CAN HUMAN RATIONALITY BE DEFENDED *A PRIORI*?

David Shier
Washington State University

ABSTRACT: In this paper, I develop two criticisms of L. Jonathan Cohen's influential *a priori* argument that human irrationality cannot be experimentally demonstrated. The first is that the argument depends crucially on the concept of a *normal human* but that no such concept suitable for Cohen's purposes is available. The second is that even if his argument were granted, his thesis of an unimpeachable human capacity for reasoning is not a defense of human reasoning, but rather amounts to the claim that we cannot make any meaningful evaluative claims about human reasoning whatsoever.

Responding to a proliferation of psychological studies purported to experimentally demonstrate widespread systematic defects in basic human reasoning, L. Jonathan Cohen famously argued that such studies do not have "bleak implications for human rationality" (1981, p. 317).[1] Although the term "rationality" is of course sometimes used very broadly (e.g., as applied to actions or even to goals), Cohen intends it only in the narrow sense of "validity in deductive or probabilistic reasoning" (p. 317). His argument, in outline, is that the normative principles of reasoning and the psychological theory of reasoning competence (i.e., our internalized reasoning principles) are both generated in the same way from the same set of intuitive judgments about inferences, and so our competence must accord with the norms and therefore cannot, in principle, be faulted. That is, "our fellow humans have to be attributed a competence for reasoning validly" (p. 317).

Cohen's argument remains influential and widely discussed. It is, for example, the subject of a sustained and highly-detailed critique by Stein (1996), who objects that Cohen's argument depends on false characterizations of the processes by which reasoning norms are justified and by which psychological theories of reasoning competence are generated. Stein's critique of Cohen is an important one which merits careful discussion, but I do not propose to discuss it at

---

[1] The "bleak implications" phrase is from Nisbett and Borgida (1975), which Cohen quotes (1981, p. 317). Throughout the present paper, page references to Cohen without a year citation are all from (1981).

length here. In this paper I raise and discuss two independent and more fundamental problems with Cohen's position.

The first problem involves the fact that Cohen restricts his thesis of an unimpeachable human competence for inference to just the "normal" humans. I argue that there is no viable criterion of normalcy available to Cohen. All the candidates fail to do the work which a concept of normalcy must do in his overall argument for this thesis, thus undermining it from the beginning. The second problem arises from the fact that Cohen's argument is, in an important sense, not even really about human reasoning in the first place. It is an *a priori* epistemological argument about the limitations on any community's judgments about certain of its own practices, and ultimately it says nothing about the actual character of human reasoning. Therefore, even if his argument were granted, Cohen should not be seen as defending human reasoning from its critics, but instead as rendering human reasoning totally *indefensible*.

For readers who may wish a reminder of the psychological studies of the sort in question, the next section briefly recaps two of the basic experiments. In subsequent sections, I outline Cohen's overall argument for an unimpeachable human rationality, develop the objection concerning the concept of normalcy on which Cohen's argument relies, and discuss the nature and significance of Cohen's conclusion.

The Psychological Research

A standard example of an experiment taken by many to raise serious questions about human rationality in the area of deductive inference is the four-card problem or "selection task," first discussed by Wason (e.g., 1968) and Wason and Johnson-Laird (1970). In the experiment, subjects can see one side of each of four cards—a "4," a "7," an "A" and a "D"—which they know to have a letter on one side and a numeral on the other. They are given the following rule about the cards: *If a card has a vowel on one side, then it has an even number on the other side*. They are then told that their task is to identify which of the cards one must see the hidden side of, in order to determine whether the given rule is true or false for these cards. Wason and Johnson-Laird (and subsequent researchers) found that very few subjects (typically college students) give the correct answer "A and 7" and that the wrong answers fall into a predictable pattern, the two most frequent answers being "A and 4" and "only A." These results have been taken to suggest a defective underlying competence—for instance, one which lacks the proper rule for contraposition. (Proper contraposition would show that the rule *If a card has an odd number on one side, then it has a consonant on the other side* is equivalent, and that the grounds for examining the "A" would equally be grounds for examining the "7.")

