Psychology," *Psychological Review* 62, no. 3 (1955).
[147] Edwards, "Human Cognitive Capabilities, Representativeness, and Ground Rules for Research," 508.
[148] Edwards goes on to explain how his frustration exhibited itself in his own research: "I remained silent about it because I believed, wrongly, that it was a fad and would die out – though those of you who have followed my work will note that I published not a word about conservatism in probabilistic inference since about 1970." Ibid.
[149] Ibid., 511.
[150] Robin M. Hogarth, "Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgmental Heuristics," *Psychological Bulletin* 90, no. 2 (1981): 197-8.
[151] Fischhoff, "Reconstructive Criticism," 517.

and why it works when it does."[152] When the phenomenon of interest is a "judgment bias, destructive testing takes the form of debiasing efforts." When we find conditions under which "a bias fails, the result is improved judgment."[153] Such a project suggested the beginnings of a more general disciplinary shift in focus – away from conditions promoting judgment biases – towards conditions promoting rational judgment.[154]

Psychologists began to scout out the robustness of Kahneman and Tversky's findings and the proper scope of Kahneman and Tversky's conclusions about human judgment. Gigerenzer's work on the use of frequencies in probability judgments provides a clear example of this genre of research. He has argued that recognizing the distinction between single-event probabilities and frequencies "unearth[s] the reasonableness hidden by the perspective of the heuristics and biases program" by making "several apparently stable cognitive illusions disappear."[155]

For example, recall the conjunction fallacy. The key experimental task used to establish the conjunction fallacy was the *Linda Problem*:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
>
> Linda is a teacher in elementary school.
> Linda works in a bookstore and takes Yoga classes.
> Linda is active in the feminist movement.
> Linda is a psychiatric social worker.
> Linda is a member of the league of Women Voters.
> Linda is a bank teller.
> Linda is an insurance salesperson.
> Linda is a bank teller and is active in the feminist movement.[156]

---

[152] Baruch Fischhoff, "Debiasing," in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (Cambridge, U.K.: Cambridge University Press, 1982), 423.
[153] Ibid.
[154] Fischoff and others were involved in applied cognitive psychological research aimed at improving medical diagnosis and informed consent. See, for example, Baruch Fischhoff, "Informed Consent in Societal Risk-Benefit Decisions," *Technical Forecasting and Social Change* 13 (1979). Baruch Fischhoff, "Clinical Decision Analysis," *Operations Research* 28 (1980).
[155] Gerd Gigerenzer, "Why the Distinction between Single-Event Probabilities and Frequencies Is Important for Psychology (and Vice Versa)," in *Subjective Probability*, ed. George Wright and Peter Ayton (New York, NY: John Wiley & Sons, 1994), 141-2.
[156] Tversky and Kahneman, "Judgments of and by Representativeness."

Kahneman and Tversky found that the vast majority of statistically naïve *and* statistically sophisticated subjects rated the conjunction of events as more probable than either conjunct, in violation of the conjunction rule.[157] Impressed by their experimental results, Kahneman and Tversky took their research to demonstrate the "massive failure of the conjunction rule" and speculated that the conjunction fallacy must affect the judgments of "political analysts, jurors, judges, and physicians."[158] In the same breath, they admit that their experimental tasks "were constructed to elicit conjunction errors, and they do not provide an unbiased estimate of the prevalence of these errors."[159] Yet, this passage continues, in a less careful manner, to suggest that the conjunction fallacy is "only a symptom of a more general phenomenon: People tend to overestimate the probabilities of representative (or available) events and/or underestimate the probabilities of less representative events."[160]

Later studies demonstrated the limited scope of the conjunction fallacy. Gigerenzer and Hertwig discovered that subjects would conform to the conjunction rule in the Linda problem when the statistical information and questions were restated in terms of frequencies:[161]

> In an opinion poll, the 200 women selected to participate have the following features in common: They are, on average, 30 years old, single, and very bright. They majored in philosophy. As students, they were deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.
>
> Please estimate the frequency of the following events.
>
> How many of the 200 women are bank tellers? ___ of 200
> How many of the 200 women are active feminists? ___ of 200
> How many of the 200 women are bank tellers and active feminists? ___ of 200[162]

Under this condition, subjects did not violate the conjunction rule.

---

[157] According to probability theory, the probability of two independent events A and B is equal to or less than the probability of each of its conjuncts: $p(A \& B) \leq p(A)$ and $p(A \& \ B) \leq p(B)$. Subjects should rank the probability that Linda's both a bank teller and active in the feminist movement as equal to or lower than the ranking each of these conjuncts taken alone. Ibid., 93.

[158] Ibid., 94.

[159] Amos Tversky and Daniel Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review* 90, no. 4 (1983): 311.

[160] Ibid.

[161] Gerd Gigerenzer, "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)," *Psychological Review* 103, no. 3 (1996).

[162] Ralph Hertwig and Gerd Gigerenzer, "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors," *Journal of Behavioral Decision Making* 12 (1999): 291.

Kahneman and Tversky were the first to discover that frequency presentations improve probability judgments.[163] And, this finding confers credibility to their claim that they never assumed that "heuristics are independent of content, task, and representation."[164] However, it was other researchers who connected this discovery with practical concerns about how to *improve* human reasoning *and* with questions about the robustness of judgmental biases. Gigerenzer and Hertwig argued that this discovery served as a counter-example to the over-generalized claim that human judgment cannot conform to the conjunction axiom (or any other probabilistic rule): if the mind did not have a heuristic for making probability judgments in conformance to the conjunction rule, then subject responses should not improve with changes in how the information is represented. Similar frequency effects (and counter-examples) were also discovered for the overconfidence fallacy[165] and the base rate fallacy.[166]

2.4 The Lesson of Context-Specificity

Just as critical researchers suspected, whether subjects exhibit rational or irrational judgment depends crucially upon the experimental conditions and tasks one chooses to study. The research that focused explicitly on identifying conditions promoting rational judgment suggested a more context-sensitive approach to understanding rational and irrational judgment: we now had conclusions about the conditions in which people did and did not exhibit the overconfidence effect,[167] the base

---

[163] Tversky and Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," 308-10.
[164] Kahneman and Tversky, "On the Reality of Cognitive Illusions," 583.
[165] Gerd Gigerenzer, "How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases"," *European Review of Social Psychology* 2 (1991), Gigerenzer, "Why the Distinction between Single-Event Probabilities and Frequencies Is Important for Psychology (and Vice Versa).", Joshua Klayman et al., "Overconfidence: It Depends on How, What, and Whom You Ask," *Organizational Behavior and Human Performance* 79, no. 3 (1999).
[166] Gigerenzer, Hoffrage, and Kleinbolting, "Probabilistic Mental Models: A Brunswikian Theory of Confidence."
[167] Gigerenzer, "How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases".", Gigerenzer, "Why the Distinction between Single-Event Probabilities and Frequencies Is Important for Psychology (and Vice Versa).", Klayman et al., "Overconfidence: It Depends on How, What, and Whom You Ask."

rate effect,[168] and the conjunction fallacy.[169] This context-sensitive approach served to underscore the importance of the lesson of context-specificity: experimental evidence demonstrates the particular ways in which we are rational or irrational in specific contexts of reasoning.

It is important to note that the lesson of context-specificity applies for both sides of the rationality debate. Gigerenzer does not always take sufficient care in his claims about the scope of rational judgment. Kahneman and Tversky rightly catch Gigerenzer at suggesting this kind of overgeneralization in his claims about frequency judgments:

> The major empirical claim in Gigerenzer's critique, that cognitive illusions "disappear" when people assess frequencies rather than subjective probabilities, also rests on a surprisingly selective reading of the evidence. Most of our early work on availability biases was concerned with judgments of frequency, and we illustrated anchoring by inducing errors in judgments of the frequency of African nations in the United Nations. Systematic biases in judgments of frequency have been observed in numerous other studies.[170]

The moral to draw from contemporary research should be that we should make sufficiently qualified claims about the scope and conditions for irrational judgment *and* for rational judgment.

2.5 Contextual Values and the Rationale for Research

Researchers who recognized that the rationality or irrationality of judgment depends crucially upon experimental conditions began to ask important questions about the direction of future research on human judgment. *Which* types of contexts of reasoning should researchers be interested in studying? Kahneman and Tversky had argued for their focus on judgment biases for the broader intellectual goal of gaining an

---

[168] Gigerenzer, Hoffrage, and Kleinbolting, "Probabilistic Mental Models: A Brunswikian Theory of Confidence."

[169] Hertwig and Gigerenzer, "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors."

[170] Kahneman and Tversky, "On the Reality of Cognitive Illusions," 584. They refer to the following papers: Paul Slovic, G. Fischhoff, and S. Lichtenstein, "Facts Versus Fears: Understanding Perceived Risk," in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (Cambridge, U.K.: Cambridge University Press, 1982), Amos Tversky, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* 5 (1973), Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases."

understanding of normal cognitive processes, just as researchers study "illusions to understand the principles of normal perception."[171] At first, Fischoff's proposed change in research agenda towards identifying conditions of rational judgment was justified by same goal: to understand normal cognitive processes.

However, researchers began to criticize the rationale behind Kahneman and Tversky's research. For example, Lawrence Phillips thought it "revealing" that "researchers in this area prefer to focus on the deficiencies, to develop explanations and models to account for these deficiencies, rather than to look for the characteristics of tasks that would enable people with different capacities to do well."[172] As early as 1979, Alan Baddeley noted "a basic change in attitudes away from the ivory-tower view of the 1960s that the pursuit of knowledge – any knowledge – for its own sake was sufficient end in itself." He suggested a trend towards wanting to do research that is "at least potentially useful." He noted that governmental institutions issuing research grants preferred "research yielding practical benefits."[173] Baddeley called psychological research focused on seeking theories that bear on real life problems *applied cognitive psychology*."[174] He observed that a focus on "real-world problems" changes the orientation of theorizing "by drawing attention to interesting and important questions and by ensuring that our theories and concepts do not become too laboratory and paradigm bound."[175]

Like Baddeley, Fischoff and Edwards both suggest that researchers change their *rationale* for psychological research and let such rationales guide the kinds of questions they pursue. Fischoff urged researchers to look to practical concerns in defining their research agendas: in 1983 he urged researchers to "study judgment not just as an intellectual curiosity, or as a key to understanding basic cognitive processes, but also as a guide to action." And, he argued that it was in relation to this interest to help people

---

[171] Kahneman and Tversky, "On the Study of Statistical Intuitions," 493.
[172] Lawrence D. Phillips, "A Theoretical Perspective on Heuristics and Biases in Probabilistic Thinking," in *Analysing and Aiding Decision Processes, Volume 14*, ed. Patrick Humphreys, Ola Svenson, and Anna Vari (Amsterdam: North-Holland Publishing Co., 1983), 533.
[173] Alan Baddeley, "Applied Cognitive and Cognitive Applied Psychology: The Case of Face Recognition," in *Perspectives on Memory Research: Essays in Honor of Uppsala University's 500th Anniversary*, ed. Lars-Goran Nilsson (New York, NY: Lawrence Erlbaum Associates, 1979), 367-8.
[174] Ibid., 369.
[175] Ibid., 368.

make good judgments that motivated the interest in whether human judgment was rational or not: he claimed that researchers' interest in providing practical guidance "requires a global appraisal of "how much do people know?" or "how good is people's judgment?""[176]

Edwards also urged psychologists to adopt the research methods of those "practically-oriented" researchers "who define their roles as being to help others to perform intellectual tasks, notably decision making."[177] He observed that a practical, applied orientation would require researchers to "learn how to get access to the populations to which we wish to generalize" and to identify "the myriad kind of tasks" that "especially deserve our attention."[178] Such research took the lesson of context-sensitivity seriously: one of Edwards's "ground rules" for research was that such research focus on "tasks representative of the kinds of tasks that we wish our generalizations to cover," which required studying "specific classes of minds performing specific kinds of tasks."[179]

2.6 The Preference for Discovering Conditions Promoting Rational Judgment

Psychologists working on judgment under uncertainty adopted an increasingly practical orientation towards research. Even Kahneman and Tversky, for example, suggested that psychologists should focus on systematic errors and inferential biases – not just because doing so improves our understanding of cognitive processes – but because doing so might "suggest ways of improving the quality of our thinking."[180] However, it is research that focused on discovering conditions that promote *rational* rather than irrational judgment that lends itself more easily to making recommendations about how to create conditions that *improve* human judgment. To see this, I will look to the still growing research on the base rate fallacy. Recall Kahneman and Tversky's famous lawyer-engineer question:

---

[176] Fischhoff, "Reconstructive Criticism," 520.
[177] Edwards, "Human Cognitive Capabilities, Representativeness, and Ground Rules for Research," 512.
[178] Ibid.
[179] Ibid., 511-2.
[180] Kahneman and Tversky, "On the Study of Statistical Intuitions," 494.

> A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.
>
> The same task has been performed by a panel of experts, who were highly accurate in assigning probabilities to the various descriptions. You will be paid a bonus to the extent that your estimates come close to those of the expert panel.[181]

The *low-engineer group* was told that there were 30 engineers and 70 lawyers. The *high-engineer group* was told that there were 70 engineers and 30 lawyers. Both groups were provided the same five personality descriptions, most of which were stereotypical of an engineer or lawyer.[182] Kahneman and Tversky found that subjects' predictions about how probable it was that a given person was an engineer or lawyer were independent of the base rates of engineers/lawyers in violation of Bayes' Rule. Kahneman and Tversky took the results of their lawyer/engineer study to demonstrate that "when worthless evidence is given, prior probabilities are ignored."[183]

A few years after Kahneman and Tversky's *Science* article, Casscells questioned the implications the base rate fallacy might have on a kind of probability judgment task with life or death consequences: namely, physicians' abilities to diagnose disease. He published an alarming study demonstrating that even Harvard medical school staff and students – individuals highly trained in diagnosing disease (i.e., estimating the conditional probability that an individual has a disease given her symptoms and the base rate of the disease) – did not incorporate base rate information properly in conformance with Bayes' Rule. Consider Casscells's *medical diagnosis problem*:

> If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? ___%[184]

Casscells found that 45% of the subjects answered 95% – a response that was way off the mark. Only 18% provided the correct Bayesian answer of 2%. Kahneman and Tversky

---

[181] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 53.

[182] Kahneman and Tversky offered the following as an example of one of the personality descriptions:

> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.
>
> The probability that Jack is one of the 30 engineers in the sample of 100 is ------%.

[183] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases."

[184] {Casscells, 1978 #513}

took Cassells's results to suggest that "even highly educated respondents often fail to appreciate the significance of outcome base rate in relatively simple formal problems" to increase the robustness and scope of the base rate fallacy:[185] they say that "[t]he failure to appreciate the relevance of prior probability in the presence of specific evidence is perhaps one of the most significant departures of intuition from the normative theory of prediction."[186] Kahneman and Tversky were content to broaden the scope of judgmental bias to include physicians and experts.

Later research reduced the scope of Kahneman and Tversky's stronger conclusion *and*, in addition, suggested ways to improve human judgment: psychologists discovered that subjects could be coaxed into providing normatively appropriate probability judgments by changing the experimental design, the information format,[187] the order in which information was presented,[188] or the way in which the personality descriptions were said to be selected.[189] Discovering the conditions that improved human judgment also seemed to have life or death consequences. For example, Cosmides and Tooby revised the diagnostic task by couching base rate information in terms of frequencies and asking subjects to provide their answers as frequencies rather than single-event probabilities greatly improved subject responses to the medical diagnosis problem:

> 1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive (i.e., the "true positive" rate is 100%). But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease (i.e., the "false positive" rate is 5%).
>
> Imagine that we have assembled a random sample of 1000 Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people. Given the information above:
>
> On average,
> How many people who test positive for the disease will *actually* have the disease? ___ out of ___"[190]

---

[185] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 154.
[186] Ibid., 243.
[187] Gigerenzer and Hoffrage, "How to Improve Bayesian Reasoning without Instruction: Frequency Formats."
[188] Jon A. Krosnick, Fan Li, and Darrin R. Lehman, "Conversational Conventions, Order of Information Acquisition, and the Effect of Base Rates and Individuating Information on Social Judgments," *Journal of Personality and Social Psychology* 59, no. 6 (1990): 1142-3.
[189] Zvi Ginossar and Yaacov Trope, "Problem Solving in Judgment under Uncertainty," *Journal of Personality and Social Psychology* 52, no. 3 (1987): 471.
[190] Cosmides and Tooby, "Are Humans Good Intuitive Statisticians after All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty," Experiment 2.

Cosmides and Tooby's experimental task elicited the correct Bayesian response in 56% of subjects, much higher than Cassells's 18%. If anything, subjects seemed to weight the base rate too heavily: 28% gave the lower 0.1% response. Only 4% of subjects provided Casscells's median response of 95%. By stating the statistical information in terms of frequencies, Cosmides and Tooby found a way to improve probabilistic judgment simply by taking advantage of cognitive abilities people already have. Hoffrage and Gigerenzer discovered similar kinds of frequency effects on physicians' judgments on the predictive power of diagnostic tests.[191]

Applied research on frequency effects suggests many fields in which we can improve human judgment. Research on how best to communicate risk information is a burgeoning, rich field of ecological rationalism.[192] The practical goals of this field are to improve patients' ability to understand and make judgments about risk, modify risk-relevant behavior, and facilitate cooperative, shared decision making.[193] Psychologists have sought to discover ways of presenting information to facilitate informed, reasoned medical decision making. Frequency formats help patients better understand risk and make better-informed medical decisions. In addition, displaying statistical information as pictographs helps patients to put anecdotal evidence in perspective and make better,

---

[191] Ulrich Hoffrage and Gerd Gigerenzer, "Using Natural Frequencies to Improve Diagnostic Inferences," *Academic Medicine* 73, no. 5 (1998). They were interested in discovering whether physicians could judge the positive predictive value of a diagnostic test: that is, the probability that a patient has a disease (in this case, breast cancer) given a positive diagnostic test, the sensitivity of the test (the probability that the test will show positive in the presence of disease), the rate of false positives (the probability the test will show positive when there is no disease), and the prevalence or base rate of the disease. Presenting the information in terms of frequencies improved physicians' estimates of the positive predictive value of a diagnostic test from 10% to 46% – an improvement, but not enough to rely on this strategy of improving diagnostic judgment alone. However, Gigerenzer suggests that physicians' estimates improved in the sense that "when the information was presented in natural frequencies, the physicians' estimates clustered around the correct answer." See Gerd Gigerenzer, "Is the Mind Irrational or Ecologically Rational?," in *The Law and Economics of Irrational Behavior*, ed. Francesco Parisi and Vernon L. Smith (Stanford, CA: Stanford University Press, 2005), 55.

[192] Eric R. Stone et al., "Foreground:Background Salience: Explaining the Effects of Graphical Displays on Risk Avoidance," *Organizational Behaviour and Human Decision Processes* 90, no. 1 (2003): 19.

[193] B. Rohrmann, "The Evaluation of Risk Communication Effectiveness," *Acta Psychologica* 81, no. 2 (1992): 170. See also Glyn Elwyn et al., "Decision Analysis in Patient Care," *Lancet* 358, no. 9281 (2001): 571, Baruch Fischhoff, "Risk Perception and Communication Unplugged: Twenty Years of Process," *Risk Analysis* 15, no. 2 (1995), Ann Fisher, "Risk Communication Challenges," *Risk Analysis* 11, no. 2 (1991). Peter A. Ubel, "Is Information Always a Good Thing? Helping Patients Make "Good" Decisions," *Medical Care* 40, no. 9 (2002), Peter A. Ubel, Christopher Jepson, and Jonathan Baron, "The Inclusion of Patient Testimonials in Decision Aids: Effects on Treatment Choices," *Medical Decision Making* 21, no. 1 (2001).

evidence-based medical judgments.[194] Applied research on frequency effects also suggest that frequency forms help jurists and judges draw statistical and Bayesian inferences from forensic DNA analyses.[195] Generally, people are less likely to convict when the DNA evidence is presented in terms of frequencies rather than single-event probabilities.[196] Identifying conditions that promote human judgment easily lend themselves to identifying the ways in which to improve politically and morally important cases of human judgment.

2.7 Ecological Rationalism: Current Research

2.71 Research in Applied Cognitive Psychology

Applied Cognitive Psychology is "concerned with understanding "real-life" problems in a theoretically satisfying way."[197] Stanovich points to two different ways in which applied cognitive psychology has sought to improve human judgment:[198] the Apologist approach and the Meliorist approach.

---

[194] Angela Fagerlin, C. Wang, and Peter A. Ubel, "Reducing the Influence of Anecdotal Reasoning on People's Health Care Decisions: Is a Picture Worth a Thousand Statistics?," *Medical Decision Making* 25, no. 4 (2005). The pictographs consisted of tiles arranged in a 10 x 10 matrix. Each tile presented an icon – a darkened silhouette of a person from the chest up, with the person's heart represented in white. Researchers represented the rate of success and failure of different medical interventions by the relative frequency of shaded icons. Successful cases were displayed the icon in the boldest contrast: black and white. Unsuccessful cases were displayed the icon in a lighter contrast: light gray and white.

[195] Jonathan J. Koehler, "When Are People Persuaded by DNA Match Statistics?," *Law and Human Behavior* 25 (2001). For example, consider the following expert testimony: "there is only a two percent chance the defendant's hair would be indistinguishable from that of the perpetrator if he were innocent." In contrast, consider the following statement: "only 2% of the people have hair that would be indistinguishable from that of the defendant and in a city of 1,000,000 people there would be 20,000 such individuals. Mock jurors are much less likely to convict when provided the second information format, which highlights a suspect's chance of matching by mere coincidence. Jonathan J. Koehler, "The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient," *Southern California Law Review* 74 (2001).

