Philosophy $2^2 3^3$: Intuitions and Philosophy — Hamilton College
Fall 2011 — Russell Marcus

Class 11 - Rationality and Experimentation
Cohen, "Can Human Irrationality Be Experimentally Demonstrated"
Replies from Evans and Pollard; Kahneman; and Stich

I. Evidence and Irrationality

In Idiotfest, we looked at a variety of experimental results concerning a phenomenon which has come to be known as irrationality.
We saw that humans tend to certain fallacious patterns of reasoning.
Tversky and Kahneman labeled three of these patterns as representativeness, availability, and adjustment and anchoring.
Other so-called cognitive biases include framing, over-reliance on emotion, and unwarranted disparagement of emotion.
It is difficult to generalize about these diverse fallacious inferences.
Many of them involve probabilities and statistical analysis.
It seems clear that human beings tend to lack command of both basic and subtle probabilities.
It is fairly easy to show, experimentally, that people commit fallacies in identifiable patterns.
The claim that such tendencies are evidence for some fundamental human irrationality is strong, and may not be supported by the evidence.
Cohen argues that the evidence we have seen does not support broad claims of human irrationality.
His defense of human rationality contains two portions.
First, he sketches a view of human rational competence.
He relies here on Chomsky's distinction between competence and performance.
Second, he shows that the experimental results about irrationality do not support the general claim that human beings are irrational.

> No doubt ordinary people often err in their reasoning, and such a mistake begins to be of scientific interest when it can be shown to instantiate some regular pattern of performance error. However, nothing in the existing literature on cognitive reasoning, or *in any possible future results of human experimental enquiry*, could have bleak implications for human rationality, in the sense of implications that establish a faulty competence (Cohen, 330).

II. Intuitions and Rational Competence

In the first half of his paper, Cohen argues that humans must be rational, by definition.
Any argument for irrationality presupposes the existence of some normative epistemic theory.
In order to claim that some inference or pattern of inferences is irrational is to assume that there is a rational set of inference patterns to which the given inference does not conform.
We have to know what is rational to infer in order to know that some inference is not rational.

Cohen acknowledges Goodman's claims about reflective equilibrium and the role of intuitions in the justification of our beliefs.

> It is necessary first of all to examine the credentials of those normative theories by reference to which investigators may legitimately evaluate the rationality or irrationality of a naive subject's inference or probability judgment. Such a normative theory, I shall argue, is itself acceptable for

> the purpose only so far as it accords, at crucial points, with the evidence of untutored intuition (Cohen, 317b).

Cohen's argument is the one we saw through the first part of the semester: there's no legitimate higher authority (foundational claims) to which we can appeal to establish our epistemic norms.
We start with intuitions and proto-theories and work toward reflective equilibrium.
Cohen's notion of intuition is exactly the philosophically appropriate one.

> An intuition that *p* is here just an immediate and untutored inclination, without evidence or inference, to judge that *p* (Cohen, 318b).

For Cohen, our intuitions form the basis for a theory of epistemic competence in precisely the same way that Chomsky invokes grammatical intuitions in the pursuit of a theory of linguistic competence.
Chomsky claimed we can ascribe our best linguistic theory (a competence) to people no matter the errors they might make in performance.
Because of the poverty of the stimulus, we must posit an innate, universal grammar.
Our best account of that universal grammar is one which arises naturally as a product of our development.
For lots of reasons, we might develop in ways that prevents our linguistic performance from matching our competence.
Similarly, our inferential performance may fall short of our competence.
Cohen uses a walking metaphor: we all know how to walk, but we all sometimes stumble.
It would be wrong to deny us knowledge of how to walk on the basis of our stumbles.
Similarly, it would be wrong to deny us rationality on the basis of our errant inferences.

To evaluate our reasoning as irrational, we have to ascribe to human reasoners an epistemic competence.
That competence may not be matched by performance.
But, we can not evaluate any individual inference without a background of epistemic competence, one which is normative in the sense that it prescribes which inferences are justifiable or rational.

> For every such normative theory, which determines how it is proper to act or reason, there is room to construct a factual theory that does take intuitions as its subject matter. This factual theory will describe or predict the intuitive judgments that formulate the data for the corresponding normative theory. It will be a psychological theory, not a logical or ethical one. It will describe a competence that normal human beings have - an ability, uniformly operative under ideal conditions and often under others, to form intuitive judgments about particular instances of right or wrong, deducibility or nondeducibility, probability or improbability. This factual theory of competence will be just as idealized as the normative theory from which it derives. And though it is a contribution to the psychology of cognition, it is a by-product of the logical or philosophical analysis of norms rather than something that experimentally oriented psychologists need to devote effort to constructing. It is not only all the theory of competence that is needed in its area. It is also all that is possible, since a different competence, if it actually existed, would just generate evidence that called for a revision of the corresponding normative theory. 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, 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* (Cohen, 321b, emphasis added).

On Cohen's Goodmanian picture, our intuitions need not be infallible.
They are just the basis for the normative theory for which we seek reflective equilibrium.
Since that normative epistemic theory is the basis for judging whether an inference is rational, it could not make sense to call human being irrational.
The worst that we can say is that our performance sometimes fails to match our competence.

> Allegations of defects in performance need to be carefully scrutinized. Some of these allegations are correct and important. But others seem to arise from a misapplication or misconception of the relevant standards of rationality by which the experimentally revealed phenomena should be judged, even when those phenomena themselves are quite robust and incontestable (Cohen, 317-8).

We need intuitions to form our epistemic norms.
We need intuitions even for developing the laws of probability.
Which probabilistic laws we choose to apply in any case has some variability.
Perhaps that was something we learned from our discussion of the Monty Hall case.
But, even if there are precise laws of probability to be chosen, the basis for that choice will be our probabilistic intuitions.