The results appear to be robust for variations of the experiment at a similar level of abstraction (say, cards with geometrical shapes on the two sides). But when researchers vary the experiment along certain other dimensions (e.g., making the rule "If a letter is sealed, then it has a stamp on it" and switching the cards for

actual envelopes that are sealed or not and stamped or not), the results vary (e.g., Johnson-Laird, Legrenzi, & Legrenzi, 1972). It does appear that performance increases with more familiar subject matters and more realistic props, although the significance and explanation of these differences are hotly debated. Cohen (1986, p. 155), for instance, suggestively calls the selection task "Wason's four-card trick" and argues that since subjects show an ability to contrapose in the more concrete tasks, they must have the correct principle available and some distracting or illusory aspect of the abstract selection task causes them to fail to apply the principle. Others, including Stein (1996, p. 92), point out that this difference could be explained in terms of the subject lacking a proper *general* principle which can be applied to the abstract task and various concrete versions.

A standard example of an experiment involving probabilistic inference is the conjunction problem, in which subjects are asked to read a description of a person and are then instructed to rank several additional statements in order of probability (Tversky & Kahneman, 1983, p. 297). One of the descriptions is the following:

> 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.

The following statements are among those the subjects are asked to rank:

i. Linda is a teacher in an elementary school.
ii. Linda is a bank teller.
iii. Linda works in a bookstore and takes yoga classes.
iv. Linda is active in the feminist movement.
v. Linda is a bank teller and is active in the feminist movement.

A very large majority of subjects ranks (v) as *more* probable than (ii), despite the fact that this ranking flouts a basic rule of probability—that the probability of a conjunction must be less than or equal to the probability of any of its conjuncts.

In subsequent versions, researchers vary the experiment in certain dimensions, such as eliminating all the statements but (ii) and (v), or restating (ii) as "Linda is a bank teller, whether or not she is active in the feminist movement" (to rule out the possibility that subjects are interpreting [ii] as "Linda is a bank teller and is *not* active in the feminist movement"), but this does not significantly affect the proportion of subjects who rank (ii) higher than (v). Tversky and Kahneman also examine the possibility that subjects have the conjunction principle available, but somehow fail to apply it in this case. Subjects are given the Linda description, together with statements (ii) and (v), and explicit *arguments* for (ii) and for (v) being the more probable alternative respectively. The argument for (v) is that Linda is more likely to be a feminist bank teller than she is to be a bank teller because she *resembles* an active feminist more than a bank teller (1983, p. 299).[2]

---

[2] The argument for (ii) being more probable than (v) is that every feminist bank teller is a bank teller, but some women bank tellers aren't feminists, and Linda could be one of them.

Most subjects endorse this argument, and thus appear to have internalized incorrect reasoning principles, rather than simply failed to apply correct ones through some mere performance error.

The disconcertingly poor performance of subjects in a variety of such experiments has been regarded by many (e.g., Kahneman & Tversky, 1972; Nisbett & Borgida, 1975) as evidence that subjects have internalized a defective or incomplete set of inference rules. Cohen, however, argues that it is *conceptually impossible* for such experimental results to impugn human rationality and that we must instead regard the underlying inference rules as perfectly in order as they stand. Let us turn now to Cohen's argument.

The Structure of Cohen's Argument

At the heart of Cohen's argument are two claims—one about the origin of the normative principles governing reasoning and one about the origin of the descriptive psychological theory of reasoning.

Cohen claims that the norms of reasoning must answer to *intuitions*—that is, the normative principles governing human reasoning can be justified only by ultimately appealing to the intuitions of ordinary people. An intuition in this sense is not something mysterious; for Cohen, an intuition that P is "just an immediate and untutored inclination, without evidence or inference, to judge that P" (p. 318). So the intuitions involved here are immediate, untutored inclinations to judge that something follows, either deductively or probabilistically, from something else. Cohen argues that with respect to both deductive and probabilistic inference all normative criteria for the evaluation of inferences ultimately derive their authority solely from their accordance with such intuitions.

He argues this by cases—that is, by setting out to eliminate the various other candidates for sources of justification, such as formal theories of mathematical logic or probabilities, or transcendental arguments. For instance, he argues that to apply formal logic theories to everyday reasoning, one must decide how the relevant formal operators are related to natural language operators, and that this requires interpreting the natural language operators. But, he continues,

> knowing the meanings of "if," "and," and "or" is indistinguishable from knowing, in principle, their legitimate deductive liaisons. So we cannot avoid appealing to intuitions of inferential validity in order to determine the claim of an interpreted formal system to constitute a theory of deducibility for everyday reasoning. (p. 319)

It is not difficult to feel the pull of the position. After all, unless the norms of reasoning are handed down to us by a god on a mountaintop, then it may seem plausible that we must ultimately start with only our own intuitive judgments about what follows from what and build our theories from there. However, even if one were to grant that the authority of reasoning norms ultimately derives from such intuitions, it would not follow that each of our intuitions was authoritative. For instance, widely held and repeatedly elicited intuitions plausibly play a more

central role in underlying the normative principles than the isolated anomalous inference. What is needed is a principled account of how a normative theory of reasoning is to be generated from the raw data of intuitions, and Cohen sets out to describe this process.