[196] Samuel Lindsey, Ralph Hertwig, and Gerd Gigerenzer, "Communicating Statistical Evidence," *Jurimetrics* 43, no. Winter (2003).

[197] Baddeley, "Applied Cognitive and Cognitive Applied Psychology: The Case of Face Recognition," 369.

[198] One of the advantages of ecological rationalism is that it is not committed at all to the claim that human judgment and/or choice is mostly or generally rational. Rather, ecological rationalism prefers discovering rational judgment for the purpose of improving human reasoning – a position that implicitly recognizes that human judgment can go awry in important ways. So, I reject what Keith Stanovich has called the *Panglossian position*: the claim that "human irrationality is a conceptual impossibility." He describes this

The *Apologist position* emphasizes "adapting the world to our cognitive machinery" by "presenting information in a way that is better suited to what our cognitive machinery is designed to do."[199] The Apologist approach has its roots in Herbert Simon's concept of bounded rationality. His bounded rationality research program sought the ways in which organisms like humans are rational given "the access to information and the computational capacities that are actually possessed by" them, "in the kinds of environments in which" they exist.[200] Fischoff suggested directing psychological research towards studies that are more "honest" in the sense that they look to "the person-task system" and focus on tasks that makes subject-environment systems "as compatible as possible."[201] By 1993, the Apologist's approach to improving human reasoning by creating more hospitable conditions of reasoning had still "received relatively little attention."[202] Joshua Klayman and Kaye Brown provided a clear statement of the Apologist approach by suggesting that researchers seek to improve human judgment by "debiasing the environment:" rather than modify "cognitive processes to fit the environment better, one can modify the environment to fit the processes that people bring to it."[203] Such a research program embraces the lesson of context-specificity: it seeks to identify "sub-environments in which people could be doing better given their goals and their resources" for the purposes of designing conditions "that avoid or compensate for anticipated errors."[204] Changes in the environment include stating statistical information as frequencies rather than single-event

---

position as being "most often represented by philosophers" and mentions in this regard Cohen and Dennett (to this list, I would add Davidson). Keith E. Stanovich, *Who Is Rational? Studies of Individual Differences in Reasoning* (Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 1999), 5-7. Paul Thagard and Richard Nisbett have rightly pointed out too strong, a charitable assumption "preempts the possibilities of criticism and improvement. If we cannot assume actions and judgments to be irrational, then we cannot hope to educate and improve choice strategies and inferential procedures." Thagard and Nisbett, "Rationality and Charity," 263. It is precisely the hope of improving human judgment that motivates ecological rationalism.

[199] Stanovich, *Who Is Rational? Studies of Individual Differences in Reasoning*, 7-8. Stanovich includes in this group of psychologists evolutionary psychologists such as Gigerenzer, Goldstein, Tooby, and Cosmides.

[200] Simon, "A Behavioral Model of Rational Choice," 99.

[201] Fischhoff, "Debiasing," 427.

[202] Joshua Klayman and Kaye Brown, "Debias the Environment Instead of the Judge: An Alternative Approach to Reducing Error in Diagnostic (and Other) Judgment," *Cognition* 49 (1993): 100.

[203] Ibid.

[204] Ibid.: 98.

probabilities. The Apologists's strategy of modifying the environment to improve human judgment is a very practical implementation of research guided by ecological rationalism.

In contrast, the *Meliorist position* "emphasizes the possibility of getting our cognitive machinery to operate differently"[205] by means of education and training. Education can help people develop analytical tools "to supplement or supplant" invalid intuitions.[206] There is some reason to think that Kahneman and Tversky were interested in a Meliorist approach, though their own research did not focus on education. Between 1974 and 1980, a group of scientists and educators at the University of Jerusalem developed a textbook for 14-year-olds "to improve their probabilistic thinking skills, introduce concepts like *uncertainty*, point out circumstances under which our thinking processes lead us astray, and suggest tools to improve our skills when dealing with uncertainty."[207] "The primary encouragement for the development of a curriculum on thinking under uncertainty – of which this book is one product – came from two people" seeking "the incorporation of scientific (educational and psychological) ideas into school curricula."[208] One of the men credited was Daniel Kahneman.

When people have intuitive understanding of a statistical concept such as the law of large numbers teaching subjects how to reason better can be relatively easy.[209] For more difficult statistical concepts, education can focus on providing heuristics that make the concepts intuitively accessible.[210] Psychological research by Richard Nisbett and

---

[205] Stanovich suggests that "early work in the heuristics and biases tradition" by researchers like Tversky, Kahneman, Nisbett, and Ross belong in this category. Stanovich, *Who Is Rational? Studies of Individual Differences in Reasoning*, 6-8.

[206] Klayman and Brown, "Debias the Environment Instead of the Judge: An Alternative Approach to Reducing Error in Diagnostic (and Other) Judgment," 99.

[207] Ruth Beyth-Maron et al., *An Elementary Approach to Thinking under Uncertainty*, trans. Sarah Lichtenstein, Benny Marom, and Ruth Beyth-Marom (Mahwah, NJ: Lawrence Erlbaum Associates, 1985), ix.

[208] Ibid., x.

[209] Psychologists suggest that people possess an abstract inferential rule system that serves as an intuitive version of the law of large numbers because, with a little bit of training, the frequency and quality of subjects' statistical reasoning increased for a wide variety of problems – even in domains beyond the examples considered during training. Geoffrey T. Fong, David H. Krantz, and Richard E. Nisbett, "The Effects of Statistical Training on Thinking About Everyday Problems," *Cognitive Psychology* 18 (1986).

[210] For example, education might also seek to provide people with less formal cognitive tools – heuristics – that make statistical notions intuitively accessible. For example, Richard Nisbett and his colleagues expressed optimism over whether we could improve human judgment by teaching inferential maxims that made scientists' inferential tools accessible to the lay person. Such inferential maxims included things like: "You can always explain away the exceptions" and "Think about evidence as if it were a sample, and reflect about sample size." Richard E. Nisbett et al., "Improving Inductive Inference," in *Judgment under*

Legman has also focused on how teaching more abstract rules improves statistical judgments[211] and on how different fields of study foster different analytical tools in undergraduate and graduate students.[212] Researchers also suggest that teaching people about their own judgment biases and judgment reliability into a larger perspective.[213] Such studies only begin to touch the practical question of how best education people to be better problem-solvers: "we know very little" about how to teach concepts and rules for good judgment. How much "we can improve reasoning by instruction" is still "a completely open question."[214] Insofar as education and training are "conditions" that improve human judgment, the Meliorist's strategy of improving human judgment is in line with the motivations of ecological rationalism.

2.72 Ecological Rationalism in the Social Sciences

The social and moral interest in promoting the interest in rational judgment provides a very different kind of conception of and rationale for psychological research. In traditional accounts of methodological rationalism, the preference for rationalizing

---

Uncertainty: Heuristics and Biases, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (Cambridge, U.K.: Cambridge University Press, 1982), 445.

[211] In later research, Nisbett and his colleagues discovered results that "purely abstract rule training produced improvement in both the frequency and the quality of statistical answers" and that providing "training on examples readily generalized to domains very different from the trained domain." Nisbett et al., "Teaching Reasoning," 304-5.

[212] Lehman and his colleagues discovered that undergraduate training in psychology and the social sciences more generally improve students' statistical reasoning about a wide range of problems and that undergraduate students of the natural sciences and humanities showed only marginally improvements in statistical reasoning. They found no disciplinary differences in the scores of the same students during their first year in college. Darrin R. Lehman and Richard E. Nisbett, "A Longitudinal Study of the Effects of Undergraduate Training on Reasoning," in *Rule for Reasoning*, ed. Richard E. Nisbett (Hillsdale, NJ: Lawrence Erlbaum Associates, 1993), 353-4. Lehman and his colleagues also discovered that graduate education in psychology in medicine used statistical ideas to solve scientific and everyday-life problems (in contrast, graduate students in Chemistry and Law did not exhibit such abilities). Darrin R. Lehman, Richard E. Nisbett, and Richard O. Lempert, "The Effects of Graduate Training on Reasoning: Formal Discipline Nad Thinking About Everyday-Life Events," in *Rule for Reasoning*, ed. Richard E. Nisbett (Hillsdale, NJ: Larence Erlbaum Associates, 1993), 330.

[213] For example, Gary Gaeth and James Shanteau evaluated the effectiveness of using two training procedures (lecture versus interactive teaching) designed to reduce the adverse influence of irrelevant information in making judgments under uncertainty. They found that people could not learn to ignore irrelevant information, but could learn to recognize their tendency to over-attend to irrelevant information and compensate for it. G. J. Gaeth and James Shanteau, "Reducing the Influence of Irrelevant Information on Experienced Decision Makers," *Organizational Behavior and Human Performance* 33 (1984).

[214] Lehman, Nisbett, and Lempert, "The Effects of Graduate Training on Reasoning: Formal Discipline Nad Thinking About Everyday-Life Events," 335-6.

interpretations bear on different methodological roles rationality plays in interpretation. In ecological rationalism, rational judgment itself becomes an object of study. The interest in discovering conditions promoting rational rather than irrational judgment is motivated by a social and moral interest in promoting cognitive health.

Ecological rationalism is not necessarily unique to interpretation in cognitive psychology. The social sciences more generally depart from the hard sciences in important ways: the social sciences are interested in explaining politically and morally significant human behavior in terms of intentional states, the environments created by social institutions, and the dynamic relationship between the two. Cognitive psychology guided by ecological rationalism is no different. Applied cognitive psychology suggests different information formats institutions might adopt to communicate risk and improve human judgment. Applied cognitive psychology also suggests ways that educational institutions teach its students about statistical and probabilistic concepts and rules.

More broadly, ecological rationalism suggests expanding the domain of study to sociological questions about how institutions are implicated in good human judgment and decision making. For example, Arthur Lupia has argued that American citizens successfully use party-affiliation as a reliable heuristic in deciding who to vote for: American political party systems are structured in ways that enable our notoriously ignorant citizenry to use this limited, but reliable information to cast reasoned votes.[215] Vernon Smith suggests that it is the job of our economic institutions to coax "Pareto-efficient behavior out of agents who do not know what that means."[216] Legal scholars have suggested that additional regulations in tort and contract law to deflect negative consequences of individuals' judgment biases.[217] Likewise, researchers in organizational

---

[215] Arthur Lupia and Mathew D. McCubbins, *The Democratic Dilemma: Can Citizens Learn What They Need to Know?*, ed. Randall Calvert and Thrainn Eggertsson, *Political Economy of Institutions and Decisions* (Cambridge, U.K.: Cambridge University Press, 1998), Arthur Lupia, Mathew D. McCubbins, and Samuel L. Popkin, eds., *Elements of Reason: Cognition, Choice, and the Bounds of Rationality, Cambridge Studies in Political Psychology and Public Opinion* (Cambridge, U.K.: Cambridge University Press, 2000).

[216] Vernon L. Smith, "Rational Choice: The Contrast between Economics and Psychology," *Journal of Political Economy* 99, no. 4 (1991): 894. See also Vernon L. Smith, *Bargaining and Market Behavior: Essays in Experimental Economics* (New York, NY: Cambridge University Press, 2000).

[217] Eric A. Posner, "Probability Errors: Some Positive and Normative Implications for Tort and Contract Law," in *The Law and Economics of Irrational Behavior*, ed. Francesco Parisi and Vernon L. Smith (Stanford, CA: Stanford University Press, 2005). Eric Posner argues that laws serve to protect individuals

psychology suggest the implementation of cognitive repairs to deflect negative consequences of judgmental biases.[218]

2.8 Conclusion

Researchers who actively sought to limit the scope of Kahneman and Tversky's claims about human irrationality mobilized a disciplinary return to studying rational judgment. This research underscored the methodological point that experimental evidence can only properly support claims about the *particular* ways in which we are rational or irrational in *specific* contexts of reasoning. Practically oriented researchers of the applied cognitive stripe began to focus on identifying the conditions promoting rational judgment because of their practical implications: such research provided better grounds for recommendations about how to change contexts, educational strategies, and institutions to improve human judgment. This change in research aim was motivated by something like an interest in promoting cognitive health: prominent researchers explicitly argued that moral and political interests should determine what kinds of judgments, tasks, and subjects should constitute significant areas of research. The contextual interest in promoting cognitive health is not unique to cognitive psychology, but can be of general interest to researchers in the social sciences more generally.

Social and moral interests are not the only motivations behind shifts in disciplinary focus, however. In chapter 3, I will discuss how psychology's interest in creating valid questionnaires created a new disciplinary focus on discovering the

---

from their insensitivity to small differences between probabilities and optimism with respect to low-probability events.

[218] Klayman provides the following illustration of an organizational-level cognitive repair: civil engineers have been shown to exhibit overconfidence in their judgments about the height at which a structure will fail. To mitigate this kind of overconfidence, the engineering profession has created specific kinds of cognitive repairs – namely, safety factors: "In an actual assessment civil engineers would precisely calculate the amount and strength of foundation materials necessary to hold a structure of a particular height, then they would multiply their precise answer by a safety factor (i.e., a number between three and eight), and use the larger figure to build the foundation." Joshua Klayman, Chip Heath, and Richard P. Larrick, "Cognitive Repairs: How Organizational Practices Can Compensate for Individual Shortcomings," *Research in Organizational Behavior* 20 (1998): 4. The original study on overconfidence was provided by M. Hynes and E. Vanmarcke, *Reliability of Embankment Performance Predictions, Proceedings of the Asce Engineering Mechanics Division Specialty Conference* (Waterloo, Ontario, Canada: University of Waterloo Press, 1976).

conditions of successful versus unsuccessful communication in experimental contexts. In this chapter, I will say more about the conditions in which we should adopt methodological rationalism and prefer charitable interpretations in interpreting subject responses in experimental contexts. In chapter 4, I will discuss how psychology's internal concerns about what counts as a legitimate or good explanation connects up with the explanatory interests of naturalized epistemology.

Chapter 3

The Gricean Turn in Psychology

Traditional accounts of charitable interpretation that rely on norms of rationality to guide interpretation have typically invoked rules of logic and probability, as well as principles of evidence or justification – while overlooking norms governing the social and communicative relationships between the interpreter and interpreted.[219]

Psychologists working on conversational pragmatics and judgment have observed the same oversight in their field: researchers, especially those from the heuristics and biases tradition, who tend to argue that subjects are systematically irrational, have neglected to consider how social or conversational norms may influence subjects' interpretations of and communications with experimenters.[220] In response, psychologists such as Norbert Schwarz, Denis Hilton, and Gerd Gigerenzer have invoked Paul Grice's principles of cooperative communication to attribute alternative interpretations of experimental tasks to subjects – interpretations for which subject responses may be said to be "conversationally rational."[221]

My account of charitable interpretation broadens traditional charitable accounts by recognizing conversational norms as rational principles of conversational inference. I will call the general method of using conversational principles to guide interpretation *Gricean charity*. Gricean charity provides a naturalized account of charity, which looks to facts about natural language and communication in the interpretation of subject

[219] This chapter is also available in published form. See Carole J. Lee, "Gricean Charity: The Gricean Turn in Psychology," *Philosophy of the Social Sciences* 36, no. 2 (2006).
[220] Denis J. Hilton, "The Social Context of Reasoning: Conversational Inference and Rational Judgment," *Psychological Bulletin* 118, no. 2 (1995): 249.
[221] Norbert Schwarz, "Judgment in a Social Context: Biases, Shortcomings, and the Logic of Conversation," *Advances in Experimental Social Psychology* 26 (1994).

responses. I take psychologists' work in conversational pragmatics as primary exemplars of Gricean charity at work. This work demonstrates that a broader perspective on rationality and the nature of subject-experimenter communication imports specific evidential requirements on psychological studies: namely, that subject responses be interpreted in light of empirical information about (i) successful and unsuccessful communication in specific experimental contexts, and (ii) the conversational norms governing communication in experimental conditions.

The genealogy of this charitable approach may be traced back to "the presentation problem" faced by Ward Edwards.[222] In order to test human performance on rational choice tasks, psychologists have to put the task and options into words. However, turning the decision task into a word problem adds an additional level of complexity for both the subject and researcher. For the subject, natural language expressions are often ambiguous, and may support any number of meanings. So, the subject must interpret the intended meaning of the stated task, and provide a response under that interpretation. As a result, the subject's choice behavior is influenced by her interpretation of the experimental task. Edwards's observation may be captured by a more general Davidsonian lesson: any psychological theory on human judgment "must *include* a theory of interpretation" about subjects' beliefs about the experimental task.[223]

In the first part of this paper, I will lay out my account of Gricean charity. First, I will argue that conversational norms are indeed relevant to the questionnaires and surveys used in psychological testing. To illustrate, I will reinterpret some portions of Kahneman and Tversky's Linda questionnaire in light of Gricean conversational maxims, and use this reinterpretation to rationalize subject responses. This analysis serves to highlight the methodological lessons of the Gricean turn in psychological research. In the second part of this paper, I will consider and respond to methodologically motivated objections to Gricean charity. In the course of responding to these objections, I will argue that Gricean charity generates new psychologically interesting questions, phenomena, and methods, without harboring scientifically illegitimate forms of bias.

---

[222] Ward Edwards discusses this kind of problem, though not under the rubric "presentation problem." Edwards, Lindman, and Phillips, "Emerging Technologies for Making Decisions."
[223] Davidson, "Belief and the Basis of Meaning (1974)," 147. The italics are mine.

3.1 Questionnaires as Forms of Cooperative Communication

Conversational pragmatics invokes normative principles of communication to account for how subjects arrive at their interpretations of the experimental task. For an account of these norms, researchers have turned to Paul Grice's account of cooperative communication. According to Grice, cooperative communication aims to use language efficiently and effectively to further a common goal or set of goals. Communication is said to be rational insofar as it conforms to conversational principles that are themselves instrumental in furthering these co-operative ends.[224] The most general principle, the *Cooperative Principle* (CP), directs conversants to "[m]ake your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."[225] Grice analyzes *CP* into the following general principles:

> *Quality*: (i) do not say what you believe to be false and (ii) do not say that for which you lack adequate evidence.
>
> *Quantity*: (i) make your contribution as informative as is required for the current purposes of the exchange and (ii) do not make your contribution more informative than is required.
>
> *Relevance*: make your contribution relevant.
>
> *Manner*: (i) avoid obscurity of expression, (ii) avoid ambiguity, (iii) be brief (avoid unnecessary prolixity) and (iv) be orderly (provide information in a sequentially accessible way).

The conversational maxims provide interpretations of utterances of natural language expressions. Natural language expressions often imply multiple meanings, or imply meanings that are not captured by the literal statement as expressed. The conversational maxims provide a way of inferring a speaker's intentionally implied meaning – that is, the *conversational implicature* – even when this meaning goes beyond the literal meaning of what she has said.[226] Conversational implicatures are *calculable* in

---

[224] Paul Grice, "Logic and Conversation," in *Studies in the Way of Words* (Cambridge, MA: Harvard University Press, 1989), 28.
[225] Ibid., 26.
[226] The special case Grice considers is one in which a conversant *flouts* a conversational maxim. Here, a conversant provides a contribution that, when taken literally, violates one of the maxims. For example,

the sense that, for every putative implicature, it is possible to construct an inductive argument showing how the implicature follows given the literal meaning of the utterance, mutually recognized facts about the particular context of communication, and the conversational maxims.[227] If the hearer discovers that any of these premises are false, then the implicature is *cancelled.*

Psychologists working in conversational pragmatics have argued that questionnaires can be understood as forms of cooperative communication. As Hilton points out, psychological experiments and surveys used to test cognitive competence "are forms of social interaction between the experimenter and participant, which invariably involve communication through ordinary language."[228] Experimenters provide subjects information via instructions, questionnaires, and surveys; and subjects provide responses in the form required. As in other conversational settings, these forms of information exchange are carried out in ordinary language.

Researchers in conversational pragmatics have argued that experimenters and subjects share a mutual recognition that their efficient, effective acts of communication serve mutually understood goals. Subjects know that experimenters aim to collect data – namely, subjects' responses to pre-designed questions and experimental tasks. Since subjects cannot ask for clarification or paraphrases of the question, the experimental

---

let's say Dick and Jane have plans to meet with a third person Tom, who is very late (and usually so). Dick asks, "What happened to Tom?" Jane responds, "Tom's watch must operate counter-clockwise." Jane's contribution, when taken literally, is false – she *knows* Tom's watch does not operate counter-clockwise. Jane flouts the maxim of quality which enjoins her to provide true or well-founded contributions. In order to construe Jane's contribution as conforming to the maxim of quality, we infer that she intends to imply something beyond the literal meaning of her statement – something that is actually true – perhaps that "Tom has profound trouble keeping track of time." Jane expects Dick to be able to infer this implied meaning, given what he knows about Jane, the conversational context, and the conversational maxims. The interpretation of a wide range of linguistic phenomena (such as figures of speech, hyperbole, metonymy, irony, and metaphor) may be subsumed under the more general problem of interpreting implicatures.

[227] John Levinson analyzes the inductive argument constructed by the hearer in the following way:

> (i) *S* has said that *p*.
> (ii) There's no reason to think *S* is not observing the maxims or at least the co-operative principle.
> (iii) In order for *S* to say that *p* and be indeed observing the maxims or the co-operative principle, *S* must think that *q*.
> (iv) *S* must know that it is mutual knowledge that *q* must be supposed if *S* is to be taken to be co-operating.
> (v) *S* has done nothing to stop me, the addressee, from thinking that *q*.
> (vi) Therefore, *S* intends me to think that *q*, and in saying that *p* has implicated *q*.

Stephen Levinson, "Conversational Implicature," in *Pragmatics* (Cambridge: Cambridge University Press, 1983), 113-4.

[228] Hilton, "The Social Context of Reasoning: Conversational Inference and Rational Judgment," 249.

conditions encourage them to assume that the meaning of the questions and tasks are self-evident.[229] With this knowledge, subjects can expect that experimenters expedite this process by stating their questions and tasks clearly, concisely, and sufficiently. That is, subjects may reasonably expect questionnaires and surveys to conform to the maxims of quantity, relevance, and manner. For subjects, experiments are conducted by their academic and epistemic authority figures: much current research is conducted on undergraduate Psychology students by graduate students and professors. Because of the asymmetry in authority, expertise, and knowledge between experimenters and subjects, subjects may reasonably expect experimenter-provided information to be especially truthful and well-supported, in accordance with the maxim of quality.[230]

Experimenters know that subjects characteristically participate in the experiments to fulfill requirements for their introductory psychology courses, or for small monetary rewards. Coming into a psychological experiment, subjects come with prior expectations about the experimenter's goals. In particular, they know that the experimenters have designed special questions, with the goal of evaluating and explaining their answers. In order for their answers to be constructive or relevant towards the experimenters' goals, the subjects are expected to answer sincerely, in conformance with the maxim of quality; and, in order to help the experimenters carry out this goal efficiently (given the number of subjects involved), they are expected to answer in conformance with the maxims of relevance, quantity, and manner.

Even when subject responses take the form of checking boxes, circling multiple choice options, or ranking outcomes, these maxims still apply. Even with such regimented forms of response, the maxim of quality dictates that subjects provide honest rather than dishonest answers. The maxim of relevance directs subjects to answer the experimenter's intended question, and not a different question that may be more amusing to entertain. The maxim of quality enjoins subjects to provide sufficient answers by answering all questions. And, in accordance with the maxim of manner, subjects are expected to provide unambiguous, clearly marked, and quickly deciphered responses: for

---

[229] Herbert H. Clark and Michael F. Schober, "Asking Questions and Influencing Answers," in *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*, ed. Judith M. Tanur (New York, NY: Russell Sage Foundation, 1992), 26.

[230] Hilton, "The Social Context of Reasoning: Conversational Inference and Rational Judgment," 254.

example, it would be inappropriate for a subject to provide a long-winded essay in answering a multiple-choice question.

In the experimental context, subjects can rely on special kinds of clues in interpreting the experimenters' intended meanings – clues such as the wording of the task, the questionnaire's previous questions, the formal structure of the questionnaire, and interactions with and assumptions about the experimenter.[231] Such evidence grounds key implicatures about the meaning of the experimental task or question.

3.2 The Linda Problem

To see how conversational norms and assumptions can inform our interpretation of subjects' responses, consider Daniel Kahneman and Amos Tversky's *Linda Problem*:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
>
> Linda is a teacher in elementary school.
> Linda works in a bookstore and takes Yoga classes.
> Linda is active in the feminist movement.
> Linda is a psychiatric social worker.
> Linda is a member of the league of Women Voters.
> Linda is a bank teller.
> Linda is an insurance salesperson.
> Linda is a bank teller and is active in the feminist movement.[232]

According to probability theory, the probability of two events A and B is equal to or less than the probability of each of its conjuncts: $p(A \,\&\, B) \leq p(A)$ and $p(A \,\&\, B) \leq p(B)$. If we identify Linda's being active in the feminist movement as event A, and Linda's being a bank teller as event B, subjects should rank the probability that Linda's both a bank teller and active in the feminist movement ($A \,\&\, B$) as equal or lower than the ranking for A or B considered alone. They found that the vast majority of statistically naïve *and*

---

[231] For a nice review of this research, see Norbert Schwarz, *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation* (Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 1996).

[232] Tversky and Kahneman, "Judgments of and by Representativeness."

statistically sophisticated subjects rated the conjunction of events as more probable than the conjunct, in violation of the conjunction rule.

However, it isn't clear that subjects can be said to have violated the conjunction rule, if we reinterpret the questionnaire in light of Grice's conversational maxims.[233] The only information required to rank the probabilities in accordance with the conjunction principle are two particular outcomes: the outcomes "Linda is a bank teller" and "Linda is a bank teller and is active in the feminist movement." The rest of the cover story, such as the personality description and the other outcomes, are irrelevant, unnecessary, and superfluous.[234] The experimenters' question, as stated, violates the maxims of relevance, quantity, and manner.[235] These violations have implications for the questionnaire's validity. If subjects enter the experiment under the assumption that experimenters are cooperative communicators, they would rule out this interpretation, since it construes experimenters as violating conversational norms.

Subjects are thus put in the position of interpreting the meaning of the experimenters' question so that it conforms to the conversational maxims. Such an interpretation must render *all* of the information provided by experimenters as useful and relevant to solving the intended question. One way to render the extra personality and outcome information useful and relevant, is to take the experimenters as asking something other than a mathematical probability question. The term "probable" is *polysemous*: it can also be interpreted as meaning "plausible," "having an appearance of truth," or "reasonable in light of the evidence."[236] Subjects, faced with the problem of inferring which of these meanings experimenters implicate, might interpret the question as a *plausibility* problem, where an outcome is said to be more "plausible" insofar as it

---

[233] The line of argument here is due to Hertwig and Gigerenzer, "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors."

[234] Jonathan E. Adler, "Abstraction Is Uncooperative," *Journal for the Theory of Social Behavior* 14, no. 2 (1984).

[235] Notice that the personality description and the rest of the outcomes *are* relevant to the question of how the outcomes "Linda is a bank teller" and "Linda is a bank teller and is active in the feminist movement" rank relative to the other outcomes. However, the personality information and other outcomes are *not* relevant to the question Kahneman and Tversky are primarily interested in: namely, how the outcomes "Linda is a bank teller" and "Linda is a bank teller and is active in the feminist movement" rank relative to each other.

[236] Hertwig and Gigerenzer, "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors." See also Gigerenzer, "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)."

has something to speak I favor of it. Under this interpretation of the question, it is not incorrect to judge that the conjunction of events as more plausible than its conjunct: given the personality description, the outcome "Linda is a bank teller" has nothing to speak in favor of it; however, the outcome "Linda is a bank teller and is active in the feminist movement" does have something to speak in favor of it, since we expect Linda's commitment to liberal values would be expressed in her choice of occupation and/or hobbies.[237]

The notion of plausibility might also be understood as having something to do with how well the personality information might explain the different outcomes: the personality description provided has more explanatory strength when it comes to explaining why "Linda is a bank teller and active in the feminist movement," as opposed to explaining why "Linda is a bank teller." If subjects interpret "probability" to mean something like "degree to which they can be explained," then subjects' responses cannot be said to be incorrect. Not only does the personality description better explain the conjunction than the conjunct, but – as Kahneman and Tversky themselves point out – best explanations and most probable outcomes often have an "inverse relationship:"[238] the value of an explanation can be improved by increasing its content and scope, even though the probability of its truth is reduced thereby.[239]

---

[237] An alternative interpretation of "plausibility" might be provided by the notion of "conceptual coherence." For a model of conceptual coherence that bears on how stereotypes and individuating information can inform each other, see Ziva Kunda and Paul Thagard, "Forming Impressions from Stereotypes, Traits, and Behaviors: A Parallel-Constraint-Satisfaction Theory," *Psychological Review* 103, no. 2 (1996).

[238] Tversky and Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," 312.

[239] Conversational implicature may also play a role in subjects' interpretation of the outcome "Linda is a bank teller." Subjects might interpret this outcome, when stated alone, as the most information the experimenters can assert with confidence according to the maxim of quality. However, when presented with the accompanying outcome "Linda is a bank teller who is active in the feminist movement," the meaning of "Linda is a bank teller" is unclear. Subjects may have taken the additional information "who is active in the feminist movement" as indicating a kind of *contrast*, where "Linda is a bank teller" is supposed to mean that "Linda is a bank teller who is *not* active in the feminist movement." This interpretation would distinguish the outcomes "Linda is a bank teller" and "Linda is a bank teller who is active in the feminist movement" as non-overlapping events. To prevent subjects from adopting this interpretation, Tversky and Kahneman rephrased "Linda is a bank teller" as "Linda is a bank teller, whether or not she is active in the feminist movement." However, Don Dulany and Denis Hilton found that only 28% of subjects interpreted "whether or not she is active in the feminist movement" in the way Tversky and Kahneman hoped. See Tversky and Kahneman, "Judgments of and by Representativeness." Don E. Dulany and Denis J. Hilton, "Conversational Implicature, Conscious Representation, and the Conjunction Fallacy," *Social Cognition* 9, no. 1 (1991).

3.3 Methodological Implications of the Gricean Turn

Gricean charity provides an interpretation of subjects that rationalizes their apparently irrational responses. Grice's conversational maxims suggest that Kahneman and Tversky's statement of the Linda problem misleads at least some subjects to solve a different problem – a problem for which subjects' responses are rational. Accordingly, Kahneman and Tversky are not entitled to describe subjects as providing irrational responses.

This is the first methodological lesson I draw from the Gricean turn in psychological research: the conclusions of any psychological study can only be as valid as its questionnaires and surveys. If experimenter communications violate the conversational maxims, leaving greater room for the interpretation of unintended implicatures, then we cannot take subjects' responses at face value – much less generalize or explain them. The Gricean turn suggests researchers studying human reasoning create semantically clear questionnaires and surveys, in conformance with conversational maxims. Even Kahneman and Tversky seem to accept this point. In the *Postscript* of their canonical book *Heuristics and Biases*, Kahneman and Tversky admit that apparently irrational responses elicited from subjects could have arisen from the subject's misunderstanding of the question.[240] And, they concede that the "conversational aspect of judgment studies deserves more careful consideration than it has received in past research, *our own included*."[241]

Another methodological lesson of the Gricean turn is that experimenters, in arguing for the validity of their questionnaires and surveys, must collect data about subjects' interpretations of the experimental tasks. Kahneman and Tversky's research up to that point had been conducted using a "*question-answering paradigm*," where "the subject is exposed to information and is asked to answer questions or to estimate values,

---

[240] Kahneman and Tversky, "On the Study of Statistical Intuitions," 493.
[241] Ibid., 504. Italics mine.

orally or in writing."[242] Subjects were provided uniform information from experimenters; and, experimenters assumed subjects' interpretations would be uniformly identical with their own. However, because communicating clearly in experimental conditions is itself a challenge, this is not a safe assumption.[243]

The Gricean turn also reminds us that effective communication requires cooperation by both experimenter and subject. Indeed, experimenters shoulder a *greater* responsibility for clear communication because of the asymmetrical nature of experimenter-subject communication. Unlike ordinary conversation, the course and content of communication in experimental conditions are predetermined by only one conversant, namely the experimenter.[244] Experimenters can direct the subjects and ask any number of clarificatory questions, though the converse is not true. As a result, experimenters are especially responsible for making "correct assumptions about the codes and contextual information that the audience will have accessible and be likely to use in the comprehension process."[245]

The Gricean turn changes the nature of the investigation – away from sweeping charges of rationality or irrationality lodged against subjects – toward exploration of the communicative conditions that either tend to mislead subjects or that tend to facilitate their successful performance.[246] Such investigation, undertaken by psychologists studying conversational pragmatics, is *situational* since it focuses on *experimental conditions* to explain the apparent rationality or irrationality of subject responses. The situational explanation here is of a special kind, in which the experimenter and the experimenter's relationship with the subject partly constitute the explanatory situation/context. The situational explanation suggested by Gricean charity is *reflexive*,

---

[242] Ibid., 501.

[243] Indeed, subsequent testing has found that subjects usually misinterpret the meaning of key words and phrases in the Linda problem and other problems from the heuristics and biases tradition. Hertwig and Gigerenzer used a multiple-choice method in checking subjects' interpreted meaning of "probability," by asking them to check which of a list of terms best reflected their understanding of "probability" in the Linda problem. They found that only 12% of checked choices were mathematical (e.g., "expectancy," "frequency," "percentage," "logicality," or "certainty"). See Hertwig and Gigerenzer, "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors," 280-2.

[244] Clark and Schober, "Asking Questions and Influencing Answers," 26.

[245] Dan Sperber and Deirdre Wilson claim this asymmetry in communication exists in all communicative contexts. I do not agree with this stronger claim. Dan Sperber and Deirdre Wilson, *Relevance: Communication and Cognition* (Oxford, U.K.: Basil Blackwell Ltd., 1986), 43.

[246] Thanks to Elizabeth Anderson for this insight.

insofar as it considers how the experimenter and the experimenter's relationship with the subject play a role in subjects' observed behavior. Such situational explanations shift investigation away from *attributional* explanations, which focus on characteristics of the subjects themselves – such as flawed or limited memory, attention, search, or reasoning strategies. The great irony here is that explanations that unfairly blame subject responses on internalized judgment heuristics risk committing the fundamental attribution error.[247]

3.4 Naturalized Conversational Norms

The great methodological insight of conversational pragmatics is that, rather than make unsubstantiated assumptions about subjects' interpretations of experimental tasks, researchers ought to gather evidence to identify those interpretations and identify conditions of successful versus unsuccessful communication. However, in hypothesizing about subjects' conversational inferences, critics might argue that researchers in conversational pragmatics have adopted a few working assumptions of their own about subjects' conversational assumptions and norms. In particular, researchers in conversational pragmatics have uncritically adopted Grice's maxims of conversation as the relevant norms of conversation in subject-experimenter communication. One possible explanation for this is that researchers have regarded Grice's conversational maxims as universal norms of conversation.

Regarding Grice's maxims as universal norms of conversation would fail to respect the context and cultural relativity of such norms. Every day experience demonstrates that the content of conversational norms varies depending on the context and goals of communication. For example, we recognize that in contexts where conversation is made for the sake of mutual entertainment, the maxims of manner, quantity, and quality do not apply: many of the anecdotes, jokes, and tangents we trade are valued precisely for the creative ways in which the description of true or merely hypothetical events are drawn out, elaborated, and exaggerated. Additionally, the conversational norms may depend on broader cultural norms or goals: the research of

[247] Hilton, "The Social Context of Reasoning: Conversational Inference and Rational Judgment," 249.

linguistic anthropologists demonstrates that conversational norms we hold dear are not universal, but reflect the goals and cultural norms of specific social milieus.[248]

Regarding Grice's maxims as universal norms of conversation would also fail to capture the dynamic relationship between conversational norms and subjects' goals, questionnaire content, and communicator identity. To illustrate, consider the following case of subject-experimenter communication. Subjects familiar with psychological testing may know that the ostensible purpose of an experiment is often different from its real purpose. Subjects who know this may think it likely that experimenters are trying to deceive them in the sense that the experimenter aims to gather information about the subject that is not directly asked for. Since this assumption is true more often than not, especially in questionnaires from the heuristics and biases tradition, it would be reasonable for subjects *not* to assume that experimenters' conversational contributions conform to the maxims of quality, quantity, or relevance. Subjects who look skeptically upon the questions asked of them may adopt the goal of figuring out what the experimenters are *really* after, and provide answers that seek to uncover or frustrate the experimenter's goals.[249] In this scenario, subjects would best be described as rejecting Grice's maxims of quality, quantity, and relevance.

Although researchers in conversational pragmatics have not explored conversational norms that differ in content from Grice's maxims, it would be unfair to construe them as adopting Gricean norms as universal norms of conversation. These researchers have gone to great lengths to cite a broad range of studies suggesting that subjects respect something like the maxims of quality, relevance, manner, and quantity.[250] They have also sought to empirically support their account of subjects' conversational assumptions. For example, they cite experiments studying how subjects'

---

[248] For example, Michelle Rosaldo's fieldwork suggests that the Ilongots do not share Grice's commitment to the maxim of quality. She argues that the primary conversational norm in Ilongot culture directs conversants to provide conversational contributions that maintain social roles and relationships, regardless of the truth of those assertions. Michelle Z. Rosaldo, *Knowledge and Passion: Ilongot Notions of Self and Social Life* (Cambridge, UK: Cambridge University Press, 1980). For similar kinds of critiques in the context of Malagasy society, see Elinor Ochs Keenan, "The Universality of Conversational Postulates," *Language in Society* 5 (1976).

[249] Thanks to Peter Railton for raising this possibility.

[250] For extensive reviews, see Hilton, "The Social Context of Reasoning: Conversational Inference and Rational Judgment." See also Schwarz, *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation.*

judgments of trustworthiness (on the part of the experimenter) vary with changes in the content of communication and the identity of the communicator.[251] The implication of citing this research is supposed to be that subjects' judgments of trustworthiness may have important effects on what conversational norms govern subject-experimenter communication.

The best interpretation of researchers' commitment to Grice's maxims is to interpret them as adopting Grice's maxims as empirically compelling formulations of the conversational norms that seem to govern subject-experimenter communication in certain types of experimental contexts. Grice's maxims are not presented as uncontestable social facts, but as working hypotheses. However, in embracing a naturalized account of conversational norms, conversational pragmatics would be well served by checking its assumptions about subjects' conversational assumptions and norms with respect to particular cases of subject-experimenter communication. Such research would build on the lessons of an older psychological literature on source effects that concerned itself with how subjects' goals influence research results.[252]

3.5 Gricean Charity and Naturalized Interpretation

Like conversational pragmatics, *Gricean charity* adopts conversational norms as norms of rationality, and uses these norms to guide the interpretation of subjects' beliefs. Gricean charity invokes these norms of conversation to justify alternate interpretations of

---

[251] See, for example, Ginossar and Trope, "Problem Solving in Judgment under Uncertainty," experiment 5. Also, Eleanor Singer, Hans-Jurgen Hippler, and Norbert Schwarz discovered that increasingly emphatic confidentiality assurances decreases the rate at which subjects are willing to respond to survey questionnaires. Their explanation is that such confidentiality assurances – when construed as relevant to the ensuing survey questions – suggest that the survey will ask questions that are personal, embarrassing, and/or incriminating. Subjects' decreased willingness to participate, and decreased willingness to provide identifying information in order to participate in future surveys, seems to indicate a decreased level of trust in the surveyors' confidentiality assurances. Eleanor Singer, Hans-Jurgen Hippler, and Norbert Schwarz, "Confidentiality Assurances in Surveys: Reassurance or Threat?," *International Journal of Public Opinion Research* 4, no. 3 (1992).

[252] This literature attributes to subjects a broader range of possible goals, including: the attainment of private rewards, the discovery of the experiment's true rationale, the presentation of the self in the best light, and the desire to contribute to the experiment's success by assisting the experimenter in proving her point. For a review, see Arie W. Kruglanski, "The Human Subject in the Psychology Experiment: Fact and Artifact," *Advances in Experimental Social Psychology* 8 (1975).

the stated question or task – an interpretation for which subjects' expressed beliefs can sometimes be construed as rational. Implicit in this approach to intentional explanation is the empirical assumption that subjects are very unlikely to violate naturalized conversational norms. Gricean charity's difference with conversational pragmatics is merely one of emphasis: Gricean charity explicitly embraces naturalized norms of conversation, and recommends the continued collection of evidence about the conversational assumptions and norms guiding subject-experimenter communication.

This naturalized approach to interpretation builds on important themes in naturalized accounts of interpretation by contemporary philosophers such as David Henderson and Mark Risjord. Like these accounts, Gricean charity is naturalized in the sense that it allows and encourages the use of empirical knowledge taken from the human sciences to guide interpretive theory. Henderson argues that empirical knowledge, especially psychological theory, should be the primary guide in interpretation. Risjord's interest in interpreting group-level events in terms of cultural norms expands the list of human sciences relevant to interpretation to include anthropological, sociological, and historical theories. Gricean charity's interest in naturalized conversational norms connects psychological research with findings and theories in linguistic anthropology and sociolinguistics.

Henderson's and Risjord's naturalized accounts of interpretation also embrace the Davidsonian lesson that any psychological theory on human judgment "must *include* a theory of interpretation" about subjects' beliefs about the experimental task.[253] Henderson paints a picture of "interpretation-*cum*-explanations," where we "construct interpretive schemes so as to be yoked with our psychological and sociological theories to the end of modeling and accounting for the behavior and behavioral dispositions of our subjects.[254] Under Risjord's account, appeals to meanings (common to a group of speakers) are crucial to the interpretation of group-level phenomena.[255] Gricean charity's contribution to this common ground is in providing a positive account of the kind of "integral role" that interpretation should play in the co-development of interpretive and

[253] Davidson, "Belief and the Basis of Meaning (1974)," 147. The italics are mine.
[254] Henderson, *Interpretation and Explanation in the Human Sciences*, 73-4.
[255] Mark Risjord, *Woodcutters and Witchcraft* (Albany, NY: State University of New York Press, 2000), 137-8.

psychological theory. In particular, Gricean charity draws important methodological lessons from current research on conversational pragmatics and recommends a naturalized account of conversational norms that respects the cultural relativity of conversational norms, and their dynamic relationship with questionnaire content and communicator identity.

3.6 Objections

3.61 The Charge of Universal Rationality

An important test of any account of charitable interpretation is whether it can allow for and even prefer interpretations that describe others as being systematically irrational. Some might object that Gricean charity fails this test, since it can be enlisted to rationalize just about any case of apparent irrationality. This possibility stems from a particular step in the inductive argument for any conversational implicature. In drawing an implicature, a hearer must arrive at the belief that, in order for some speaker *S* to say that *p*, and still be observing the maxims or the co-operative principle, *S* must think that *q*. This premise about *q*, is itself arrived at by means of an inductive argument about what claim *q* is implicated, given the literal meaning of *p*, the conversational context, and beliefs mutually held by speaker and addressee.[256] Without any principled way of deciding what gets to count as a reasonable or best candidate implicature *q*, acceptable conversational implicatures are restricted only by the limitations of the human imagination. Since we can always find some reinterpretation for which subject responses can be said to be rational, Gricean charity always provides a way to rationalize subject responses.