On Cohen's account, a normative theory of inference for a population is produced by cataloging the population's intuitions about what follows from what and then idealizing this set of intuitions (throwing out the isolated oddball inference to screen for distractions and so on). This idealized set of intuitions then forms the basis for generating a set of inference rules by the process of narrow reflective equilibrium, introduced by Goodman as follows: "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" (1955, p. 67). The normative theory will simply be the set of inference principles that has the overall best fit or coherence with the idealized body of intuitions.

Cohen then considers the origin of the descriptive theory of reasoning, which "will describe a competence that normal human beings have . . . to form intuitive judgments about particular instances of right or wrong, deducibility or nondeducibility, probability or improbability" (p. 321). He asserts that psychological researchers who would construct such a theory must begin by taking as their data people's actual intuitive judgments about inferences and, in the course of constructing the theory, must also idealize by screening for anomalous inferences, etc. That is, the empirical psychological theory that predicts people's judgments will have to be generated from the same data (the idealized set of intuitions) and the same process (narrow reflective equilibrium) that generate the normative theory.

But, in light of these considerations, Cohen's thesis of unimpeachable human reasoning seems there for the taking, since the psychological theory ascribes to people a competence that consists in forming and maintaining the same set of intuitive judgments about inference that determine correctness. And if people's intuitions had been different, then the normative theory would have been correspondingly different, since the process generating it would have taken those different intuitions as its inputs. Therefore, according to Cohen, practice is guaranteed to accord with standards, for there will always be a perfect correspondence between the empirical theory of inference and the normative theory. But if reasoning competence is guaranteed *a priori* to accord with the norms of reasoning, then it cannot make any sense to speak of a faulty competence for reasoning.

> In other words, where you accept that a normative theory has to be based ultimately on the data of human intuition, you are committed to the acceptance of human rationality as a matter of fact in that area, as in the sense that it must be correct to ascribe to normal human beings a cognitive competence—however often faulted in performance—that corresponds point by point with the normative theory. (p. 321)

However, it should be emphasized that Cohen does not deny that humans are prone to a variety of systematic errors in reasoning. When Cohen denies that the results of the four-card problem, for example, impugn human rationality, he is not suggesting that the standard principles of logic texts are false or that all subjects are correct in their card selections. He is content to call the mistakes made by subjects just that—mistakes, or fallacies. What Cohen does dispute is the theoretical interpretation of these errors. While some researchers interpret systematic errors as stemming from a defective set of internalized inference rules, Cohen instead proposes a perfect underlying set of rules and endeavors to attribute all genuine errors to the existence of various performance obstacles and cognitive illusions, which inhibit the proper exercise of this flawless capacity for reasoning.[3]

Cohen's assertion, that the normative theory and the descriptive theory are both generated by narrow reflective equilibrium from the same set of intuitions, has been widely discussed in the literature and remains controversial. Stein (1996, chap. 5), for example, agrees that the norms of reasoning are justified by a process of reflective equilibrium but argues that it is not simply narrow reflective equilibrium between the intuitions of the population and the inference principles. Instead he suggests that the process might be one of wide reflective equilibrium (involving coherence with broader epistemological and metaphysical theories, etc.), or that it might involve a restricted set of intuitions (e.g., only the intuitions of inference experts), or both.[4] Stein also argues that the descriptive theory of reasoning competence must be constrained by various data beyond just the intuitions of the population—for example, results from neurophysiology, computational theory, and evolutionary theory. For instance, physical facts about the brain such as the number of neurons and their speed of operation could show some principles of reasoning not to be part of the actual competence on the grounds that they are simply physically unrealizable in humans (1996, p. 161).[5] Since, according to Stein, the normative theory is not generated from these additional scientific data as inputs, the "point by point correspondence" alleged by Cohen will not follow.