This worry overlooks an important lesson of the Gricean turn, namely, the role of evidence. Gricean charity puts the onus on the researcher to identify subjects' interpretations of experimental tasks and to identify conditions of successful communication. This evidence can sometimes undermine experimenters' claims that

[256] Levinson, "Conversational Implicature," 113-4.

subjects do in fact draw particular conversational inferences. Additionally, such evidence can speak in favor of describing subjects as engaging in irrational lines of reasoning.

For example, *open-ended protocols* ask subjects to describe their interpretations of the experimental task and/or explain their answers, in their own words. This method is helpful because it allows experimenters to capture the cognitive processes associated with specific semantic inferences, problem-reasoning, and judgment.[257] Using this method, Don Dulany and Denis Hilton asked subjects "What did 'Linda is a bank teller' *mean* or *imply* to you? Be as clear as you can." They found that many subjects indicated that Linda's being a bank teller was irrelevant to answering the question. Their reasoning seemed to go as follows: because Linda's being a bank teller was a property shared between the two outcomes, it was "a constant that cancelled," and was not supposed to affect the outcomes, thus reducing the problem to judging whether Linda was a feminist or not.[258] This line of reasoning is very problematic. Events common to different outcomes do not cancel out in the way subjects imagine. This additional self-reported information about subjects' beliefs and inferences suggests describing their judgments as profoundly mistaken.[259]

Before my response to this objection, it seemed that Gricean charity's *modus operandi* was to rationalize what seemed to be irrational responses in famous studies on human judgment. However, some studies do seem to demonstrate some kind of systematic irrationality in human reasoning. This is a great strength of Gricean charity – that it recommends the collection of evidence that may speak in favor of the rationality *or* irrationality of human reasoning under various conditions. It is Gricean charity's deference to evidence about communication in experimental contexts that prevents the hyper-rationalization of subject responses.

---

[257] How experimenters ought to use information from open-ended protocols is not clear cut, since subjects do not always have direct access or insight into their conversational inferences. Psychologists are generally skeptical about whether subjects have insight about their inferences or about the factors that do and do not influence their judgments/choices. For the classic paper on this, see Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Models," *Psychological Review* 84, no. 3 (1977).

[258] Dulany and Hilton, "Conversational Implicature, Conscious Representation, and the Conjunction Fallacy," 102.

[259] This self-reported information also suggests that subject responses cannot be rationalized, even if they interpret "probable" to mean "plausible:" even if subjects are concerned with the relative plausibility of outcomes, the outcomes *still* do not cancel in the way they believe they do. Thanks to James Joyce for this point.

The symmetrical treatment of rational and irrational beliefs can also be found in Henderson's and Risjord's accounts. Under Henderson's account, *both* rational and irrational belief are held to the same standard of explicability, which seeks to explain beliefs and actions in terms of the subject's causally relevant intentional states and psychological dispositions. In cases where we happen to adopt a "rationalizing explanation," the explanatory force of this explanation derives not from the rationality of what the subject believes, but from citing causal antecedents and dispositions in terms that feature in psychological generalizations. Gricean charity provides an example of what Henderson would identify as a kind of "modest" rationalizing explanation: Gricean charity hypothesizes that subjects are likely to arrive at a particular interpretation of an experimental task, in virtue of their disposition to conform to conversational principles (where claims about their dispositions and about the conversational principles are empirical claims).[260] The force of Gricean charity's explanations draws strength from empirical evidence about subjects' interpretations and their conversational assumptions, goals, and norms.

For Risjord, theories that attribute rational rather than irrational beliefs are held to the same standard of explanatory coherence – a standard that is not committed to interpreting all beliefs and actions as rational.[261] In contrast to Henderson, however, Risjord recognizes that norms of rationality have a legitimate place in intentional explanation. Interpreters bring with them "interests constitutive to the interpretive enterprise" such as interests in the agents' point of view and in the structure of the society in which they live.[262] Risjord observes that our interest in these perspectives requires explanations invoking norms.[263] This is because intentional explanations and group-level explanations invoke reasons, where – for Risjord – reasons may count as reasons only insofar as they conform to norms recognized by the agent or the society.[264] It is in this indirect way that norms figure in the content of explanations.[265] Gricean charity draws strength from Risjord's analysis. As an account of interpretation, Gricean charity has an

[260] Henderson, *Interpretation and Explanation in the Human Sciences*, 135-6.
[261] Risjord, *Woodcutters and Witchcraft*, 182.
[262] Ibid., 177.
[263] Ibid., 187-8.
[264] Not all reasons are like this. We may have instrumental reasons that do not have anything to do with norms.
[265] Ibid., 155.

interest in understanding communication and psychological experimentation from the subject's point of view. As such, Gricean charity is interested in the reasons subjects have for their interpretations of the task, where these reasons count as reasons insofar as they relate to conversational norms governing subject-experimenter communication. These conversational norms figure in the content of Gricean charity's intentional explanations.

3.62 Biased Applications of Gricean Charity

Keith Stanovich and Richard West observe that some charitable strategies seeking to rationalize subject responses function by reacting to findings of the heuristics and biases research approach: they aim to restore the rationality of subject responses in the face of research purporting to demonstrate otherwise. What is suspicious about this pattern of theorizing is that it criticizes studies purporting to demonstrate irrationality, but rarely – if ever – critiques those where modal subject response *coincides* with the normative response.[266] The worry is that charitable researchers in psychology are biased insofar as they hold higher standards of experimental design and evidence for psychological theories claiming to demonstrate human irrationality.

However, Gricean charity does not require a higher standard of evidence for theories claiming that human reasoning is irrational in some way. Rather, what speaks for the strength of Gricean charity is that it recommends an evidential standard that applies generally to studies on human judgment, irrespective of their conclusions about the rationality or irrationality of subject responses. The interviewing methods used for attaining key interpretive evidence is also shared across the rationality divide. I will discuss such interviewing techniques in greater detail in subsequent sections of this paper.

3.63 How to Test the Effects of Irrelevant Information

[266] Stanovich and West make this observation in the context of the "reject-the-norm" strategy, which rejects the experimenter's normative theory for a different one to which modal subject response conforms. Keith E. Stanovich and Richard F. West, "Individual Differences in Reasoning: Implications for the Rationality Debate?," *The Behavioral and Brain Sciences* 23 (2000): 650.

Some object that conversational norms overly constrain what it is that cognitive psychologists can test since these norms may circumvent asking questions in ways that prove psychologically interesting. In particular, Kahneman and Tversky expressed the concern that Grice's maxim of relevance poses an "exceptionally difficult" problem for experimenters interested in studying the effects of irrelevant information" on cognition.[267] Citing Richard Nisbett et. al.'s work, they observe that subjects can mistakenly construe nearly *any* piece of information as relevant.[268] From a methodological point of view, the worry is that the maxim of relevance's constraint on questionnaire and survey design precludes the very possibility of testing the influence of irrelevant information on human cognition.

However, there are ways in which to test the effect of irrelevant information on cognition, *without having to violate conversational norms*. Psychologists working on conversational pragmatics have forged ingenious experimental methods to do this. One way is to undermine subjects' assumption that the source of information in the experimental context is intentional and cooperative. Recall that the conversational maxims apply in the special case where information-exchange occurs between *intentional, cooperative communicators* in ordinary language. If experimenters can undermine this key assumption, then all conversational implicatures should be cancelled, which would allow experimenters to study the effects of irrelevant information on human cognition.

Norbert Schwarz and his colleagues (Fritz Strack, Denis Hilton, and Gabi Naderer) discovered an ingenious way of using a computer interface to do just this. They focused on Kahneman and Tversky's famous lawyer-engineer question:

> A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.

---

[267] Kahneman and Tversky, "On the Study of Statistical Intuitions," 501-2.
[268] Richard E. Nisbett, Henry Zukier, and Ronald E. Lemley, "The Dilution Effect: Nondiagnostic Information Weakens the Implications of Diagnostic Information," *Cognitive Psychology* 13 (1981).

> The same task has been performed by a panel of experts, who were highly accurate in assigning probabilities to the various descriptions. You will be paid a bonus to the extent that your estimates come close to those of the expert panel.[269]

In Kahneman and Tversky's original study, the *low-engineer group* was told that there were 30 engineers and 70 lawyers. The *high-engineer group* was told that there were 70 engineers and 30 lawyers. Both groups were provided the same five personality descriptions, most of which were stereotypical of an engineer or lawyer.[270] Kahneman and Tversky found that subjects' predictions about how probable it was that a given person was an engineer or lawyer were independent of the base rates of engineers/lawyers described in the questionnaire: subjects ignored the base rate information in violation of Bayes' Rule.

Norbert Schwarz et. al. predicted that subjects' violation of Bayes' Rule resulted from conversational implicatures reasonably inferred from the original wording of the questionnaire. They point out that subjects who identify the experimenter as cooperative are in the position of trying to render the communicated information about Jack's personality relevant to their interpretation of the experimental task; and, by the maxim of quantity, subjects are left to infer that *all* the detailed information provided about Jack's personality are meant to play into the proper solution of the task.[271]

---

[269] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 53.

[270] Kahneman and Tversky offered the following as an example of one of the personality descriptions: "Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

The probability that Jack is one of the 30 engineers in the sample of 100 is ------%."

[271] Special clues in the question underscore the relevance and importance of the personality description in solving the task. The first paragraph of the instructions "informs subjects that the individuating information was compiled by psychologists on the basis of respected procedures of their profession, namely interviews and tests." Schwarz et. al. observe that, since psychologists are stereotypically perceived as experts on issues of personality rather than probability and base rates, identifying the authors of the personality descriptions as psychologists emphasizes the relevance and informativeness of the individuating information rather than the base rate information in solving the experimental task. Kahneman and Tversky reinforce the importance and relevance of the personality descriptions by going on to state that "[t]he same task has been performed by a panel of experts, who were highly accurate in assigning probabilities to the various descriptions." This sentence underscores the relevance of the individuating information by pointing out that the stereotypical descriptions are sufficiently diagnostic for experts to succeed in solving the experimental task. Although the professional identity of the experts is left unspecified, subjects might reasonably infer the experts are psychologists, based on the following facts: the experts are highly accurate in personality-based predictions, and the experts are so-called by experimental psychologists in the context where they seem to use the personality tests to predict outcomes. The further claim that "[y]ou will be paid a bonus to the extent that your estimates come close to those of the expert panel" suggests that the subjects are encouraged to make judgments in a *similar manner* as the expert panel. If subjects have already identified the experts as psychologists, this statement would encourage subjects to study the personality

To undermine the assumption that the individuating information was relevant and informative, Schwarz et. al. ingeniously created a *Computer Communicator* condition, where subjects were told that a computer – an uncooperative and unintentional communicator – had created the personality description by randomly drawing sentences from psychologists' or researchers' files pertaining to the target person.[272] By undermining the assumption of cooperative communication, subjects were freed from having to construe the personality information as relevant to the experimental task. They found that in the computer-communicator condition, subjects weighed the individuating information less, and considered the base rate more: the mean probability estimate for subjects judging the probability that the target was an engineer was only 40%, compared to the control group's mean probability estimate of 76%.

3.64 Conversational Clarity and Conceptually Difficult Tasks

Kahneman and Tversky have objected that efforts to make the experimental task as semantically unambiguous as possible reveals key clues about solving the task. Such clues, the objectors maintain, compromise researchers' abilities to test whether subjects can solve the task *without* undue help. The best way to understand this objection perhaps is by way of example. In the Linda case one way to clear up the ambiguity about what "probable" means, is to paraphrase with the more precise mathematical word "frequency" in the following way:[273]

> In an opinion poll, the 200 women selected to participate have the following features in common: They are, on average, 30 years old, single, and very bright. They majored in philosophy. As students, they were deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.
>
> Please estimate the frequency of the following events.

---

traits described, and use them to diagnose professional identity. Norbert Schwarz et al., "Base Rates, Representativeness, and the Logic of Conversation: The Contextual Relevance of "Irrelevant" Information," *Social Cognition* 9, no. 1 (1991).

[272] To undermine the assumption that the individuating information was relevant and informative, Schwarz and his colleagues told subjects that a computer had created the provided personality description by randomly drawing sentences from psychologists' or researchers' files pertaining to the target person. Ibid.: 74.

[273] This idea is due to Hertwig and Gigerenzer, "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors."

How many of the 200 women are bank tellers? ___ of 200
How many of the 200 women are active feminists? ___ of 200
How many of the 200 women are bank tellers and active feminists? ___ of 200[274]

As the researchers Ralph Hertwig and Gerd Gigerenzer expected, none of the subjects provided answers in violation of the conjunction rule under this formulation of the Linda problem.

However, Kahneman and Tversky suggest this formulation of the question makes judgments of probability a piece of cake. They ask subjects for a numerical estimate of the ratio of women who are bank tellers and/or active feminists in the total population of Linda-like people. That is, they ask subjects to estimate the number of people belonging to: the set of bank tellers, the set of active feminists, and the intersection of both these sets, within a fixed population. Being able to judge the probability of a conjunction of events requires conceptualizing how classes of events are related in this set theoretical way. At the very least, understanding the conjunction rule requires seeing that the number of members in the *intersection* of two sets, must be *less than* or equal to the number of members in each of those individual sets. Hertwig and Gigerenzer's way of posing the question not only clarifies the meaning of the question, it takes subjects through the hardest step of understanding how to think about and solve the task.[275] So the question becomes: are there ways of clarifying the meaning of the Linda question *without* giving away what Kahneman and Tversky take to be key clues?[276]

---

[274] Ibid.: 291.

[275] This finding replicates Kahneman and Tversky's discovery that subjects are much less likely to violate the conjunction rule when the conjunctions were represented by the intersection of concrete, finite classes, than by an abstract combination of properties. They found that only 25% of subjects violated the conjunction fallacy when they were asked to estimate the frequencies rather than single-event probabilities. But, by representing the question in this frequency format, Tversky and Kahneman took themselves as encouraging "subjects to set up a representation of the problems in which class inclusion is readily perceived and appreciated." Tversky and Kahneman take their findings to demonstrate that "[t]he formal equivalence of properties to classes is apparently not programmed into the lay mind" – that subjects do not "evaluate compound probabilities by aggregating elementary ones" – unless provided a representation "in which different relations and rules are transparent." Tversky and Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," 294-309.

[276] Notice that Kahneman and Tversky's question and objection presupposes no obligation to satisfy Gricean norms. From a Gricean perspective, the lesson to be drawn is not that "people are irrational," but that "if you do not want to mislead people, then speak more clearly." An experiment without clearly specified questions does not demonstrate that subjects are innately irrational, but that certain representations are less effective at communicating a task than others.

Hertwig and Gigerenzer discovered a way of clarifying the intended meaning of the probability question, without giving away key clues to solving the problem. They observed that, in the original question, the meaning of "probability" was obscured by the maxim of relevance: in particular, the maxim directs subjects to make irrelevant personality information pertinent to the experimental task. They hypothesized that one way to indicate that the personality information was not necessarily relevant to the probability task, is to ask a different question for which Linda's personality description *is* relevant, *before* asking the probability question. This leaves subjects free to interpret the meaning of the subsequent "probability" question without reference to the personality description. To test this, Hertwig and Gigerenzer provided subjects with Linda's personality description, and asked subjects to judge *how good an example* Linda's personality was of "an active feminist," "a bank teller," and "a bank teller and an active feminist." Subjects were then asked the probability question, "Which of the following statements is most probable?": "Linda is an active feminist," "Linda is a bank teller," and "Linda is an active feminist and bankteller." Under this version of the Linda problem, they found that subjects were less likely to violate the conjunction rule: only 59% of subjects violated the conjunction rule, compared to 88%. Notice that the effect of successful communication here is not to manipulate subjects into providing the normatively correct answer, since the majority of subjects continue to violate the conjunction rule. Rather, the effect is to get a more accurate picture of subjects' probability judgments.

3.7 Conclusion

Gricean charity builds on recent accounts of naturalized interpretation by issuing in positive methodological recommendations for proper interpretation and explanation in psychological experimentation. In particular, the Gricean turn reminds us that researchers are responsible for successful communication in the experimental context. In order to promote successful communication, Gricean charity recommends creating questionnaires in conformance to naturalized conversational norms, and recommends

seeking evidence about (i) the conditions of successful communication and (ii) the conversational norms governing subject-experimenter communication. The Gricean turn also shifts psychological explanations toward situational explanations that focus on the experimental conditions responsible for subject responses. And, this situational perspective engages in reflexive analysis about the influence of the experimenter (and experimenter's relationship with the subject) on observed behavior.

Gricean charity's evidential recommendations are unbiased in the sense that they apply symmetrically across the rationality divide: they apply to studies on human judgment, irrespective of the studies' conclusions about the rationality or irrationality of subjects' responses. Evidence gained by these methods does not necessarily speak in favor of the rationality or irrationality of subject responses. Thus, Gricean charity passes an important test for any account of charitable interpretation: namely, that it allows for the possibility of and even recommends interpretations that construe others as being systematically irrational. As a result, Gricean charity does not illegitimately favor psychological theories that construe human reasoning as being generally rational.

Gricean charity *is* biased in its special focus on conversational pragmatics. However, this is not a problematic bias, but a focus of research. Such a focus is found in normal scientific research programs that use background theory and hard-core assumptions as tools to generate new questions, predictions, evidence, and hypotheses.[277] Gricean Charity suggests new questions about the nature of conversational inference: Under what conditions can we undermine subjects' assumption of the experimenter's cooperativeness? How can we undermine the assumption of intentionality in questionnaires, surveys, or experimental communications in general? How can we emphasize the right information so as to elicit valid ways of reasoning and correct judgment? How is task interpretation tied into reasoning about how to solve the task? And how do conversational norms vary under different experimental conditions?

These questions motivate the creation of new methods for collecting evidence about subjects' conversational inferences. For example, Schwarz et. al.'s computer communication condition allowed researchers to undermine subjects' assumption of

[277] Philip E. Tetlock, "The Impact of Accountability on Judgment and Choice: Toward a Social Contingency Model," *Advances in Experimental Social Psychology* 25 (1992). Gigerenzer, "From Tools to Theories: A Heuristic Discovery in Cognitive Psychology."

cooperativeness and intentionality on the part of experimenters. Hertwig and Gigerenzer's subtle double-question technique provided a way of manipulating the assumption of relevance.[278]

Finally, the questions, hypotheses, and interests motivated by Gricean charity, have lead to the discovery of new kinds of psychologically interesting phenomena. For example, Schwarz's computer communicator condition revealed new conditions that sensitize subject probability judgments to base rate information. His computer communicator condition also revealed other fascinating evidence. When the wording of the questionnaire was framed as a statistics problem, the computer communicator condition *impaired* subjects' probability judgments: subjects again ignored or underweighted the base rate information.[279] Schwarz et. al. suggest that subjects relied more on the individuating information "presumably because in a statistical framework random sampling suggests that the resulting selection is representative of the population of descriptive information from which it is drawn."[280] By communicating more successfully with subjects, researchers in conversational pragmatics have opened the door to getting a clearer picture of the surprising and muddled ways in which we weigh evidence, draw inferences, and make choices.

---

[278] In addition to these methods of checking subject task construal, experimenters may employ a *multiple-choice check* method, which asks subjects to check which of a provided set of interpretations best matches their interpretation of the task. One weakness of this approach is that subjects may draw different implicatures from the more precisely stated multiple-choice options. For examples relevant to the conjunction fallacy, see Hilton, "The Social Context of Reasoning: Conversational Inference and Rational Judgment," 260. Experimenters may also check subject interpretations by asking subjects to *paraphrase* the experimenter's original question. One problem with this kind of method of checking subject interpretations is that the paraphrases provided may themselves be vague, polysemous, and/or indeterminate. For examples relevant to the conjunction fallacy, see Hertwig and Gigerenzer, "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors."

[279] Schwarz et al., "Base Rates, Representativeness, and the Logic of Conversation: The Contextual Relevance of "Irrelevant" Information," 72-3.

[280] Ibid.: 76.

Chapter 4

Naturalized Epistemology Rationalized

The connections between naturalized epistemology and psychology are quite different from the kind of continuity that naturalized epistemologists have traditionally acknowledged. Willard Van Orman Quine, for example, famously claims that we are prompted to study "the relation between the meager input and the torrential output" of human behavior, "in order to see how evidence relates to theory."[281] In this and other passages, Quine seems to propose replacing talk of the normative evidential relation with the naturalistic, causal talk of cognitive psychologists. It is because of claims like these that Quine is accused of advocating the strong replacement of epistemology by psychology.[282] The *strong replacement* thesis claims that evidential relationships can be reduced to descriptive psychological claims without any loss of content.[283]

I will argue that the strong replacement thesis and the critiques launched against it are misguided for different reasons than those typically offered.[284] The most common objection against the strong replacement thesis is that it eliminates the normative role of justification and verification by replacing it with purely descriptive, causal-nomological

---

[281] W. V. O. Quine, "Epistemology Naturalized," in *Naturalizing Epistemology*, ed. Hilary Kornblith (Cambridge, MA: The MIT Press, 1997), 25.

[282] This way of construing Quine's position is inaccurate. In *Pursuit of Truth*, Quine distinguishes mere *descriptive* science from *theoretical* science, the latter of which reflects science's more general concern about the nature of evidence, verification, and justification. It is clear that Quine does not argue for a strong replacement of epistemology by *descriptive* science, as the strong replacement theory would require, but a replacement by *theoretical* science, which includes work from the philosophy of science, sociology, psychology, linguistics, and the hard sciences. I will say more on this later in the chapter. See W. V. O. Quine, "Evidence," in *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990), 19.

[283] Hilary Kornblith, "Introduction: What Is Naturalistic Epistemology?," in *Naturalizing Epistemology*, ed. Hilary Kornblith (Cambridge, MA: The MIT Press, 1997), 3-7.