Stich, as another example, argues against Cohen by rejecting the reflective equilibrium account of the normative theory altogether. He argues that the very possibility of "strange inference principles" being in reflective equilibrium with intuitions is enough to show that this approach doesn't capture the ordinary sense of "justification." "For surely," he says, "we are not at all inclined to say that person is justified in using any inferential principle—no matter how bizarre it may be—simply because it accords with his reflective inferential practice" (1990, p. 84).

---

[3] Cohen explains away some of the alleged errors as not being errors at all, but as resulting from experimenters misapplying appropriate normative theory or applying inappropriate normative theory.

[4] Stich and Nisbett (1980) proposed a version of the expert reflective equilibrium account, which was in turn critiqued by Conee and Feldman (1983).

[5] Cherniak makes essentially the same point, saying "the antagonism between formal correctness and tractability makes the ideal competence account seem particularly unrealistic" (1986, p. 145).

These are important criticisms, but in the remaining sections I would like to develop two independent criticisms of Cohen's argument which I take to be even more fundamental. The first is raised by consideration of the question "to whom must we attribute this flawless competence, according to Cohen?"

To Whom Must We Attribute This Flawless Competence?

Cohen claims that "our fellow humans have to be attributed a competence for reasoning validly" (p. 317), but to *which* of our fellow humans does Cohen believe we must attribute this competence? Literally every human? Consideration of this question reveals problematic aspects of Cohen's argument.

Suppose we discovered an individual whose intuitions about inference were radically and systematically different from our own. This could perhaps be revealed by the subject's recalcitrant deviation from our intuitive judgments, despite elaborate and prolonged efforts to eliminate the possibility of mere performance errors. We would be forced to conclude that this person does not share our competence—that this person is "following a different program" for inference and is not rational. Surely the existence of such an anomalous reasoner is at least *possible*; it is a prospect that cannot be ruled out prior to empirical investigation. Since Cohen appears to defend human rationality, would he insist *a priori* that we endorse the competence of such an anomalous reasoner?

Cohen does not seem committed to the rationality of such an anomalous subject, since he restricts the data that he takes to generate the normative theory to just the intuitions of *normal* humans, and he accordingly restricts the scope of his "flawless competence" thesis to normal humans. This is evident in statements such as "it must be correct to ascribe to normal human beings a cognitive competence that corresponds point by point with the normative theory" (p. 321) and " . . . ordinary people cannot be regarded as intrinsically irrational in regard to any such cognitive activity" (p. 322).

What are the criteria of normalcy here? The concept of *normal human* might well involve cognitive normalcy, which would in turn have implications involving rationality. But if the concept of normalcy, on which Cohen's argument rests, turns out to presuppose rationality, then Cohen's thesis that normal humans must be regarded as rational would appear to be trivial. However, there does not at first glance appear to be such a presupposition of cognitive capacity in general, or of rationality in particular, since Cohen identifies the population of normal human reasoners simply as "adults who have not been systematically educated in any branch of logic or probability theory" (p. 317). But why should normalcy be linked to adulthood? Presumably the idea is that children have not reached cognitive maturity, and that their reasoning skills in particular may be undeveloped compared to those of adults. But, of course, not all adults have matured cognitively either; some are cognitively impaired despite qualifying biologically as adults.

By "adult" Cohen can mean either someone who has matured biologically (specified in age or by physiological development) or someone who has developed certain psychological capacities; nothing else could be relevant here. If, on the one

hand, Cohen means "adult" in the biological sense, then an anomalous reasoner of adult biological age would (if untrained in logic) automatically qualify as normal. Such a reasoner would therefore have to be regarded as fully rational, according to Cohen's conclusion that we must "accept the inherent rationality of [our] fellow adults" (p. 321). But, as discussed above, the isolated anomalous reasoner is at least possible, and so Cohen cannot define "normal" in terms of biological development.

If, on the other hand, Cohen intends "adult," and hence "normal," in the sense of psychological characteristics, then psychological criteria cannot legitimately presuppose the capacity to reason correctly. In other words, if normal people were, in effect, identified as those who are disposed to reason correctly, then Cohen's conclusion that normal humans must be regarded as rational would just be trivial, amounting to the claim that "rational humans are rational." To avoid the threatened triviality, the proposed criteria of normalcy would have to be specified neutrally— that is, independently of any prior concept of correct reasoning. But how would one specify these psychological characteristics? I believe there are only two general approaches that might be capable of underlying a normative theory of inference.

On the first approach, one would identify the criterial intuitions as those of the *majority* with respect to their neutrally described psychological characteristics (or perhaps just with respect to their reasoning capacities)—for example, simply as the psychological majority (perhaps a very large majority). So, a normal human would be one of adult age who conforms to the psychological majority (and who is untrained in formal logic).[6] This avoids the earlier problem with anomalous reasoners, for Cohen's thesis would now amount to the claim that everyone in the psychological *mainstream* would thereby count as rational.

But there is no good reason for taking membership in the psychological majority to ensure rationality, and simply to assume the rationality of the majority is to beg the central question. Nor can rationality be taken as *consisting in* conformity with majority psychological characteristics or with majority inferential capacities, for this would entail that the majority could not possibly be wrong about an inference. But it is at least *possible* for the majority of humans to endorse a bad inference. Suppose that the conjunction principle for probabilities was among the internalized inference procedures of all humans, but that fifty-percent-minus-one of the population was then exposed to brain-damaging chemicals which somehow resulted in the loss of that principle. Now, if the next two people to die happened to be among the unaffected individuals, this would give the majority to the brain-damaged group, but it is absurd to say that this chance occurrence alone would make their reasoning correct.[7] In any event, to say of an individual's

---

[6] Richard Feldman, who gave me helpful suggestions on an early version of this paper, pointed out the need to respond to this possible analysis of "normal."

[7] An advocate of Cohen's position could bite the bullet and insist that it would thereby *become* correct, but I know of no one who argues this position. Cohen, at any rate, is content to recognize that such inferences are indeed fallacious; he only argues about the interpretation of the error.

reasoning that it is in the majority is to say nothing at all about the *nature* of that reasoning—a point to which I will return below.

The second approach for specifying psychological characteristics criterial for normal inference would be to identify the normal humans as those whose inferential mechanisms (or psychological mechanisms in general) are functioning properly. Such functional accounts of organic mechanisms frequently have an evolutionary basis, so that proper function gets defined in terms of natural design. (E.g., see Wakefield, 1992, who defines "mental disorder" as the failure of a mental mechanism to perform a function for which it was designed by nature; or Boorse, 1975, who defines a "healthy" organism as one whose systems are functioning in accord with their natural design as determined by natural selection.)[8]

There are nonteleological accounts of proper function available as well, such as those offered by Wright (1973) and by Cummins (1975). On Cummin's account, for instance, when a system has a capacity that is appropriately explained by analyzing it in terms of the interaction of the capacities of various system components, then it is appropriate to identify the relevant capacity of a given component as its function.[9] So the function of the circular blade on a can-opener would be to cut the lids, since a satisfactory analytical explanation of the can-opener's capacity to open cans is possible only by appealing, in part, to the blade's capacity to cut. But although the blade has other capacities, such as serving as an RPM indicator for the motor, none of these capacities are *functions* of the blade (relative to the explanation at hand), since an appeal to such capacities plays no role in an adequate and appropriate explanation of the can-opening capacity.

The strategy of identifying normal human intuitions as those produced by properly functioning inferential mechanisms can avoid the earlier problems—that is, the inadequate correlation of statistics with norms (the majority approach), the possibility of the anomalous reasoner, and the threat of triviality. Consider the evolutionary approach. Membership in the statistical majority would not ensure rationality, since it is possible for the majority to have improperly functioning inference mechanisms. The anomalous reasoner is no problem, since a subject whose inferential mechanism is not functioning properly would not qualify as normal. Of course, proper-functioning inferential mechanisms cannot simply be identified as those that, *ceteris paribus*, produce correct inferences, for this would trivialize Cohen's thesis. But the trivialization problem is avoided if the function for which inferential mechanisms are selected by nature is independently specified, and this means identifying the fitness-enhancing qualities of these mechanisms.

The obvious candidate would be something like the tendency to produce truths and avoid falsehoods. But, as several have noted, mechanisms which reliably generate true beliefs from true premises need not always be as fitness-enhancing as alternative "shortcut" mechanisms, such as approximation strategies and heuristics which, despite being technically incorrect, can be employed by subjects to generate

---

[8] Two referees at *Behavior and Philosophy* reminded me of the need to address such proper-function accounts of normalcy.