[284] See, for example, Jaegwon Kim, "What Is "Naturalized Epistemology"?," in *Epistemology: An Anthology*, ed. Ernest Sosa and Jaegwon Kim (Oxford, U.K.: Blackwell Publishers, Ltd., 1997).

input-output patterns constituting reliably formed "justified" belief.[285] However, this critique and the strong replacement thesis wrongly assume that psychological "facts" and "questions" are not themselves steeped in epistemic concepts and normative discourse – an assumption that is deeply mistaken in light of research and changing methodological standards in contemporary psychology.

In the first part of this chapter, I will argue that some of the theories arising from the literature on rational human judgment suggest that cognitive psychology and epistemology are continuous in the sense that our normative theories of good inferential practice guide and even figure in the actual *content* of psychological claims. Since the cognitive revolution of the mid-1900s, cognitive psychologists have used theories of inference and rationality to model the cognitive processes linking evidence to belief: that is, "psychological facts" and "psychological theory" have been and continue to be couched in terms of and guided by our normative discourse. I will refer to Gerd Gigerenzer's research to demonstrate how the normative discourse of epistemology informs psychology's interests, questions, and claims about human judgment.

In the second part of the chapter, I will discuss some of the methodological standards Gigerenzer seems to embrace in his critiques of the heuristics and biases research program. In particular, I will focus on ways in which he seems to think cognitive processes should be specified for the sake of explaining human judgment. These methodological critiques suggest that cognitive psychology and naturalized epistemology are disciplines with shared explanatory goals: in particular, both invoke cognitive processes to explain the psychological transformation of inputs to output-beliefs; and both seek to explain the epistemic status of output-beliefs by reference to the same cognitive process invoked to explain its production.

In the third part of this chapter, I will make a few observations on how the shared explanatory goals between cognitive psychology and naturalized epistemology recasts traditional challenges facing reliabilism. I will revisit questions about whether reliabilism should or can adopt epistemic naturalism and the threat of reducing process reliabilism to a parasitic account of pseudo-reliability. I will also discuss how the problem of

[285] Ibid., 305.

generality presents itself in psychological theorizing. My main exegetical goal is not to provide definitive solutions to these issues but to revisit them in light of the kind of continuity I have argued for between naturalized epistemology and psychology.

4.1 Building Epistemic Norms into Cognitive Processes

I argued previously that we should understand ecological rationalism as a preference for discovering the conditions that promote rational judgment. *Prima facie*, this might seem to suggest that psychologists might seek situational explanations for subject responses – explanations that primarily invoke experimental conditions rather than characteristics of the subjects themselves – in explaining human judgment. However, Gigerenzer's psychological hypotheses often involve claims about the psychological processes by which subjects arrive at their judgments, and the role of circumstances in triggering different psychological processes.[286]

In this section, I will argue that Gigerenzer uses statistical models to inform the content of the cognitive algorithms and processes invoked to explain human judgment. Gigerenzer ultimately embraces these statistical models as epistemic norms. For this, I will look to the fast and frugal heuristics research program's recognition heuristic.

4.11 The Recognition Heuristic Suite

The *fast and frugal heuristics* research program founded by Gigerenzer, Peter Todd, and the ABC Research Group seeks to discover heuristics that are *ecologically rational* – that is, cognitive strategies that exploit the information occurring in natural environments to support a disproportionately high frequency of true or normatively correct beliefs (in relation to the total frequency of correct/true and false/incorrect beliefs)

[286] Thanks to Peter Railton for this way of stating the contrast.

for a given reference class.[287] These ecologically rational heuristics provide models of bounded rationality: they identify rational strategies of reasoning for creatures with limited information and computational capacities faced with particular kinds of environments.[288] The *recognition heuristic*, "the most frugal" of all, makes inferences from patterns of missing knowledge.[289] This heuristic is particularly frugal because the search for information extends only to recognition: search is stopped whenever one object is recognized and the others are not. The recognition heuristic simply directs one to choose the recognized object.

For simplicity's sake, Goldstein and Gigerenzer test the recognition heuristic on two-alternative forced choice tasks, where one is asked to choose which of two provided options taken from a reference class (such as the set of German cities) has a higher value on some criterion (such as population size). In two-alternative forced choice tasks, the basic modus operandi of the recognition heuristic may be formulated in this way: "If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion."[290]

The ecological validity of the recognition heuristic depends in part on the strength of the correlation between recognition and the criterion of interest. This correlation is accounted for by the *recognition validity* α. The recognition validity α can defined as follows:

$$\alpha = \frac{R}{(R+W)}$$

where R is the number of correct inferences the recognition heuristic would achieve, computed across all pairs in which one object is recognized and the other is not, and W is

---

[287] Gigerenzer explicitly defines the ecological validity of a cue as a "true relative frequency" Gigerenzer, Hoffrage, and Kleinbolting, "Probabilistic Mental Models: A Brunswikian Theory of Confidence," 508. For example, the validity of the recognition heuristic is measured by the frequency with which the heuristic would lead to a correct inference, divided by the frequency with which the heuristic would or actually lead to a correct inference or an incorrect inference. Daniel G. Goldstein and Gerd Gigerenzer, "Models of Ecological Rationality: The Recognition Heuristic," *Psychological Review* 109, no. 1 (2002). For the canonical text for the fast and frugal heuristics research program, see Gerd Gigerenzer, Peter M. Todd, and ABC Research Group, *Simple Heuristics That Make Us Smart* (Oxford, UK: Oxford University Press, 1999).

[288] Simon, "A Behavioral Model of Rational Choice.", Simon, "Rational Choice and the Structure of the Environment."

[289] Goldstein and Gigerenzer, "Models of Ecological Rationality: The Recognition Heuristic," 75.

[290] Ibid.: 76.

the number of incorrect inferences under the same circumstances.[291] When α is positive, the heuristic leads to the inference that the unrecognized object has the lesser criterion value (e.g., that the unrecognized city has a smaller population than the recognized city). When α is negative, the heuristic leads to the inference that the unrecognized object has the higher criterion value (e.g., that the unrecognized city has a larger population than the recognized city).[292] For reference classes where α is no better than chance, the recognition heuristic cannot be said to be ecologically valid.

In two-alternative forced choice tasks, we can only make use of the recognition heuristic in cases where we recognize one object but not the other. However, not all two-alternative forced choice tasks involve one recognized and one unrecognized object. In drawing pairs of objects from a reference class of N objects, there are three ways the pairs can turn out: one recognized and one unrecognized; both unrecognized; or, both recognized. Let's say there are n recognized objects and thus $N - n$ unrecognized objects. This means that there are:

$n(N-n)$ pairs where one object is recognized and the other is unrecognized;

$\frac{(N-n)(N-n-1)}{2}$ pairs in which neither object is recognized; and,

$\frac{n(n-1)}{2}$ pairs where both objects are recognized.

To transform these absolute numbers into proportions in an exhaustive test of all possible pairs, it is necessary to divide each by the total number of possible pairs $\frac{N(N-1)}{2}$.

We can calculate the expected proportion of correct inferences in all three cases: for pairs where one object is recognized and one not; for pairs where both are unrecognized; and for pairs where both are recognized. When one object is recognized and the other is not, the probability that one arrives at a correct answer depends on the recognition validity α. When neither object is recognized, a guess must be made, and the probability of getting a correct answer is at the level of chance or 1/2. When both objects

[291] Ibid.: 78.
[292] Ibid.: 76.

are recognized, the probability of getting a correct answer depends on the *knowledge validity* β.

For two-alternative forced choice tasks with reference class N, where n objects are recognized, the overall expected proportion of correct inferences relative to the proportion of correct and incorrect inferences $f(n)$ can be calculated as follows:

$$f(\mathrm{n}) = 2\left(\frac{n}{N}\right)\left(\frac{N-n}{N-1}\right)\alpha + \left(\frac{N-n}{N}\right)\left(\frac{N-n-1}{N-1}\right)\frac{1}{2} + \left(\frac{n}{N}\right)\left(\frac{n-1}{N-1}\right)\beta$$

The leftmost term on the right side of the equation is the proportion of correct inferences made by the recognition heuristic (with recognition validity α) when one object is recognized and the other is not. The middle term refers to the proportion of correct inferences resulting from guessing when both objects are unrecognized at the rate of chance. And, the rightmost term refers to the proportion of correct knowledge-mediated inferences (with knowledge validity β) when both objects are recognized.[293]

So, for the reference class N of which n objects are recognized, we have three distinct strategies of inference for three different conditions of ignorance: when we recognize one object but not the other, we rely on the recognition heuristic; when we don't recognize either object, we make a guess; and, when we recognize both, we use a knowledge-mediated inference. I will refer to the totality of these distinct strategies the *recognition heuristic suite*. The recognition heuristic suite is a kind of mixed strategy in which different pure strategies are played with distinct probabilities and distinct outcomes. $f(\mathrm{n})$ measures the reliability of the total recognition heuristic suite's mixed strategy. The pure strategies include using the recognition heuristic in cases where one recognizes one item but not the other (with recognition validity α), using knowledge in cases where one recognizes both items (with knowledge validity β), and taking a guess in cases where one recognizes none of the items (at the level of chance).

Notice that $f(\mathrm{n})$ provides the same kind of statistical measure of reliability that Goldman proposes: Goldman suggests that "(as a first approximation) reliability consists in the tendency of a process to produce beliefs that are true rather than false."[294] The

---

[293] Ibid.: 78.

[294] Alvin I. Goldman, "What Is Justified Belief?," in *Epistemology: An Anthology*, ed. Hilary Kornblith (Oxford, U.K.: Blackwell Publishers Ltd., 2000), 345.

conditions in which the representativeness heuristic suite is ecologically valid are the same conditions in which the representativeness heuristic suite is a reliable strategy or process of belief-formation. In addition, notice that the recognition validity α also provides a statistical measure of reliability Goldman might condone: the recognition validity is measured by the number of correct inferences the recognition heuristic would achieve, divided by the total number of correct and incorrect inferences, computed across all cases in which one item is recognized while the other is not.

Notice that the statistical model of the recognition heuristic serves to explain why it is that the recognition heuristic suite or recognition heuristic (as a pure strategy) is valid for certain kinds of problems and contexts. This model of the recognition heuristic allows us to mathematically determine the conditions that raise the ecological validity $f(n)$ of the recognition heuristic suite: a high recognition validity α increases the total ecological validity, as does being as close as possible to half ignorance with respect to the recognized/unrecognized objects. The ecological validity of the recognition heuristic suite is built into and thus explained by the cognitive process as specified by the cognitive model.

4.12 Heuristics and Norms with Built-In Conditions of Validity

The recognition heuristic suite suggests a general methodological principle Gigerenzer seems to adopt: namely that, we should be able to explain the validity of a heuristic or cognitive process by looking to how that cognitive process is specified and/or defined:

> GM1. The specification of a cognitive process should capture or explain the conditions of its own validity.

For the recognition heuristic suite, the conditions of the cognitive processes' validity are built-in. The recognition heuristic suite allows us to model situations in which the ecological validity or reliability increases or decreases depending on the number of items in the reference class, the number of items recognized, the recognition validity α, and the knowledge validity β. In this sense, the conditions in which the recognition heuristic suite is valid or reliable are built-into the specification of the model.

It is not an accident that these strategies or heuristics have built into them the conditions of their own validity. Gigerenzer is interested in discovering the conditions promoting rational judgment and the cognitive processes underlying those rational judgments. Those judgments cannot be said to be rational unless they are sustained or produced by functional procedures that instantiate epistemic norms. For Gigerenzer, epistemic norms differ from a priori rules in the sense that epistemic norms are "*constructed* for a specific situation, not *imposed* upon it in a content-blind way:" unlike a priori rules, epistemic norms do not "disregard relevant structural properties of the given situation" such as the available information, the information structure (the ways in which the available cues are correlated), and the salient reference classes.[295]

> GM2. A cognitive process must be justified for specific problem contents, contexts, and information formats in order to count as an epistemic norm.

Epistemic norms are "constructed and justified for" different contexts and contents.[296] So, when Gigerenzer specifies cognitive processes that capture the conditions of their own validity, he is thereby justifying those cognitive processes as epistemic norms in the contexts and conditions in which they are valid.

It is clear that Gigerenzer takes the recognition heuristic suite to be an epistemic norm that confers something like justification on beliefs. The fast and frugal heuristics instantiate functional processes that serve as "normative model[s]" constructed and justified for problems with different contents, contexts, and information formats: "sound normative thinking leads us into the world of bounded rationality, of fast and frugal heuristics, satisficing, and other robust strategies that can do surprisingly well when used in the appropriate situation."[297]

---

[295] Gerd Gigerenzer, "Content-Blind Norms, No Norms, or Good Norms? A Reply to Vranas," *Cognition* 81 (2001): 93.
[296] Ibid.: 94.
[297] Ibid.: 102. Gigerenzer takes the question of what counts as an epistemic norm as a major point of contention between himself and Kahneman and Tversky. As Gigerenzer puts the debate: "The first issue on which Kahneman and Tversky and I disagree concerns the question of what counts as sound statistical reasoning. Most practicing statisticians start by investigating the content of a problem, work out a set of assumptions, and, finally, building a statistical model based on these assumptions. The heuristics-and-biases program starts at the opposite end. A convenient statistical principle, such as the conjunction rule or Bayes's rule, is chosen as normative, and some real-world content is filled in afterward, on the assumption that only structure matters. The content of the problem is not analyzed in building a normative model, nor are the specific assumptions people make about the situation." Gigerenzer, "On Narrow Norms and Vague

4.13 On Psychologism

These cognitive processes speak to a kind of possibility – the possibility that cognitive processes are specified in terms of inferential systems and rules that instantiate epistemic norms. Elliott Sober recognized this possibility in his 1979 account of psychologism which underscored "that the principles of right reason which philosophy seeks to discover are used in the information-processing systems of thinking organisms."[298] This form of psychologism is different from the sort found in contemporary discussion on naturalized epistemology. *Contemporary psychologism* claims that the results of epistemology and psychology are couched at different levels of generality – in particular, psychological results involve a specificity and attention to detail unnecessary in answering epistemology's more general and abstract concerns about the nature of justification.[299] By identifying epistemology as a more abstract inquiry than psychology, contemporary psychologism is supposed to provide epistemology the freedom to issue in general epistemic norms abstracted away from naturalistic details.

Sober's form of *psychologism* seems to recognize that cognitive psychology and epistemology can be couched at similar levels of generality. The recognition heuristic suite is specified in a way that allows us to evaluate the reliability of using that suite: naturalized epistemology's justificatory standard of reliability is embedded in the cognitive processes proposed. The recognition heuristic suite explains why subjects arrive at particular beliefs; and, the fact that the recognition heuristic suite is reliable in particular contexts explains the justificatory status of those beliefs.

Heuristics: A Reply to Kahneman and Tversky (1996)," 592-3. It is not clear that Kahneman and Tversky would necessarily disagree about what should count as an epistemic norm. Kahneman and Tversky's theoretical contribution was to suggest a psychological process by which subjects arrive at posterior probability judgments – a process that is invalid in the sense that it leads to beliefs that violates Bayes' Theorem in predictable ways. This kind of theoretical approach – in which imputed psychological processes are shown to violate some rule or reasoning – does not necessarily take Bayes' Theorem to be an epistemic norm. Rather, it relies on Bayes's Theorem as a standard of correctness that *any* epistemic norm or heuristic procedure must meet in order to be said to be valid.

[298] Elliott Sober, "Psychologism," *Journal for the Theory of Social Behavior* 8, no. 2 (1979): 189.

[299] Kornblith, "Introduction: What Is Naturalistic Epistemology?," 3-7.

4.2 Standards of Explanation in Psychology

Gigerenzer has provided methodological critiques of the heuristics and biases research program that have important connections to naturalized epistemology. To understand these critiques, I will look in particular at Kahneman and Tversky's representativeness heuristic and contrast this with Gigerenzer and Hoffrage's frequency algorithm.

4.21 Kahneman and Tversky's Representativeness Heuristic

In discussing the representativeness heuristic, I will turn to papers that their 1996 review article identifies as showing "that the rankings of outcomes by representativeness and by probability were nearly identical."[300] The first article they cite, "On the Psychology of Prediction," contains the engineer-lawyer study which – at the time – they took to be the "more stringent test of the hypothesis that intuitive predictions are dominated by representativeness and are relatively insensitive to prior probabilities."[301] However, in the 1996 paper "On the Reality of Cognitive Illusions" – a response to Gigerenzer's critiques of their research program – Kahneman and Tversky identified the less famous Tom W. study as "the most direct evidence for the role of representativeness in prediction."[302]

The crucial difference between the designs of these studies is that the lawyer-engineer study divided subjects into groups according to differentiating base-rate information and asked both groups to estimate conditional probabilities without asking them to make judgments of similarity. In contrast, the Tom W. experiment divides subjects differently. The researchers provide two control groups. One control group, called the *similarity* group, was presented with the following personality description:

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate

[300] Kahneman and Tversky, "On the Reality of Cognitive Illusions," 585.
[301] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 53.
[302] Ibid., 585.

> place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by splashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feeling and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.[303]

The similarity group was then asked, "how similar is Tom W. to the typical graduate student in each of the following fields of specialization?"[304] The second column in Table 1 below presents the mean similarity ranks assigned to the various fields.

A second control group consisting of 69 subjects, the *base-rate* group, was asked to rank the relative probabilities of the same outcomes without the personality sketch of Tom W. They were given the following directions:

> Consider all first-year graduate students in the U.S. today. Please write down your best guesses about the percentage of these students who are now enrolled in each of the following nine fields of specialization," where the fields include business administration, computer science, engineering, humanities and education, law, library science, medicine, physical and life sciences, and social science and social work. [305]

The first column in Table 1 presents the mean base rate judged by subjects.

The experimental group, called the *prediction* group, consisted of 114 graduate students in psychology at three major U.S. universities. They were given the same personality description of Tom W., but were instructed to rank probable outcomes:

> The preceding personality sketch of Tom W. was written during Tom's senior year in high school by a psychologist, on the basis of projective tests. Tom W. is currently a graduate student. Please rank the following nine fields of graduate specialization in order of the likelihood that Tom W. is now a graduate student in each of these fields.[306]

The third column in Table 1 presents the means of the ranks assigned to the outcomes by subjects in the prediction group.[307]

Table 1. Estimated base rates of the nine areas of graduate specialization and summary of similarity and prediction data for Tom W.

Graduate specialization area	Mean judged base rate (in %)	Mean similarity rank	Mean likelihood rank
Business Administration	15	3.9	4.3
Computer Science	7	2.1	2.5
Engineering	9	2.9	2.6

[303] Ibid., 49.
[304] Ibid.
[305] Ibid.
[306] Ibid., 50.
[307] Ibid.

Humanities and Education	20	7.2	7.6
Law	9	5.9	5.2
Library Science	3	4.2	4.7
Medicine	8	5.9	5.8
Physical and Life Sciences	12	4.5	4.3
Social Science and Social Work	17	8.2	8.0

Kahneman and Tversky found that over 95% of graduate students judged that Tom W. is more likely to study computer science than humanities or education, *even though* "they were surely aware of the fact that there are many more graduate students in the latter field."[308] Because Kahneman and Tversky found that the correlation between judged likelihood and similarity to be 0.97, and the correlation between judged likelihood and estimated base rate to be -0.65, they conclude that "[e]vidently, judgments of likelihood essentially coincide with judgments of similarity and are quite unlike the estimates of base rates."[309] They take this result as providing "a direct confirmation of the hypothesis that people predict by representativeness, or similarity"[310] rather than by base rates. Kahneman and Tversky impute the representativeness heuristic as the mental process responsible for certain kinds of statistical and probability judgment. From the start, Kahneman and Tversky defined representativeness in terms of similarity.[311]

4.22 Gigerenzer's Methodological Critique: Standards of Explanatory Adequacy

Gigerenzer has criticized Kahneman and Tversky's explanation for the base rate effects. Gigerenzer has aimed his critique at a deeper, methodological issue: the question

[308] Ibid.
[309] Ibid.
[310] Ibid.
[311] The first time that Kahneman and Tversky invoke the notion of representativeness is in their 1971 study on intuitions about random sampling from a larger population: "We submit that people view a sample randomly drawn from a population as highly *representative*, that is, *similar* to the population in all essential characteristics. Consequently, they expect any two samples drawn from a particular population to be more similar to one another and to the population than sampling theory predicts, at least for small samples." The article is reproduced in Amos Tversky and Daniel Kahneman, "Belief in the Law of Small Numbers (1971)," in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (Cambridge, U.K.: Cambridge University Press, 1982), 28.

of what constitutes "a satisfactory answer in psychological research on reasoning."[312] The problem with Kahneman and Tversky's heuristics, as Gigerenzer sees it, is that they are too vague to provide a real explanation. The vagueness is of different, related types that arise from single source: their heuristics do not identify process models that specify the antecedent conditions of their use.[313] In particular, they say nothing about how heuristic processes relate to "specific content[s], context[s], or representation[s]" of information.[314] For Gigerenzer, a satisfactory explanation in psychological theorizing should seek to meet the following standard:

> GM3. Psychological Explanations should invoke functionally specified process models of cognition that specify how the processes relate to specific contents, contexts, and information formats.

Because the conditions of Kahneman and Tversky's heuristics remain unspecified in these ways, the heuristics "at once explain too little and too much." They explain "[t]oo little because we do not know when these heuristics work and how; too much, because, post hoc, one of them can be fitted to almost any experimental result."[315] Without specified conditions of use, the heuristics can be selectively invoked to explain nearly any kind of judgment: if one heuristic fails to account for an observed pattern of judgment, another heuristic can be invoked to save the day. Without specified conditions of use, theorists have plenty of flexibility to selectively invoke the heuristics in ways that "resist attempts to prove, disprove, or even improve them."[316]

There does seem to be some confusion and ambiguity in Kahneman and Tversky's conceptions of these heuristics. For example, consider the representativeness heuristic. In their 1973 paper "On the Psychology of Prediction," Kahneman and Tversky found a way to operationalize the representativeness heuristic: in the Tom W. study, Kahneman and Tversky measured the degree to which Tom W's personality is

---

[312] Gigerenzer, "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)," 592.
[313] Ibid.
[314] Ibid.: 594.
[315] Ibid.: 592. "For example, consider Kahneman and Tversky's major theoretical claims: "judgments of probability or frequency are sometimes influenced by what is similar (representativeness), comes easily to mind (availability), and comes first (anchoring)."[315] Gigerenzer points out how this can happen: "base-rate neglect is commonly attributed to representativeness. However, the opposite result, overweighting of base rates (conservatism), is as easily "explained" by saying the process is anchoring (on the base rate) and adjustment."
[316] Ibid.: 596.

representative of different outcomes by measuring the degree to which Tom W's personality was judged by subjects to be "similar to" the different outcome.[317] However, in the very same paper, Kahneman and Tversky define representativeness as "the degree to which the *outcomes* represent the. . . *evidence*."[318] Here, "represent" takes the direct object "evidence." In a rhetorical move, Kahneman and Tversky nominalize the verb "to represent" into "representative:" for example, "*X* is representative of *Y*." In doing this, we would expect that the direct object of the verb "to represent," would become the adjectival phrase (subjective complement): that is, we would expect "*X* represents *Y*" to become "*X* is representative of *Y*." Hence, we would expect that the claim that "outcomes represent the evidence" would become "outcomes are representative of the evidence:" in the case of the Tom W. study, we would expect that Kahneman and Tversky would be concerned with the degree to which the different outcomes are representative of the description of Tom W. However, Kahneman and Tversky get the order of modification backwards: they aim to discover the degree to which Tom W.'s personality (the evidence) is representative of the outcomes. The reversed order of modification appears in the engineer-lawyer study described in the same paper: subject judgments were "controlled, we suggest, by the degree to which the descriptions (the evidence) appeared representative of these stereotypes (the outcomes)."[319]

The confusion here is theoretically important: the degree to which a model is judged to be similar to an outcome may not be psychologically identical to the degree to which an outcome is judged to be similar to a model – just as the judged similarity between concepts can be asymmetric. For example, "pomegranate" may be judged to be more similar to "apple" than vice versa.[320] Likewise, an individual may be judged to be more similar to a social category to which he or she belongs than vice versa. For example, consider the degree to which George is similar to "middle-income occasional golfers" versus the degree to which the category "middle-income occasional golfers" is

---

[317] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 49.
[318] Ibid., 48. Italics mine
[319] Ibid., 56. Incidentally, this reversed pattern also shows up in their *Science* article (published the same year): "the probability that Steve is a librarian, for example, is assessed by the degree to which he [the evidence] is representative of, or similar to, the stereotype of a librarian [the outcome]." Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases," 4.
[320] Edward Smith, "Concepts and Reasoning," in *Thinking*, ed. Edward Smith and Daniel Osherson, *An Invitation to Cognitive Science* (Cambridge, MA: The MIT Press, 1995), 11.

similar to George.[321] It seems that the method Kahneman and Tversky used to test whether subjects relied on the representativeness heuristic relied on a measure that did not capture the formalized conceptualization. Perhaps we could chalk this up to a slight, understandable mistake in operationalizing the notion of representativeness. However, Kahneman and Tversky's 1982 theoretical paper on the nature of the notion of representativeness seem to accept a bi-directional notion of "representativeness:"

> Representativeness is a directional relation: We say that a sample is more or less representative of a particular population and that an act is representative of a person. . . In some problems, however, it is possible to reverse the roles of model and outcome. For example, one may evaluate whether a person is representative of the stereotype of librarians or whether the occupation of librarian is representative of that person.[322]

These passages seem to suggest that the conditions in which the representativeness heuristic is elicited and the process involved in making judgments of representativeness are somewhat obscured; yet, the text manages to leave readers with the impression that the heuristic is quite representative of subject responses to probability judgments.

Gigerenzer challenges the heuristics and biases research program to move toward more precise models that identify precise, detailed models of cognitive processes that identify the conditions of their use.[323] However, this statement of the critique is not quite right because Kahneman and Tversky do provide some specification of the conditions under which the representativeness heuristic is valid and invalid: the representativeness heuristic is valid in conditions where the judged representativeness (between a model and outcome) is positively correlated with the conditional probability that the outcome is true given the evidence (where this evidence includes the model).

> In many situations, representative outcomes are indeed more likely than others. However, this is not always the case, because there are factors (e.g., the prior probabilities of outcomes and the reliability of the evidence) which affect the likelihood of outcomes but not their representativeness.[324]

---

[321] Thanks to Railton for this example.

[322] Tversky and Kahneman, "Judgments of and by Representativeness," 85. This paper also suggests that, although Kahneman and Tversky accept a bi-directional notion of representativeness, they are interested in a formalized notion that is unidirectional. For example, consider the following, conflicting formalization: "Representativeness is a relation between a process or a model, M, and some instance or event, X, associated with the model. Representativeness, like similarity, can be assessed empirically, for example, by asking people to judge which of two events, X_1 or X_2, is more representative of some model, M, or whether an event, X, is more representative of M_1 or of M_2." Tversky and Kahneman, "Judgments of and by Representativeness," 85.. Italics mine.

[323] Gigerenzer, "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)."

[324] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 48-9.

When representativeness and conditional probabilities are not correlated, then the representativeness heuristic "lead[s] to systematic and predictable errors" under certain conditions.[325]

Kahneman and Tversky sought experimental tasks in which the conditional probability is not well tracked by judgments of representativeness. In order to increase the representativeness of the outcomes, Kahneman and Tversky played on cultural stereotypes of different professions in their personality descriptions. For example, in the engineer/lawyer study, Kahneman and Tversky did not ask subjects which professional identity (engineer or lawyer) is more representative of the personality description offered. Instead, they designed the questionnaire so that "distinct stereotypes were associated with the alternative outcomes," and thought it safe to assume that subjects would judge most of the personality descriptions to be highly representative of one of the outcomes (profession).[326] And, because judgments of representativeness or similarity are "not influenced by several factors that should affect judgments of probability," such as base rate information, they predicted that subjects' judgments would violate Bayes' Rule.[327]

So, in their defense, Kahneman and Tversky say a great deal about the conditions in which judgments of representativeness do and do not apply. And, they provide information at a level of functional description common in psychology: for example, consider the cognitive dissonance and stereotype threat research.[328] Gigerenzer would agree that "[i]t is understandable that when heuristics were first proposed as the underlying cognitive processes in the early 1970s, they were only loosely characterized." His complaint is that "25 years and many experiments later, explanatory notions such as *representativeness* remain vague, undefined, and unspecified with respect both to the antecedent conditions that elicit (or suppress) them and also to the cognitive processes that underlie them."[329] What Gigerenzer seeks is a new level of progress in which the mechanisms are more carefully characterized and made amenable to experimental testing.

---

[325] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases," 20.
[326] Kahneman and Tversky, "On the Psychology of Prediction (1973)," 56.
[327] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases."
[328] Thanks to Railton for this point.
[329] Gigerenzer, "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)," 592.

4.23 The Frequency Algorithm

Gigerenzer has a very specific concern in mind. He worries that, without a more specific model of how judgments of representativeness work, we have no insight into how these judgments relate to different kinds of content or information formats. It is here that Gigerenzer's critique has bite: unlike Kahneman and Tversky, he and his colleagues have specified cognitive models that predict and explain why certain kinds of probability judgments vary with different kinds of information formats.

There are two central insights that Gigerenzer brings to bear here. The first invokes Richard Feynman's appreciation of the fact that formally equivalent representations of the same problem or mathematical statement may be psychologically different.[330] From the perspective of research on judgment under uncertainty, the implication Gigerenzer draws is that mathematical procedures used to produce computationally equivalent solutions may be psychologically different: in particular, one of those mathematical procedures model or describe an evolved cognitive mechanism while the others do not.[331]

Gigerenzer's second insight is that certain algorithms *require* that information be represented in particular kinds of ways: "cognitive algorithms are tuned to certain information formats."[332] There are a few important methodological implications of this additional insight. The first is that, unless we discover what representation of information a cognitive algorithm works on, we will not discover its existence. To illustrate, he suggests the reader "[c]ontemplate for a moment long division in Roman numerals."[333] In order to discover whether we have cognitive algorithms that carry out probabilistic computations, we need to think about the kinds of information formats to which those algorithms are most likely tuned. A further methodological implication is

---

[330] For Feynman, the psychological difference between mathematically equivalent representations has implications on how discovering new laws or insights: he suggests deriving different formulations of the same physical law in order to evoke different mental pictures or understandings. Richard Feynman, *The Character of Physical Law* (Cambridge, MA: The MIT Press, 1967), 53.

[331] Gigerenzer and Hoffrage, "How to Improve Bayesian Reasoning without Instruction: Frequency Formats," 685.

[332] Gerd Gigerenzer, "The Psychology of Good Judgment: Frequency Formats and Simple Algorithms," *Medical Decision Making* 16 (1996): 273.

[333] Gigerenzer and Hoffrage, "How to Improve Bayesian Reasoning without Instruction: Frequency Formats," 685.

that, in order to demonstrate that human judgment is incapable of conforming to particular standards or rules, it is not sufficient to demonstrate that human judgment violates those standards of rules for one type of information format. There may be other information formats – other conditions – upon which existing cognitive algorithms work.

The explanatory strategy Gigerenzer pursues is this: we can explain format effects by invoking a cognitive algorithm that is exclusively tuned to particular information formats. If Gigerenzer were to take this to be the *only* legitimate strategy for explaining format effects, then the methodological upshot would be the strong claim that *properly specified* cognitive models specify a cognitive algorithm that rely on some information formats and not others. Because there is no textual evidence to suggest that Gigerenzer holds such a strong view, I will take Gigerenzer to adopt the weaker claim that specifying cognitive algorithms that exclusively rely on particular information formats is one of however many other legitimate strategies for explaining format effects.

Gigerenzer and Hoffrage, on their search for cognitive algorithms that produce judgments that are computationally equivalent to probabilistic rules, invoke evolutionary theory. They surmise that the kind of probabilistic information available to our ancestors was not single-event probabilities or percentages, but "*frequencies* as actually experienced in a series of events," without base rate information.[334] So, if humans ever evolved a cognitive algorithm carrying out probabilistically consistent judgment, we would expect it to be tuned to the "sequential encoding and updating of event frequencies without access or reference to the base rate."[335] They illustrate how frequency information and the algorithm needed to make use of the information in probabilistically appropriate ways are connected to one another. They ask the reader to suppose there is a physician in an illiterate society who has discovered a symptom that signals a severe disease that has begun to afflict her people:

> In her lifetime, she has seen 1,000 people, 10 of whom had the disease. Of those 10, 8 showed the symptom; of the 990 not afflicted, 95 did. A new patient appears. He has the symptom. What is the probability that he actually has the disease?[336]

[334] Ibid.: 686.
[335] Ibid.
[336] Ibid.: 686-7.

To calculate the conditional probability that the patient has the disease given the symptom, all she needs is the number of cases that had both the symptom and disease (8) and the number of symptom cases (8 + 95):

> [The] Bayesian algorithm for computing the posterior probability $p(H \mid E)$ from the frequency format involves solving the equation:
>
> $$p(H \mid E) = \frac{e \,\&\, h}{(e \,\&\, h) + (e \,\&\, -h)} = \frac{8}{8 + 95}$$
>
> where e & h (*e*vidence and *h*ypothesis) is the number of cases with symptom and disease, and e & -h is the number of cases having the symptom but lacking the disease.[337]

The frequency algorithm Gigerenzer proposes is, formally speaking, equivalent to Bayes' Rule in the sense that it produces the same conditional probability solutions (population statistics) for fixed samples. However, the frequency algorithm allows the physician to make posterior probability calculations without information needed to use traditional formulations of Bayes' Theorem. For example, a traditional formulation of Bayes' law defines the posterior probability in terms of the prior probability (base rate) H, the likelihood or inverse probability of E given H (also known as the hit rate), and the expectedness of E:

$$p(H \mid E) = p(H) \bullet \frac{p(E \mid H)}{p(E)}$$

In contrast, notice that the frequency format algorithm does not require individual information about the base rate H or the expectedness of the evidence E to calculate the posterior distribution.[338] Notice too that the frequency format algorithm has the advantage of computational simplicity and ease compared to traditional formulation of Bayes' Rule. The validity of the algorithm guarantees that its output will report a descriptive population statistic, namely the relative frequency within a finite sample.[339] If subjects actually rely on the frequency algorithm in their conditional probability judgments, the cognitive algorithm explains why subjects' probability judgments tend to conform to valid statistical inference when statistical information is described in terms of frequencies and they are asked to calculate a population statistic for a fixed sample.

---

[337] Ibid.: 687.

[338] Note that one can calculate the expectedness E from $p(E \,\&\, H)$ and $p(E \,\&\, -H)$. Thanks to James Joyce for this.

[339] Thanks to Railton for this way of stating the issue.

However, in the example Gigerenzer discusses in explaining the evolution of the frequency algorithm and in the questions he poses to subjects, he does not look at cases in which we are using a mechanical procedure to identify a population statistic for a fixed population. Rather, the problems described are ones in which we are asked to make a judgment about a new case. In the pre-linguistic physician illustration, he says, "[a] *new* patient appears. He has the symptom. What is the probability that he actually has the disease?"[340] Here, the physician's problem is not simply to report the relative frequency within her previously collected sample. Rather, she must use her sample to make a *projection* about a new case.[341]

To do this, Gigerenzer suggests we use the frequency algorithm as a procedure for going from the sample data to a projection about a new case. However, this cognitive procedure could be unreliable if used with inappropriate samples, i.e. samples that are inappropriately small, statistically heterogeneous, not randomly sampled, etc. There is nothing about the frequency algorithm that guarantees that the sample data acquired is appropriate for the normatively correct use of the frequency algorithm; and so the frequency algorithm cannot be guaranteed to be reliable in the projection or prediction of new cases. The fact that the frequency algorithm can yield mathematically correct results for descriptive statistics in finite samples does not and cannot imply that it provides a rational degree of expectation in projective cases.[342]

Gigerenzer claims that the frequency algorithm is adaptive in the sense that evolution selected for the cognitive process. This is a descriptive, historical claim. Gigerenzer then demonstrates that the frequency algorithm yields mathematically correct results for descriptive statistics in finite samples. However, this mathematical demonstration is not sufficient to close the gap between the descriptive, historical claim and the normative appropriateness of the frequency algorithm in arriving at a rational degree of expectation about cases outside the sample. If Gigerenzer wants to claim that the frequency algorithm outputs a correct degree of expectation in projective cases of inference, then he will need to introduce a norm.

---

[340] Italics mine.
[341] Thanks to Railton for these insights.
[342] Thanks again to Railton for these insights.

We can get a sense of the kind of norm Gigerenzer might have in mind from the kinds of examples and cases he considers. He and Hoffrage pose the following question:

> 103 out of every 1,000 women at age forty get a positive mammography in a routine screening.
> 8 out of every 1,000 women at age forty who participate in routine screening have breast cancer *and* a positive mammography.
> Here is a new representative sample of women at age forty who got a positive mammography in routine screening. How many of these women do you expect to actually have breast cancer? ___out of ___

They discovered that correct conditional probability judgments when statistical information was stated in this format nearly doubled from 28% to 50%.[343] Because the question refers to a sample collected in a medical context (where, presumably, the sample was randomly selected and sufficiently large for the purpose of making predictions about new cases), it seems normatively appropriate to use this information for projective judgments about a different population of individuals.

The norm Gigerenzer and Hoffrage have in mind might be something like this: use the frequency algorithm to arrive at judgments of conditional probability (or rational expectation) in cases where we have statistical information from a sample that is sufficiently large, homogeneous, randomly selected, etc. That is, use the frequency algorithm as a statistician would. This kind of norm is not applied a priori without regard for "relevant structural properties of the given situation," but is tuned to the available information, the information structure (the ways in which the available cues are correlated), and the salient reference classes.[344] The frequency norm stated in this way fulfills GM2, the claim that a cognitive process must be justified for specific problem contents, contexts, and information forms in order to count as an epistemic norm.

The normative appropriateness of using the frequency algorithm for the projective case is not built into the algorithm itself, but is "built-in" by the selection of the experimental task. What is built-in to the frequency algorithm is the information format in which statistical information must be provided in for the algorithm to work: namely, in terms of frequencies rather than single-event probabilities. The frequency norm specifies

---

[343] Gigerenzer and Hoffrage, "How to Improve Bayesian Reasoning without Instruction: Frequency Formats," 688.
[344] Gigerenzer, "Content-Blind Norms, No Norms, or Good Norms? A Reply to Vranas," 93.

conditions of its own normative appropriateness. The frequency algorithm merely specifies conditions of its use. This example demonstrates that just because a psychological theory fulfills GM3 does not imply that it fulfills GM1: that is, just because a psychological explanation invokes process models of cognition that specify how the processes relate to specific contents, contexts, and information formats does not imply that the cognitive process captures or explains the conditions of its own validity. Indeed, the conditions of the epistemic process's validity may be different from the conditions in which the process is used. So, the frequency algorithm sometimes violates the frequency *norm*.

4.3 Reliabilism and Psychology

Gigerenzer has said much about the nature of epistemic norms and has promoted the idea that psychologists should seek to model rational cognitive processes. In this section I will connect Gigerenzer's methodological claims with the explanatory goals of naturalized epistemology.

4.31 The Spirit of Reliabilism

Naturalized epistemologists share the meta-epistemic view that epistemic principles are informed by the a posteriori concepts, reasons, and methods of psychologists, linguists, and other scientists.[345] The account of naturalized epistemology most widely discussed is reliabilism, founded by Alvin Goldman. Goldman thinks that the distinguishing feature that justified beliefs seem to share is that they are *causally* initiated or sustained in what he takes to be acceptable ways: "correct principles of justified belief must be principles that make causal requirements, where "cause" is

[345] Quine, "Epistemology Naturalized," 30.

construed broadly to include sustainers as well as initiators of belief."[346] He diagnoses that the "species of belief-forming (or belief-sustaining) processes" that "are intuitively justification-conferring" have in common the property of "*reliability*: the beliefs they produce are generally true." So, under Goldman's account of justification, "[t]he justificational status of a belief is a function of the reliability of the type of process or processes that cause it, where (as a first approximation) reliability consists in the tendency of a process to produce beliefs that are true rather than false."[347] His basic reliabilist claim is this:

> R1. If S's believing *p* at *t* results from a reliable cognitive belief-forming process (or set of processes), then S's belief in *p* at *t* is justified.[348]

Belief-forming processes, under Goldman's account, are functional operations that generate "a *mapping* from certain states – "inputs" – into other states – "outputs."" The outputs he has in mind "are states of believing this or that proposition at a given moment."[349] The belief-forming process is considered a more general *type* of which a particular belief-output is a *token* or instance. The reliability of the belief-forming process type is measured by the degree to which that process type produces true beliefs (so long as any beliefs constituting the "input" are true).

I take the central insights Goldman's account of reliabilism brings to bear on the nature of justification is this: our theory of justification should be explanatory in two senses. First of all, the reliable belief-forming process's functional procedure for transforming inputs to output beliefs must be able to explain the causal history of a belief: "when we say that a belief is caused by a given process, understood as a functional procedure, we may interpret this to mean that it is caused by the particular *inputs* to the process (and by the intervening events "through which" the functional procedure carries the inputs into the output) on the occasion in question."[350]

> R2. The cognitive belief-forming process should be able to explain S's belief in *p* at *t*.

---

[346] Goldman, "What Is Justified Belief?," 344-5.
[347] Ibid., 345.
[348] Ibid., 347.
[349] Ibid., 346.
[350] Ibid. This specification allows for non-algorithmic models such as probabilistic associative models. Thanks to Railton for this comment.

Notice that the functional belief-forming process does not require that the facts justifying the belief be consciously accessible to the believer.

Second, the cognitive process we invoke should be able to capture or explain the distinctive feature of a person's belief-forming process that confers justifiability onto her belief.

> R3. The cognitive belief-forming process should be able to explain the justificatory status of S's belief in *p* at *t*.

This principle captures the idea that a cognitive belief-forming process can capture or explain the epistemic status of the output-belief by clarifying "the underlying source of justificational status:" in particular, the cognitive belief-forming process's reliability.[351] I take the spirit underlying reliabilism to be encapsulated by these two principles with respect to what counts as justification-conferring cognitive belief-forming process.

4.32 Reliabilism and Psychology: Shared Explanatory Goals

When we understand process reliabilism in terms of R1-R3, and we keep in mind the methodological claims GM1 – GM3 that Gigerenzer has proposed, we can see that cognitive psychology and reliabilism have similar explanatory goals. Recall Gigerenzer's third methodological point:

> GM3. Psychological explanations should invoke functionally specified process models of cognition that specify how the processes relate to specific contents, contexts, and information formats.

Psychologists are interested in explaining output-beliefs by imputing cognitive functions as the processes responsible for transforming inputs to output-beliefs. Gigerenzer in particular seems to suggest that these cognitive processes should, in addition, have built-into them the contexts and conditions for which the cognitive processes are valid or reliable:

---

[351] Ibid., 340.

> GM1. The specification of a cognitive process should capture or explain the conditions of its own validity.

And, because Gigerenzer is interested in identifying the cognitive processes underlying rational judgment, the ways in which these cognitive processes interact with environments, and epistemic norms that are constructed and justified for environments, Gigerenzer proposes cognitive processes whose functional specifications serve as epistemic norms:

> GM2. A cognitive process must be justified for specific problem contents, contexts, and information formats in order to count as an epistemic norm.