[9] This is oversimplified. The capacity of the component would be its function relative to that particular explanation. Relative to another explanation this may not be its function.

reasonably correct conclusions quickly and easily.[10] A mechanism which does a pretty good, though not reliable, job of producing truths (or close approximations to truths), but which operates very quickly and demands relatively few cognitive resources, may well be favored by natural selection over a mechanism which does a better job of generating true beliefs but which is much more resource- and time-intensive, especially if the first mechanism tends to overestimate risks—thus erring on the side of caution. As Stich says, "from the point of view of reproductive success, it's often better to be safe (and wrong) than sorry" (1990, p. 62).

I should note that Stich and others use this point to critique certain evolutionary arguments for rationality (e.g., Quine, 1969; Fodor, 1981; Millikan, 1984) which, roughly, conclude that we must be rational because creatures who do not tend to draw true conclusions would have died out. However, I am putting this point to a different use as part of my argument that Cohen cannot cash out the concept of *normal human* as one whose inferential mechanisms are functioning properly, thus undermining his reflective equilibrium argument for rationality.

To see why an evolutionary proper-function account of normalcy is not available to Cohen, consider again the nature of the dispute over the experimental evidence. All parties agree that there are genuine systematic errors in the actual performance of many subjects; that is not contested. What is contested is the theoretical interpretation of these errors. Researchers such as Kahneman and Tversky (1973) interpret many sorts of errors as stemming from a flawed reasoning competence that lacks certain correct principles of reasoning but which includes in their stead various fallacious heuristics—for example, the representativeness heuristic, according to which people predict the outcome that appears most representative of the evidence (sometimes ignoring prior probabilities). Cohen grants that such heuristics would be fallacious but instead proposes a perfect underlying competence, that includes only correct reasoning principles, so that any genuine errors must be attributable to experimental conditions that pose obstacles to the proper exercise of this flawless underlying competence.

An analysis of normalcy in terms of proper function, and of proper function in terms natural selection, would undermine Cohen's explanation. As discussed above, it is a legitimate possibility that the selected-for features of inferential mechanisms do in fact include heuristics of the very sort at issue. Indeed, some have argued that it is not merely possible, but highly likely. According to Cherniak, for example, elementary complexity theory suggests that even some simple deductive inferences cannot be executed in a metatheoretically adequate manner, and that "our quick and dirty shortcut strategies are required to avoid intractability" (1986, p. 95).

Given this possibility, the evolutionary proper-function account of normalcy entails that *normal* (i.e., proper-functioning) humans might well have a competence which in fact incorporates some of the very sorts of heuristics which Cohen himself identifies as incorrect. (He calls the representativeness heuristic "intrinsically fallacious," for instance, p. 325.) This possibility, however, is

---

[10] See, for example, Sober (1981), Stich (1990), and Stein (1996).

irreconcilable with his thesis that normal humans must be regarded *a priori* as having a flawless competence. I should emphasize that human competence need not *actually* include such shortcut strategies in order to pose this problem for Cohen. The mere possibility is inconsistent with Cohen's position, since his view is that such flaws are *conceptually impossible*.[11] Therefore, the evolutionary proper function account of normalcy is not consistent with Cohen's overall argument.

Although the evolutionary proper-function approach is unavailable to Cohen, there are still the nonteleological accounts of proper function, such as those developed by Wright and by Cummins. I won't take up a detailed discussion of these, but here is a sketch of why these accounts of function would be similarly problematic for Cohen. Consider the Cummins-style analysis, on which a capacity possessed by the component of a system is a function of that component just in case an analysis of the system's capacity to do such-and-such accounts for that system-capacity adequately and appropriately by, in part, appealing to the component's capacity (1975).[12] On this approach, then, the function of human inferential mechanisms would be identified in terms of the role such mechanisms play in explaining the capacities of the containing systems. Here the containing system is the human organism, the component in question is the inferential mechanism, and the system capacities that we explain in terms of the components' capacities are the human's complex behavioral capacities. So, to identify the function of inferential mechanisms, we are to determine what capacities of the inferential mechanisms we need appeal to in the course of adequately and appropriately explaining the complex behavioral capacities of humans.