Like reliabilists, psychologists like Gigerenzer are interested in explaining the epistemic status of output-beliefs by reference to the validity or reliability of the cognitive functions responsible for transforming inputs into output-beliefs.

Recall, for example, the recognition heuristic. Gigerenzer invokes the recognition heuristic to explain subject responses to two-alternative forced choice tasks, where one is asked to choose which of two provided options taken from the reference class of German cities has a larger population: he invokes this cognitive process to explain subject output-beliefs. He specifies conditions in which the recognition heuristic suite is ecologically valid: mathematically speaking, high recognition validity increases the total ecological validity, as does being as close as possible to half ignorance with respect to the recognized/unrecognized objects. The conditions in which the recognition heuristic suite is ecologically valid or reliable are defined by and built into the cognitive process specified. If we were to uphold the recognition heuristic suite as a kind of reliabilist norm of justification, the cognitive process would do three things: (1) serve as a type of cognitive process responsible for conferring justificational status upon beliefs; (2) explain why S believes *p* at *t*; and (3) because of its built-in conditions of validity, the cognitive process should be able to capture and explain the justificatory status of S's belief *p* at *t*.

4.4 Recasting Reliabilism and its Challenges

The kind of psychologism I introduce in this chapter provides insight into but does not solve classic problems raised against reliabilism. I will discuss these challenges, not with an eye to solving them, but to understand them from the perspective of this form of psychologism.

4.41 On the Generality Problem

I think the kinds of cognitive processes proposed by Gigerenzer bring to bear interesting insights into the debate over the generality problem launched by Richard Feldman. The challenge, in Feldman's words, is this:

> [T]he specific process token that leads to any belief will always be an instance of many process types. For example, the process token leading to my current belief that it is sunny today is an instance of all the following types: the perceptual process, the visual process, processes that occur on Wednesday, processes that lead to true beliefs, etc. Note that these process types are not equally reliable. Obviously, then, one of these types must be the one whose reliability is relevant to the assessment of my belief. Intuitively, it seems clear that the general reliability of processes that occur on Wednesday or processes that lead to true beliefs is not relevant to the assessment of my belief. The reliability of the visual process or of the perceptual process may well be important. Let us say, then, for each belief-forming process token there is some "relevant" type such that it is the reliability of that type which determines the justifiability of the belief produced by that token. Thus, the reliability theory can be formulated as follows:
>
> (RT) S's belief that *p* is justified if and only if the process leading to S's belief that *p* is a process token whose relevant process type is reliable.
>
> In order to evaluate (RT), we need some account of what the relevant types of belief-forming processes are. Without such an account, we simply have no idea what consequences the proposal has since we have no idea which process types are relevant to the evaluation of any particular beliefs.[352]

The problem here is not simply that there are multiple ways of describing the same process token. The problem is that different process types have different degrees of reliability; and, unless we can identify which process type is the most relevant to a process token, we have no way of determining the epistemic status of S's belief that *p*.[353]

---

[352] Richard Feldman, "Reliability and Justification," *The Monist* 68 (1985): 159-60.

[353] Feldman stipulates that an account of relevance must not define the belief-forming processes so narrowly that it covers only a single instance (the token itself), which would make that belief-forming process perfectly reliable or unreliable. The problem with perfect reliability or unreliable is that it entails that all true beliefs are justified and all false beliefs are unjustified. This is the single-case problem. On the other hand, an account of relevance must not be so broad as to lump beliefs of different epistemic status

This critique is somewhat analogous to the objection Gigerenzer raised against Kahneman and Tversky's heuristics: without a specification of the conditions in which one heuristic rather than another causes subject responses, it is always possible to invoke any heuristic to explain any particular subject response.[354]

One of the reasons why the generality problems gets off the ground is because Feldman holds a very strict standard of how the relevant type should be specified: he and Conee suggest that the theory must be "elaborated at least enough to imply exactly what process type has to be reliable in the case in question. A fully general reliabilist theory of justification has to do this for all cases in which there is a fact of the matter" about the epistemic status of a belief.[355] I do not propose a solution to this problem that meets these standards. However, I think there are some interesting connections between the problem of relevance and the problem of disagreement among psychologists. Instead of seeking necessary and sufficient conditions for determining whether a token belief and belief-forming process belongs to a particular type of belief-forming process, the debate among psychologists suggests that we look to the criteria cognitive psychologists use in identifying cognitive processes.

The key disagreement between these psychologists lies in the difference in the *causal explanations* or *cognitive models* they propose. Recall Ralph Hertwig and Gigerenzer's findings in the Linda problem. According to these researchers, subjects conform to the conjunction rule when the question is worded in terms of frequencies because we have evolved an algorithm for the relevant statistical reasoning. The proper functioning of the algorithm requires that the statistical information be represented in terms of frequencies, since information in the natural environment involved the sequential encoding of discrete cases that were enumerated as frequencies rather than as percentages or single-event probabilities.[356] This account commits Hertwig and Gigerenzer to the claim that we have evolved cognitive mechanisms that carry out statistical algorithms, only when provided frequency information.

---

under the same reliable/unreliable cognitive process, which would confer those different beliefs with equal degrees of justification. Feldman dubs this result "the problem of generality. Ibid.: 160-1.

[354] Thanks to Elizabeth Anderson for this insight.

[355] Earl Conee and Richard Feldman, "The Generality Problem for Reliabilism," in *Epistemology: An Anthology*, ed. Ernest Sosa and Jaegwon Kim (Oxford, UK: Blackwell Publishers, Ltd., 2000), 373.

[356] Gigerenzer, "Why the Distinction between Single-Event Probabilities and Frequencies Is Important for Psychology (and Vice Versa)," 142.

When it comes to subjects' probability judgments when the statistical information is presented as frequencies, Kahneman and Tversky provide a competing explanation. They argue that subjects conform to the conjunction rule when the question is worded in terms of frequencies – not because we have such an evolved algorithm for doing so – but because formulating the problem in terms of frequencies walks subjects through the hardest part of the problem. To understand this, consider Hertwig and Gigerenzer's reformulation of the Linda problem: they asked subjects for a numerical estimate of the ratio of women who are bank tellers and/or active feminists in the total population of Linda-like people. That is to say, they asked subjects to estimate the number of people belonging to: the set of bank tellers, the set of active feminists, and the intersection of both these sets, within a fixed population. Being able to judge the probability of a conjunction of events requires conceptualizing how classes of events are related in this set theoretical way. At the very least, understanding the conjunction rule requires seeing that the number of members in the *intersection* of two sets, must be *less than* or equal to the number of members in each of those individual sets. By encouraging subjects to think about the nesting of classes of events, we walk subjects through the hardest step of understanding how to think about and solve the probability task.

By reformulating the probability problem, Tversky and Kahneman took themselves as encouraging "subjects to set up a representation of the problems in which class inclusion is readily perceived and appreciated."[357] Kahneman and Tversky do not see this as supporting the claim that we have evolved a mechanism for calculating probabilities described in terms of frequencies. Rather, they argue that since subjects do not "evaluate compound probabilities by aggregating elementary ones" unless provided a representation "in which different relations and rules are transparent,"[358] we should conclude that thinking in terms of nesting classes of events is *not* "programmed into the lay mind."[359] Gigerenzer would disagree that the mind naturally does think in terms of nesting classes of events when statistical information is stated as frequencies. (Tversky and Kahneman do not take the further, positive step of specifying the cognitive process

[357] Tversky and Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," 309.
[358] Ibid.: 294.
[359] Ibid.: 309.

responsible for frequency format judgments, but point in the direction of further theorizing.)

In addition, the explanations offered by Gigerenzer and Kahneman and Tversky disagree over the primary *cause* of subjects' conformance to the conjunction rule when the Linda problem is framed as a frequency problem.[360] (Is it the transparency of the nesting of events that is responsible for this result? Or, is it the existence of a cognitive mechanism that has evolved to be sensitive to frequencies?) In addition, they disagree over the *cognitive mechanism* responsible for the observed behavior. (Is it a reasoning process that follows Bayesian reasoning, but only for information represented as frequencies? Or, is it a reasoning process that depends crucially upon judgments of representativeness?) So, the disagreements here are not disagreements over who has the appropriate interpretation of probability or who has the correct conceptualization of norms. The disagreement is over what is *causally responsible* (i.e., what features of the input are relevant to the output-beliefs) and what cognitive processes *are* or *are not* responsible for the observed patterns of judgment and choice.[361]

Recasting the generality problem as the problem psychologists face in identifying cognitive processes suggests that a naturalized epistemology should only be interested in causal processes, not logical classifications of processes. Recall that the generality problem, as stated, begins with the recognition that it is possible to classify processes according to many different logical schemes. The logical puzzle is to fix upon one of those logical types. In contrast, the dispute between Gigerenzer and Kahneman and Tversky is not a disagreement over logical schemes. The disagreement is causal: they disagree over what causal process is bringing about observed outcomes. The

---

[360] A similar conclusion is arrived at in Samuels, Stich, and Bishop, "Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear."

[361] The idea that psychologists disagree on what the relevant type is, and how these disagreements are disagreements over the functional type and over the relevant features of the input – is captured by Alston's insight that the psychological "function involved will determine both what features of the input have a bearing on the belief output and what bearing they have, i.,e., how the content of the belief is determined by those features." This idea also touches on Gigerenzer's methodological injunction for psychologists to appreciate how different formal representations of cognitive algorithms may be tuned to different kinds of information or information formats. Or, as Alston states the matter, "it is part of the constitution of the psyche to be so disposed that upon being presented with certain kinds of input a belief is generated with a content that is a certain function of certain features of that input." William P. Alston, "How to Think About Reliability," in *Epistemology: An Anthology*, ed. Ernest Sosa and Jaegwon Kim (Oxford, U.K.: Blackwell Publishers, Ltd., 2000), 361.

disagreement in psychological theorizing suggests that a naturalized epistemology focus on disagreements over causal processes, not in logical classifications of processes.[362]

There are many questions I have left unanswered. For example, I have left open questions about how psychologists should deal with the problem that individual belief-tokens belong to a multiplicity of cognitive belief-forming processes (such as, electrochemical processes, organic process, and neurological processes) which are invoked by different sciences that invoke causal processes couched at different levels of generality.[363] Unfortunately, I will have to these questions off for later discussion.

4.42 On Epistemic Naturalism

Naturalized epistemologists generally adopt the thesis of substantive naturalism, or *epistemic naturalism*, which claims that the necessary and sufficient conditions for justified belief are constituted by naturalistic criteria. Notice that unlike the "naturalism" of "ethical naturalism," substantive naturalism does not claim that our valuational terms definitionally reduce to naturalistic terms. Rather, substantive naturalists make the more modest claim that our valuational epistemic terms *supervene* on naturalistic ones: if a belief is justified, it is so in virtue of having certain factual, non-epistemic properties.[364] Recall Goldman's account of justification:

> R1. If S's believing *p* and *t* results from a reliable cognitive belief-forming process (or set of processes), then S's belief in *p* and *t* is justified.[365]

Goldman claims that, since he defines "reliable belief-forming process" in terms of "such notions as belief, truth, statistical frequency, and the like, it is not an epistemic term."[366] One advantage of defining justified belief in terms of nonepistemic terms is that our theory of justification avoids making reference to difficult normative notions: invoking normative notions in the criteria for justified belief would invite an infinite regress of

---

[362] Thanks to Railton for pointing this out.
[363] Conee and Feldman, "The Generality Problem for Reliabilism," 377.
[364] Kim, "What Is "Naturalized Epistemology"?," 310.
[365] Goldman, "What Is Justified Belief?," 347.
[366] Ibid.

valuational concepts, each of which depends on the one below it in its criterion of application.

The psychologism I consider in this chapter might be considered worrisome for reliabilism in the sense that it might open up the objection that this kind of reliabilism is parasitic upon other, more fundamental accounts of justification. For example, consider the theory Goldman refers to as *evidence proportionalism*. The thesis of evidence proportionalism is that "justifiedness consists in proportioning your degree of credence in a hypothesis to the weight of your evidence," where the weight of evidence is scaled on an interval of 0 to 1, the cognizer fixes her degree of belief as a function of evidential weight, and the "weights of evidence are derivable from certain formal facts together with the cognizer's present evidential corpus."[367] Goldman's primary concern with evidence proportionalism is that it suggests that it is ultimately one's evidence for a hypothesis that confers justification onto beliefs – *not* the reliability of the cognitive processes responsible for sustaining and producing beliefs.

One might worry that the kind of cognitive processes I consider are reliable because their functional characterization finds reliable ways of proportioning available evidence to belief. The frequency algorithm proportions statistical information described in terms of frequencies in weighing posterior probabilities. And, the recognition heuristic suite proportions evidential weight or validities to different types of judgments made on the basis of recognition, knowledge, or guesswork. I have claimed that the frequency algorithm and recognition heuristic suite should count as norms of justification insofar as they are reliable belief-forming processes. However, because of the nature of the functional characterization of these cognitive belief-forming processes, objectors might well wonder whether it is the ways in which these belief-forming processes proportion and weight evidence that serves as the proper or legitimate source of justification.

Goldman did not think that this kind of worry was important because he thought it "highly unlikely that our native cognitive architecture would realize" ideal statistical methods." In support of this claim he cited research from the heuristics and biases tradition which has been "engaged in arguing that native inference propensities do not

[367] Alvin I. Goldman, *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986), 89.

match those of 'normative' statistical methodology."[368] However, he argues that even if it turned out that cognitive belief-forming processes instantiated statistical or other forms of what is traditionally seen as good reasoning, Goldman argues that this possibility is not inconsistent with reliabilism: psychologism's "talk of 'using' the method implies agreement with the claim that psychological processes are critical for justifiedness" since "statistical methods must be psychologically instantiated in order to yield justifiedness."[369]

Earl Conee and Feldman anticipate this kind of concern:

> A solution [to the generality problem] cannot identify the relevant types for a process in a way that merely smuggles a non-reliabilist epistemic evaluation into the characterization of relevant types. For instance, one could develop a form of "reliabilism" that just restates an evidentialist theory of justification in a roundabout way. Pseudo-reliabilism of this sort holds that there are only two relevant types of belief-forming process. One type is "belief based on adequate evidence" and the other type is "belief based on inadequate evidence." Assuming that the first of these is reliable and the second is not, this version of reliabilism will get plausible results (or at least results that an evidentialist would find plausible). But this theory is only verbally a version of reliabilism. It mentions the processes of belief formation only in order to characterize the quality of the evidence for the belief. This is obviously incompatible with the spirit of process reliabilism.[370]

If psychologists really should specify cognitive processes in ways that have built-into them conditions of validity and reliability, then does reliabilism – by invoking these as relevant, reliable cognitive processes – implicitly invoke an evidentialist or some other kind of theory of justification?

4.5 Conclusion

In this chapter I consider a form of psychologism for naturalized epistemology which is suggested by contemporary psychological research undertaken by the fast and

---

[368] Ibid., 92. The work he cites include Daniel Kahneman, Paul Slovic, and Amos Tversky, *Judgment under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press, 1982). Richard E. Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, ed. James J. Jenkins, Walter Mischel, and Willard W. Hartup, *The Century Psychology Series* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1980).

[369] Goldman, *Epistemology and Cognition*, 90-1.

[370] Conee and Feldman, "The Generality Problem for Reliabilism," 374.

frugal heuristics research program. This form of psychologism is significant for naturalized epistemology because it underscores a particular type of continuity between psychology and epistemology: namely, a continuity in which epistemic concepts and norms show up in the functional characterization of cognitive processes and in the content of psychological theories.

Some of the methodological standards and critiques Gigerenzer has proposed for psychological explanations and cognitive processes also have implications for naturalized epistemology. Gigerenzer's form of ecological rationalism – the preference for discovering conditions that promote rational judgment – has brought with it stricter methodological standards about how cognitive processes should be specified: in particular, they should connect with contents, contexts, and information formats; and, they should have built-in conditions of validity. These methodological claims have interesting connections with the explanatory goals of reliabilism: both invoke cognitive belief-forming processes to explain the production of output-beliefs and to explain the justificatory status of output-beliefs. And, by casting contemporary psychological theorizing as an extension of naturalized epistemology, we recast the puzzles facing traditional accounts of reliabilism.

I hope that the analysis in this chapter broadens the connections between methodological critiques in cognitive psychology and naturalized epistemology more generally: in the kind of psychologism I suggest, the interests of theoretical science – methodological issues about how to specify cognitive processes – have a very intimate connection and kinship with reliabilist theories of justification.

This dissertation has analyzed the many ways in which philosophy and psychology inform one another. I began this project by inquiring into how philosophers have used rationality as a methodological tool in interpretation and suggested that research in psychology successfully challenges those accounts. In the second chapter, I took a philosophical perspective on contemporary psychological research and argued for an account of ecological rationalism that seeks to identify the conditions promoting rational judgment, where the preference for discovering rational judgment is motivated

by moral and social interests in promoting cognitive health. In the third chapter, I looked at how Grice's account of cooperative communication has motivated improved methodological standards in research on human judgment. And, in the final chapter, I have argued that the methodological standards adopted by cognitive psychologists can have important connections to naturalized epistemology.

Bibliography

Adler, Jonathan E. "Abstraction Is Uncooperative." *Journal for the Theory of Social Behavior* 14, no. 2 (1984): 165-81.

Alston, William P. "How to Think About Reliability." In *Epistemology: An Anthology*, edited by Ernest Sosa and Jaegwon Kim, 354-71. Oxford, U.K.: Blackwell Publishers, Ltd., 2000.

Anderson, Elizabeth. "Feminist Epistemology: An Interpretation and Defense." *Hypatia* 10 (1995): 50-84.

———. "Knowledge, Human Interests, and Objectivity in Feminist Epistemology." *Philosophical Topics* 23, no. 2 (1995): 27-58.

Axelrod, R., and W. Hamilton. "The Evolution of Cooperation." *Science* 211 (1981): 1390-6.

Bacon, Francis. *The New Organon and Related Writings*. Edited by Fulton H. Anderson. New York, N.Y.: The Liberal Arts Press, 1960.

Baddeley, Alan. "Applied Cognitive and Cognitive Applied Psychology: The Case of Face Recognition." In *Perspectives on Memory Research: Essays in Honor of Uppsala University's 500th Anniversary*, edited by Lars-Goran Nilsson, 367-. New York, NY: Lawrence Erlbaum Associates, 1979.

Bazerman, M. H. *Judgment in Managerial Decision Making*. New York, NY: Wiley, 1990.

Berkeley, Diana, and Patrick Humphreys. "Structuring Decision Problems and the 'Bias Heuristic'." *Acta Psychologica* 50, no. 3 (1982): 201-52.

Beyth-Maron, Ruth, Shlomith Dekel, Ruth Gombo, and Moshe Shaked. *An Elementary Approach to Thinking under Uncertainty*. Translated by Sarah Lichtenstein, Benny Marom and Ruth Beyth-Marom. Mahwah, NJ: Lawrence Erlbaum Associates, 1985.

Boyd, Richard. "Kinds as the 'Workmanship of Men': Realism, Constructivism, and Natural Kinds." In *Rationalitat, Realismus, Revision: Proceedings of the Third International Congress, Gesellschaft Fur Analytische Philosophie*, edited by Julian Nida-Rumelin. Berlin, Germany: de Gruyter, 1999.

Brunswik, Egon. "Symposium on the Probability Approach in Psychology." *Psychological Review* 62, no. 3 (1955): 193-217.

Cascells, W., A. Schoenberger, and T. B. Grayboys. "Interpretation by Physicians of Clinical Laboratory Results." *New England Journal of Medicine* 299 (1978): 999-1000.

Chomsky, Noam. *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press, 1965.

Christensen-Szalanski, Jay J. J., and Lee Roy Beach. "The Citation Bias: Fad and Fashion in the Judgment and Decision Literature." *American Psychologist* 39 (1984): 75-8.

Clark, Herbert H., and Michael F. Schober. "Asking Questions and Influencing Answers." In *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*, edited by Judith M. Tanur, 15-48. New York, NY: Russell Sage Foundation, 1992.

Cohen, L. Jonathan. "Can Human Irrationality Be Experimentally Demonstrated?" *The Behavioral and Brain Sciences* 4 (1981): 317-70.
———. "On the Psychology of Prediction: Whose Is the Fallacy?" *Cognition* 7 (1979): 385-407.
Conee, Earl, and Richard Feldman. "The Generality Problem for Reliabilism." In *Epistemology: An Anthology*, edited by Ernest Sosa and Jaegwon Kim, 372-86. Oxford, UK: Blackwell Publishers, Ltd., 2000.
Cosmides, Leda, and John Tooby. "Are Humans Good Intuitive Statisticians after All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty." *Cognition* 58 (1996): 1-73.
Davidson, Donald. "Belief and the Basis of Meaning (1974)." In *Inquiries into Truth and Interpretation*, 141-54. New York, NY: Oxford University Press, 2001.
———. "A Coherence Theory of Truth and Knowledge." In *Epistemology: An Anthology*, edited by Ernest Sosa and Jaegwon Kim, 154-63. 2000: Blackwell Publishers Ltd., 2000.
———. "On the Very Idea of a Conceptual Scheme (1974)." In *Inquiries into Truth and Interpretation*, 183-98. New York, NY: Oxford University Press, 2001.
———. "Psychology as Philosophy (1974)." In *Essays on Actions and Events*, 229-38. Oxford, UK: Oxford University Press, 1980.
———. "Radical Interpretation (1973)." In *Inquiries into Truth and Interpretation*, 125-40. New York, NY: Oxford University Press, 2001.
———. "Thought and Talk (1975)." In *Inquiries into Truth and Interpretation*, 155-70. New York, NY: Oxford University Press, 2001.
———. "Truth and Meaning." In *Inquiries into Truth and Interpretation*, 17-36. New York, NY: Oxford University Press, 2001.
Dennett, Daniel C. "Making Sense of Ourselves." In *Mind and Cognition: A Reader*, edited by William Lycan, 184-98. Oxford, UK: Basil Blackwell Ltd, 1990.
———. "Three Kinds of Intentional Psychology." In *Reduction, Time, and Reality*, edited by R. A. Healey. New York, NY: Cambridge University Press, 1981.
———. "True Believers: The Intentional Strategy and Why It Works." In *The Intentional Stance*, 13-35. Cambridge, MA: The MIT Press, 1987.
Dulany, Don E., and Denis J. Hilton. "Conversational Implicature, Conscious Representation, and the Conjunction Fallacy." *Social Cognition* 9, no. 1 (1991): 85-110.
Eddy, David M. "Probabilistic Reasoning in Clinical Medicine." In *Judgment under Uncertainty*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 249-67. Cambridge, UK: Cambridge University Press, 1982.
Edwards, Ward. "Behavioral Decision Theory." *Annual Review of Psychology* 12 (1961): 473-98.
———. "Human Cognitive Capabilities, Representativeness, and Ground Rules for Research." In *Analysing and Aiding Decision Processes*, edited by Patrick Humphreys, Ola Svenson and Anna Vari, 507-13. Amsterdam, Holland: North-Holland Publishing Company, 1983.
———. "Nonconservative Information Processing Systems." Ann Arbor, MI: University of Michigan, Institute of Science and Technology, 1966.

Edwards, Ward, Harold Lindman, and Lawrence D. Phillips. "Emerging Technologies for Making Decisions." In *New Directions in Psychology 2*, 259-326. New York: Holt, Rinehard and Winston, Inc., 1965.

Elstein, Arthur S., Lee S. Shulman, and Sarah A. Sprafka. "Medical Problem Solving: A Ten-Year Retrospective." *Evaluation and the Health Professions* 13, no. 1 (1990): 5-36.

Elster, Jon. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge, U.K.: Cambridge University Press, 1983.

Elwyn, Glyn, Adrian Edwards, Martin Eccles, and David Rovner. "Decision Analysis in Patient Care." *Lancet* 358, no. 9281 (2001): 571-4.

Fagerlin, Angela, C. Wang, and Peter A. Ubel. "Reducing the Influence of Anecdotal Reasoning on People's Health Care Decisions: Is a Picture Worth a Thousand Statistics?" *Medical Decision Making* 25, no. 4 (2005): 398-405.

Feldman, Richard. "Reliability and Justification." *The Monist* 68 (1985): 159-74.

Feynman, Richard. *The Character of Physical Law*. Cambridge, MA: The MIT Press, 1967.

Fischhoff, Baruch. "Clinical Decision Analysis." *Operations Research* 28 (1980): 28-43.

———. "Debiasing." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 422-44. Cambridge, U.K.: Cambridge University Press, 1982.

———. "Informed Consent in Societal Risk-Benefit Decisions." *Technical Forecasting and Social Change* 13 (1979): 347-58.

———. "Reconstructive Criticism." In *Analysing and Aiding Decision Processes*, edited by Patrick Humphreys, Ola Svenson and Anna Vari, 515-23. Amsterdam, Holland: North-Holland Publishing Company, 1983.

———. "Risk Perception and Communication Unplugged: Twenty Years of Process." *Risk Analysis* 15, no. 2 (1995): 137-45.

Fisher, Ann. "Risk Communication Challenges." *Risk Analysis* 11, no. 2 (1991): 173-9.

Fong, Geoffrey T., David H. Krantz, and Richard E. Nisbett. "The Effects of Statistical Training on Thinking About Everyday Problems." *Cognitive Psychology* 18 (1986): 253-92.

Frank, Robert H. *Passions within Reason: The Strategic Role of the Emotions*. New York, NY: W. W. Norton & Company, 1988.

Gaeth, G. J., and James Shanteau. "Reducing the Influence of Irrelevant Information on Experienced Decision Makers." *Organizational Behavior and Human Performance* 33 (1984): 263-82.

Gigerenzer, Gerd. "Content-Blind Norms, No Norms, or Good Norms? A Reply to Vranas." *Cognition* 81 (2001): 93-103.

———. "From Tools to Theories: A Heuristic Discovery in Cognitive Psychology." *Psychological Review* 98, no. 2 (1991): 254-67.

———. "How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases"." *European Review of Social Psychology* 2 (1991): 83-115.

———. "Is the Mind Irrational or Ecologically Rational?" In *The Law and Economics of Irrational Behavior*, edited by Francesco Parisi and Vernon L. Smith, 37-67. Stanford, CA: Stanford University Press, 2005.

———. "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)." *Psychological Review* 103, no. 3 (1996): 592-96.
———. "The Psychology of Good Judgment: Frequency Formats and Simple Algorithms." *Medical Decision Making* 16 (1996): 273-80.
———. "Why the Distinction between Single-Event Probabilities and Frequencies Is Important for Psychology (and Vice Versa)." In *Subjective Probability*, edited by George Wright and Peter Ayton, 129-61. New York, NY: John Wiley & Sons, 1994.
Gigerenzer, Gerd, and Daniel G. Goldstein. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* 103, no. 4 (1996): 650-69.
Gigerenzer, Gerd, and Ulrich Hoffrage. "How to Improve Bayesian Reasoning without Instruction: Frequency Formats." *Psychological Review* 102, no. 4 (1995): 684-704.
Gigerenzer, Gerd, Ulrich Hoffrage, and Heinz Kleinbolting. "Probabilistic Mental Models: A Brunswikian Theory of Confidence." *Psychological Review* 98, no. 4 (1991): 506-28.
Gigerenzer, Gerd, Peter M. Todd, and ABC Research Group. *Simple Heuristics That Make Us Smart*. Oxford, UK: Oxford University Press, 1999.
Ginossar, Zvi, and Yaacov Trope. "Problem Solving in Judgment under Uncertainty." *Journal of Personality and Social Psychology* 52, no. 3 (1987): 464-74.
Goldman, Alvin I. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press, 1986.
———. "What Is Justified Belief?" In *Epistemology: An Anthology*, edited by Hilary Kornblith, 340-53. Oxford, U.K.: Blackwell Publishers Ltd., 2000.
Goldstein, Daniel G., and Gerd Gigerenzer. "Models of Ecological Rationality: The Recognition Heuristic." *Psychological Review* 109, no. 1 (2002): 75-90.
Goodman, Nelson. *Ways of Worldmaking*. Indianapolis, IN: Hackett Publishing Company, 1978.
Grice, Paul. "Logic and Conversation." In *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 1989.
Henderson, David K. *Interpretation and Explanation in the Human Sciences*. Albany, NY: State University of New York Press, 1993.
Hertwig, Ralph, and Gerd Gigerenzer. "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors." *Journal of Behavioral Decision Making* 12 (1999): 275-305.
Hilton, Denis J. "The Social Context of Reasoning: Conversational Inference and Rational Judgment." *Psychological Bulletin* 118, no. 2 (1995): 248-71.
Hoffrage, Ulrich, and Gerd Gigerenzer. "Using Natural Frequencies to Improve Diagnostic Inferences." *Academic Medicine* 73, no. 5 (1998): 538-40.
Hogarth, Robin M. "Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgmental Heuristics." *Psychological Bulletin* 90, no. 2 (1981): 197-217.
Hynes, M., and E. Vanmarcke. *Reliability of Embankment Performance Predictions, Proceedings of the Asce Engineering Mechanics Division Specialty Conference*. Waterloo, Ontario, Canada: University of Waterloo Press, 1976.

Jou, Jerwen, James Shanteau, and Richard Jackson Harris. "An Information Processing View of Framing Effects: The Role of Causal Schemas in Decision Making." *Memory & Cognition* 24, no. 1 (1996): 1-15.

Kahneman, Daniel, Paul Slovic, and Amos Tversky. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press, 1982.

Kahneman, Daniel, and Amos Tversky. "On the Interpretation of Intuitive Probability: A Reply to Jonathan Cohen." *Cognition* 7 (1979): 409-11.

———. "On the Psychology of Prediction (1973)." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 48-68. New York, NY: Cambridge University Press, 1982.

———. "On the Reality of Cognitive Illusions." *Psychological Review* 103, no. 3 (1996): 582-91.

———. "On the Study of Statistical Intuitions." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 493-508. New York, NY: Cambridge University Press, 1982.

———. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47, no. 2 (1979): 263-92.

Keenan, Elinor Ochs. "The Universality of Conversational Postulates." *Language in Society* 5 (1976): 67-80.

Kim, Jaegwon. "What Is "Naturalized Epistemology"?" In *Epistemology: An Anthology*, edited by Ernest Sosa and Jaegwon Kim, 301-13. Oxford, U.K.: Blackwell Publishers, Ltd., 1997.

Klayman, Joshua, and Kaye Brown. "Debias the Environment Instead of the Judge: An Alternative Approach to Reducing Error in Diagnostic (and Other) Judgment." *Cognition* 49 (1993): 97-122.

Klayman, Joshua, Chip Heath, and Richard P. Larrick. "Cognitive Repairs: How Organizational Practices Can Compensate for Individual Shortcomings." *Research in Organizational Behavior* 20 (1998): 1-37.

Klayman, Joshua, Jack B. Soll, Claudia Gonzalez-Vallejo, and Sema Barlas. "Overconfidence: It Depends on How, What, and Whom You Ask." *Organizational Behavior and Human Performance* 79, no. 3 (1999): 216-47.

Koehler, Jonathan J. "The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient." *Southern California Law Review* 74 (2001): 1275-8.

———. "When Are People Persuaded by DNA Match Statistics?" *Law and Human Behavior* 25 (2001): 493-503.

Kornblith, Hilary. "Introduction: What Is Naturalistic Epistemology?" In *Naturalizing Epistemology*, edited by Hilary Kornblith, 1-14. Cambridge, MA: The MIT Press, 1997.

Korobkin, R. B., and T. S. Ulen. "Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics." *California LAw Review* 88 (2000): 1051-144.

Krosnick, Jon A., Fan Li, and Darrin R. Lehman. "Conversational Conventions, Order of Information Acquisition, and the Effect of Base Rates and Individuating Information on Social Judgments." *Journal of Personality and Social Psychology* 59, no. 6 (1990): 1140-52.

Kruglanski, Arie W. "The Human Subject in the Psychology Experiment: Fact and Artifact." *Advances in Experimental Social Psychology* 8 (1975): 101-47.

Kuhn, Thomas S. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, 320-39. Chicago: University of Chicago Press, 1977.

Kunda, Ziva, and Paul Thagard. "Forming Impressions from Stereotypes, Traits, and Behaviors: A Parallel-Constraint-Satisfaction Theory." *Psychological Review* 103, no. 2 (1996): 284-308.

Lee, Carole J. "Gricean Charity: The Gricean Turn in Psychology." *Philosophy of the Social Sciences* 36, no. 2 (2006): 193-218.

Lehman, Darrin R., and Richard E. Nisbett. "A Longitudinal Study of the Effects of Undergraduate Training on Reasoning." In *Rule for Reasoning*, edited by Richard E. Nisbett, 340-57. Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.

Lehman, Darrin R., Richard E. Nisbett, and Richard O. Lempert. "The Effects of Graduate Training on Reasoning: Formal Discipline Nad Thinking About Everyday-Life Events." In *Rule for Reasoning*, edited by Richard E. Nisbett, 315-39. Hillsdale, NJ: Larence Erlbaum Associates, 1993.

Leslie, A. "The Theory of Mind Impairment in Autism: Eidence for a Modular Mechanism of Development?" In *Natural Theories of Mind*, edited by A. Whiten, 51-62. Oxford, U.K.: Blackwell Publishers, Ltd., 1991.

Levinson, Stephen. "Conversational Implicature." In *Pragmatics*, 97-166. Cambridge: Cambridge University Press, 1983.

Lindsey, Samuel, Ralph Hertwig, and Gerd Gigerenzer. "Communicating Statistical Evidence." *Jurimetrics* 43, no. Winter (2003): 147-63.

Longino, Helen E. "Gender, Politics, and the Theoretical Virtues." *Synthese* 104, no. 3 (1995): 383-97.

———. *Science as Social Knowledge*. Princeton, N.J.: Princeton University Press, 1990.

Lopes, L. Lola. "The Rhetoric of Irrationality." *Theory and Psychology* 1, no. 1 (1991): 65-82.

Lupia, Arthur, and Mathew D. McCubbins. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Edited by Randall Calvert and Thrainn Eggertsson, *Political Economy of Institutions and Decisions*. Cambridge, U.K.: Cambridge University Press, 1998.

Lupia, Arthur, Mathew D. McCubbins, and Samuel L. Popkin, eds. *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*. Edited by James H. Kuklinski and Dennis Chong, *Cambridge Studies in Political Psychology and Public Opinion*. Cambridge, U.K.: Cambridge University Press, 2000.

Nisbett, Richard E., David H. Krantz, Christopher Jepson, and Geoffrey T. Fong. "Improving Inductive Inference." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 445-59. Cambridge, U.K.: Cambridge University Press, 1982.

Nisbett, Richard E., David H. Krantz, Christopher Jepson, and Ziva Kunda. "The Use of Statistical Heuristics in Everyday Inductive Reasoning." *Psychological Review* 90, no. 4 (1983): 339-63.

Nisbett, Richard E., Darrin R. Lehman, Geoffrey T. Fong, and Patricia W. Cheng. "Teaching Reasoning." In *Rules for Reasoning*, edited by Richard E. Nisbett, 297-314. Hillsdale, NJ: Larence Erlbaum Associates, 1993.

Nisbett, Richard E., and Lee Ross. *Human Inference: Strategies and Shortcomings of Social Judgment*. Edited by James J. Jenkins, Walter Mischel and Willard W. Hartup, *The Century Psychology Series*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1980.

Nisbett, Richard E., and Timothy DeCamp Wilson. "Telling More Than We Can Know: Verbal Reports on Mental Models." *Psychological Review* 84, no. 3 (1977): 231-59.

Nisbett, Richard E., Henry Zukier, and Ronald E. Lemley. "The Dilution Effect: Nondiagnostic Information Weakens the Implications of Diagnostic Information." *Cognitive Psychology* 13 (1981): 248-77.

Peterson, C. R., and Lee Roy Beach. "Man as an Intuitive Statistician." *Psychological Bulletin* 68 (1967): 29-46.

Phillips, Lawrence D. "A Theoretical Perspective on Heuristics and Biases in Probabilistic Thinking." In *Analysing and Aiding Decision Processes, Volume 14*, edited by Patrick Humphreys, Ola Svenson and Anna Vari, 525-43. Amsterdam: North-Holland Publishing Co., 1983.

Piaget, Jean, and Barbel Inhelder. *The Origin of the Idea of Chance in Children*. New York, NY: Norton, 1951. Reprint, 1975.

Posner, Eric A. "Probability Errors: Some Positive and Normative Implications for Tort and Contract Law." In *The Law and Economics of Irrational Behavior*, edited by Francesco Parisi and Vernon L. Smith, 456-73. Stanford, CA: Stanford University Press, 2005.

Quine, W. V. O. "Epistemology Naturalized." In *Naturalizing Epistemology*, edited by Hilary Kornblith, 15-31. Cambridge, MA: The MIT Press, 1997.

———. "Evidence." In *Pursuit of Truth*, 1-21. Cambridge, MA: Harvard University Press, 1990.

Risjord, Mark. *Woodcutters and Witchcraft*. Albany, NY: State University of New York Press, 2000.

Rohrmann, B. "The Evaluation of Risk Communication Effectiveness." *Acta Psychologica* 81, no. 2 (1992): 169-92.

Rosaldo, Michelle Z. *Knowledge and Passion: Ilongot Notions of Self and Social Life*. Cambridge, UK: Cambridge University Press, 1980.

Saks, Michael J., and Robert F. Kidd. "Human Information Processing and Adjudication: Trial by Heuristics." *Law and Society Review* 15, no. 1 (1980): 123-60.

Samuels, Richard, Stephen P. Stich, and Michael Bishop. "Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear." In *Common Sense: Reasoning and Rationlity*, edited by Renee Elio, 236-68. Oxford: Oxford University Press, 2002.

Schwarz, Norbert. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 1996.

———. "Judgment in a Social Context: Biases, Shortcomings, and the Logic of Conversation." *Advances in Experimental Social Psychology* 26 (1994): 123-62.

Schwarz, Norbert, Fritz Strack, Denis Hilton, and Gabi Naderer. "Base Rates, Representativeness, and the Logic of Conversation: The Contextual Relevance of "Irrelevant" Information." *Social Cognition* 9, no. 1 (1991): 67-84.

Simon, Herbert A. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69, no. 1 (1955): 99-118.

———. "Rational Choice and the Structure of the Environment." *Psychological Review* 63, no. 2 (1956): 129-38.

———. *The Sciences of the Artificial.* 3 ed. Cambridge, MA: The MIT Press, 1996.

Singer, Eleanor, Hans-Jurgen Hippler, and Norbert Schwarz. "Confidentiality Assurances in Surveys: Reassurance or Threat?" *International Journal of Public Opinion Research* 4, no. 3 (1992): 256-68.

Slovic, Paul, G. Fischhoff, and S. Lichtenstein. "Facts Versus Fears: Understanding Perceived Risk." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 463-89. Cambridge, U.K.: Cambridge University Press, 1982.

Smith, Edward. "Concepts and Reasoning." In *Thinking*, edited by Edward Smith and Daniel Osherson, 1-33. Cambridge, MA: The MIT Press, 1995.

Smith, Vernon L. *Bargaining and Market Behavior: Essays in Experimental Economics.* New York, NY: Cambridge University Press, 2000.

———. "Rational Choice: The Contrast between Economics and Psychology." *Journal of Political Economy* 99, no. 4 (1991): 877-97.

Sniderman, P. M., R. A. Brody, and Philip E. Tetlock. *Reasoning and Choice: Explorations in Political Psychology.* New York, NY: Cambridge University Press, 1991.

Sober, Elliott. "Psychologism." *Journal for the Theory of Social Behavior* 8, no. 2 (1979): 165-91.

Sperber, Dan, and Deirdre Wilson. *Relevance: Communication and Cognition.* Oxford, U.K.: Basil Blackwell Ltd., 1986.

Stanovich, Keith E. *Who Is Rational? Studies of Individual Differences in Reasoning.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 1999.

Stanovich, Keith E., and Richard F. West. "Individual Differences in Reasoning: Implications for the Rationality Debate?" *The Behavioral and Brain Sciences* 23 (2000): 645-726.

Stein, Edward. *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science.* Oxford, U.K.: Oxford University Press, 1996.

Stich, Stephen P. "Could Man Be an Irrational Animal? Some Notes on the Epistemology of Rationality." *Synthese* 64 (1985): 115-35.

———. "Dennett on Intentional Systems." In *Mind and Cognition: A Reader*, edited by William Lycan, 167-84. Oxford, UK: Basil Blackwell Ltd, 1990.

Stone, Eric R., Winston R. Sieck, Benita E. Bull, J. Frank Yates, Stephanie C. Parks, and Carolyn J. Rush. "Foreground:Background Salience: Explaining the Effects of Graphical Displays on Risk Avoidance." *Organizational Behaviour and Human Decision Processes* 90, no. 1 (2003): 19-36.

Sunstein, C. R. "Behavioral Analysis of Law." *University of Chicago Law Review* 64 (1997): 1175-95.

Tanner, Wilson P., and John A. Swets. "A Decision-Making Theory of Visual Detection." *Psychological Review* 61, no. 6 (1954): 401-9.

Taylor, Charles. "Interpretation and the Sciences of Man." In *Philosophy and the Human Scienecs*, 15-57. Cambridge, UK: Cambridge University Press, 1985.

———. "Understanding and Ethnocentricity." In *Philosophy and the Human Sciences*, 116-33. Cambridge, UK: Cambridge University Press, 1985.

Taylor, Shelley E. "The Social Being in Social Psychology." In *The Handbook of Social Psychology*, edited by Daniel T. Gilbert, Susan T. Fiske and Gardner Lindzey, 58-95. Boston, MA: The McGraw Hill Companies, Inc., 1998.

Tetlock, Philip E. "The Impact of Accountability on Judgment and Choice: Toward a Social Contingency Model." *Advances in Experimental Social Psychology* 25 (1992): 331-76.

Thagard, Paul, and Richard E. Nisbett. "Rationality and Charity." *Philosophy of Science* 50 (1983): 250-67.

Thompson, William C., and Edward L. Schumann. "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy." *Law and Human Behavior* 11 (1987): 167-87.

Tversky, Amos. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5 (1973): 207-32.

Tversky, Amos, and Daniel Kahneman. "Belief in the Law of Small Numbers (1971)." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 23-31. Cambridge, U.K.: Cambridge University Press, 1982.

———. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90, no. 4 (1983): 293-315.

———. "The Framing of Decisions and the Psychology of Choice." *Science* 211, no. 4481 (1981): 453-8.

———. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (1974): 1124-31.

———. "Judgments of and by Representativeness." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 84-98. New York, NY: Cambridge University Press, 1982.

———. "Rational Choice and the Framing of Decisions." *The Journal of Business* 59, no. 4, Part 2: The Behavioral Foundations of Economic Theory (1986): S251-S78.

Ubel, Peter A. "Is Information Always a Good Thing? Helping Patients Make "Good" Decisions." *Medical Care* 40, no. 9 (2002): V39-V44.

Ubel, Peter A., Christopher Jepson, and Jonathan Baron. "The Inclusion of Patient Testimonials in Decision Aids: Effects on Treatment Choices." *Medical Decision Making* 21, no. 1 (2001): 60-8.

Weber, Max. *The Theory of Social and Economic Organization*. Translated by A. M. Henderson and Talcott Parsons. New York, NY: Oxford University Press, 1947.