But which capacities are these? An obvious candidate would be something like the tendency to produce truths and avoid falsehoods. But it is far from obvious that we must appeal to mechanisms which reliably generate true beliefs in order to explain our behavioral capacities. Indeed, we may ultimately even be constrained from appealing to some such mechanisms by theoretical considerations of complexity and computability and empirical findings about the brain, as mentioned above. A capacity to process information relatively quickly, resource-unintensively, and fairly accurately might very well turn out to be capable of playing the requisite role in a satisfactory explanation of our capacities. Therefore, there is reason to think that on a Cummins-style proper-function account, some of the possible approximation strategies, inferential heuristics, and the like could qualify as functions of inferential mechanisms. That is, as with the evolutionary accounts, properly functioning inferential mechanisms could turn out to include some of these quick and dirty strategies.[13]

---

[11] It might be thought that this begs the question against Cohen by assuming some objective norms of reasoning according to which these heuristics would be fallacious, but I do not assume any such norms here. I rely only on the claim that heuristics of the sort in question could turn out to be naturally selected (and hence *normal*, because part of proper function, so-defined), and the fact that *Cohen* regards them as fallacious.

[12] Again this is oversimplified, as this would be its function relative to that particular explanation.

[13] A similar argument can be made involving the Wright analysis of functions in the obvious way.

The problem for such an account is the same as for the evolutionary accounts. That is, it would remain a legitimate possibility that normal (i.e., proper-functioning) humans have a reasoning competence that incorporates some of the same inference strategies Cohen acknowledges to be incorrect, yet this very possibility cannot be reconciled with his thesis that normal humans *must* be regarded *a priori* as rational.

Therefore, none of the candidate analyses for "normal human" are available to Cohen. Unless a viable analysis of normalcy capable of sustaining Cohen's argument can be provided (and to my knowledge there are no other candidates even available), his argument is undermined. I take this to be a more fundamental criticism of Cohen's argument than those raised by Stein and by Stich since even if reflective equilibrium were the correct method by which the norms are justified (*contra* Stich), and even if both the normative and the descriptive theories were generated by narrow reflective equilibrium from intuitions alone (*contra* Stein), there is no unproblematic way of specifying the reference class of those whose intuitions are to be the input of the process and to whom we are to attribute an unassailable competence for reasoning.

The Significance of Cohen's Conclusion

I have argued for the existence of a serious and perhaps insurmountable problem with Cohen's argument. But I think that there is also a fundamental problem involving the nature and significance of the argument's conclusion. Even if we were to grant the argument (e.g., if a workable analysis of normalcy could be provided), Cohen's conclusion, properly understood, should not make us confident of human reasoning. Rather, it should be seen as leaving us incapable of saying *anything at all* in its defense.

Cohen poses the question "Can human irrationality be experimentally demonstrated?" The answer he gives is, of course, "No." Yet he is not actually claiming that humans cannot be irrational, nor is he making the weaker claim that such irrationality cannot be demonstrated. He is advancing the conclusion, weaker yet, that ordinary humans cannot demonstrate systematic defects in ordinary human inference. This is not a claim about human rationality *per se;* rather it is about the epistemic predicament of those whose judgments (according to Cohen) set the standards of correctness.

Suppose that intelligent beings have been discovered on Io, a moon of Jupiter. A group of human researchers has been charged with studying these beings, including a team of psychologists who have been asked to describe and evaluate the inferential capacities of the Ionians. These psychologists have read their Cohen, and they reason as follows:

"When we get to Io, we will first catalogue the normal Ionian intuitions about inference.[14] We will then take these intuitions and idealize them to screen for

---

[14] Of course, I've already argued that there are problems in identifying the "normal" intuitions, but let's waive that objection for the moment.

performance errors, taking the idealized intuitions as input for the process of narrow reflective equilibrium, and thus derive the normative theory of Ionian inference. We will then generate the descriptive, empirical theory of Ionian reasoning competence, and compare this to the Ionian norms to evaluate their reasoning. But, luckily for us, Cohen has already shown that this comparison is guaranteed to reveal a perfect fit between (idealized) Ionian inferences and the norms which govern them. Our work is done, and we don't even have to leave home!"[15]

In high spirits, they put the budget to other uses and submit a report on the completely flawless reasoning competence of the inhabitants. But, of course, such a conclusion of Ionian rationality would be totally uninteresting since it does not in any way concern the actual *character* of Ionian reasoning. Following this method, our psychologists would have to pronounce them rational no matter what the actual Ionian inferences were. Cohen would presumably agree that such a verdict about the Ionians is pointless.

But what, if anything, makes Cohen's conclusion about humans any more significant than the conclusion about the Ionians? Well, we are in a different position with regard to our judgments about the Ionians and our judgments about humans. In the first case, there are two distinct issues—whether the Ionians are rational by *our* standards, and whether they are rational by their own standards. (Presumably, when our psychologists were asked to evaluate Ionian reasoning, it is evaluation relative to our norms that is sought.) In the case of our judgments about human inference, however, it seems that there are no longer two distinct issues, for asking whether humans meet "our" norms for inference appears no different from asking whether they meet "their" norms for inference.

But a sort of difference remains. When Cohen concludes that "our fellow humans have to be attributed a competence for reasoning validly," this is like an answer to the question of whether the Ionians meet their own standards, since (according to Cohen) this conclusion would *have* to be drawn regardless of the actual character of human inferences. So on his account, the psychological findings certainly do not impugn human rationality, but this has nothing to do with the character of human reasoning. Indeed, Cohen's argument would apply to any situation in which a community has a practice for which both the standards of correctness and the theory of competence ultimately derive from the population's own intuitions, and therefore his conclusion is no more about human reasoning than it is about Ionian reasoning—or about Ionian morality, for that matter.

---

[15] According to Quine's (1960) principle of charity, an apparent absurdity in translation should normally (perhaps always) be chalked up to mistaken translation rather than irrationality on the part of the other person. In evaluating Ionian inference, if, on initial translation, they said things which appeared logically absurd, we would adjust the translation to make it come out coherently. If the principle requires us to make the necessary adjustments *ad infinitum*, then it would appear to rule out an attribution of irrationality. Yet another reason for filing our report without leaving home! Of course, this all depends on whether such a principle of charity really requires unlimited adjustments and whether it is correct, questions which must remain outside the scope of the present paper.

Indeed, on Cohen's view we have no choice but to endorse our own underlying competence, yet as far as we know every civilization in the universe would condemn our competence. So even if Cohen *could* establish *a priori* that we can't find fault with human reasoning, this would not constitute a genuine defense of human rationality in the face of "bleak implications." What his argument would show is that ordinary humans can make no substantive evaluation of ordinary human reasoning at all—negative or positive. If he is right that we are fated to endorse our own reasoning regardless of its character, then we can no more meaningfully vindicate human reasoning than we can condemn it.

REFERENCES

Boorse, C. (1975). On the distinction between disease and illness. *Philosophy and Public Affairs, 5,* 49-68.

Cherniak, C. (1986). *Minimal rationality*. Cambridge, MA: MIT Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317-331.

Cohen, L. J. (1986). *The dialogue of reason*. Oxford, England: Oxford University Press.

Conee, E., & Feldman, R. (1983). Stich and Nisbett on justifying inference rules. *Philosophy of Science, 50,* 326-331.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72,* 741-764.

Fodor, J. (1981). Three cheers for propositional attitudes. In J. Fodor (Ed.), *Representations: Philosophical essays on the foundations of cognitive science* (pp. 100-123). Cambridge, MA: MIT Press.

Goodman, N. (1955). *Fact, fiction, and forecast.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P., Legrenzi, P., & Legrenzi, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395-400.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 4,* 237-251.

Millikan, R. (1984). Naturalist reflections on knowledge. *Pacific Philosophical Quarterly, 65,* 315-334.

Nisbett, R., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32*, 932-943.

Quine, W. (1960). *Word and object.* Cambridge, MA: MIT Press.

Quine, W. (1969). Epistemology naturalized. In W. Quine (Ed.), *Ontological relativity and other essays* (pp. 69-90). New York: Columbia University Press.

Sober, E. (1981). The evolution of rationality. *Synthese, 46,* 95-120.

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science.* Oxford, England: Oxford University Press.

Stich, S., & Nisbett, R. (1980). Justification and the psychology of human reasoning. *Philosophy of Science, 47,* 188-202.

Stich, S. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge, MA: MIT Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychology Review, 90,* 293-315.

Wakefield, J. (1992). The concept of mental disorder. *The American Psychologist, 47,* 373-388.

Wason, P., & Johnson-Laird, P. (1970). A conflict between selecting and eliminating information in an inferential task. *British Journal of Psychology, 61,* 509-515.

Wason, P. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20,* 273-281.

Wright, L. (1973). Functions. *Philosophical Review, 82,* 139-168.