Given Cohen's Chomskyan picture, one might raise questions about how much deviation between performance and competence is allowable before the view buckles.
We can't ascribe competence to agents who fail completely to demonstrate competence.
If our reasoning is really terrible, then perhaps there is no sense in which we can ascribe competence to ourselves.
On the other hand, we would still need some normative theory from which to determine whether our reasoning really is terrible.
Thus, Cohen believes that his point is conceptual and not experimental.
Human irrationality, he says, is just not experimentally demonstrable.

III. Evaluating the Irrationality Evidence

In order to show that experimental evidence does not weigh against ascriptions of human rationality (in competence), Cohen divides the kinds of evidence used to support irrationality claims into four categories:

> Cognitive Illusion
> Tests of Intelligence or Education
> Misapplications of Appropriate Normative Theories
> Applications of Inappropriate Normative Theories

Cohen intends his categorization to be comprehensive.
His claim is that any evidence for human irrationality fits into one of these four categories.
For cases in each category, Cohen argues that we best understand the evidence for irrationality as showing only that we have some errors in performance.
So, no evidence will support the irrationality claim.

Kahneman, in his response, derides Cohen's application of this rubric.

> Cohen's categories can be used as a handy kit of invective in encounters with psychologists. He also offers subjects in psychological studies a handy kit of defenses that they may use if accused of errors: temporary insanity, a difficult childhood, entrapment or judicial mistakes - one of them will surely work, and will...restore the presumption of rationality (Kahneman 340b).

But it's worthwhile to see how Cohen uses it to support his argument.
Under the heading of cognitive illusions, Cohen discusses the Wason test, as well as some of the Tversky and Kahneman results.
We could put the Monty Hall case in this category.
Cohen's view is that we all give up our errant intuitions once the illusion is revealed, once we are shown the correct inference.
(Relatedly, check out Dan Ariely's claim that we are unable to give up our perceptual illusions even when we recognize that they are illusions, from 2:30 to 5:00 in this. It's followed by a review of the data about organ donation that Susannah related in class. The whole video is really worth watching.)

Under 'tests of intelligence or education', Cohen lists many of the errors in probabilistic reasoning we saw in Idiotfest.
This category thus seems central to the debate between defenders of rationality and its deniers.
Those who argue that humans are irrational often rely on results about our misunderstanding or misapplication of probability and statistics.
For Cohen, who distinguishes between competence and performance, such misapplication is not evidence for the irrationality thesis.
It is evidence of a lack of talent, experience, or education.

> Possession of a competence for deductive or probabilistic reasoning entails the possession of a mechanism that must include not only certain basic procedures, corresponding to a set of axioms or primitive rules for the normative system concerned, but also a method of generating additional procedures, corresponding to the proof of theorems or derived rules in that normative system. But the actual operation of this method, beyond its simplest forms, may require skills that are relatively rare, just as a particular talent is required for the discovery of proofs in logic or mathematics wherever no mechanical decision procedure is known (Cohen, 325b).

We don't claim that human beings are irrational just because we can't prove all mathematical theorems.
There are always outstanding mathematical problems on which many smart people work hard.
We accept that our limitations in intelligence do not undermine our view of ourselves as competent reasoners.
Defenders of the irrationality thesis, though, believe that our limitations in knowing and applying theories of probability and statistics are evidence for irrationality.
For example, Tversky and Kahneman invoke people's insensitivity to sample size in predicting results about dispersion as evidence for human cognitive shortcomings.
Cohen argues that proper sensitivity should not be expected without training.

> It required the genius of a great mathematician (Bernoulli 1713) to discover and prove that, if you estimate the probability of a certain characteristic's incidence in a population from its frequency in a sample, then the probability of your estimate's being correct, within a specifiable interval of approximation, will vary with the size of the sample. So it is easily understandable

> that psychological experiment finds a tendency among ordinary people, untutored in statistical theory, to be ignorant of this principle and its applications (Tversky and Kahneman 1971).

There are legions of similar human inabilities to grasp mathematical and logical facts.
Cohen argues that our ignorance of facts is one thing.
An inability to reason properly, given facts, is another.

Thus in the categories of cognitive illusions and ignorance, Cohen argues that human failings should not be taken as evidence of irrationality.
In the third and fourth categories, Cohen argues that the evidence used by those who defend the irrationality thesis does not even show that we make errors.
In these cases, people misapply an appropriate normative (epistemic) theory or apply an inappropriate theory.
These are not errors of performance, he claims.
For example (related to Cohen's, and close to my heart), philosophers often worry about students' committing the fallacy of denying the antecedent, FA.

FA $\alpha \supset \beta$
$\sim\alpha$
So, $\sim\beta$

There is a related valid inference which we can call biconditional modus tollens, or BMT.

BMT $\alpha \equiv \beta$
$\sim\alpha$
So, $\sim\beta$

People who are not well-versed in logic, and I see this especially in beginning logic students, have difficulty distinguishing the material conditional, $\supset$, from the biconditional, $\equiv$.
It may very well be the case that people who commit the fallacy FA are doing so only because they believe that they are inferring reasonably using BMT.
Indeed, even logicians and mathematicians tend to use 'if' and 'if and only if' interchangeably.

> What are to be taken as the actual, concrete premises that are represented by the initial formulas in a primitive or derived rule for natural deduction, when such rules are taken to be the norms relevant to some actual sequence of human reasoning? The mere sentences uttered do not normally constitute all of the premises conveyed by the total act of communication, since we are presumptively entitled to take the latter as including also any judgments that are implied by the act of uttering those sentences in the contextual circumstances (Cohen, 326).

People may naturally, from context, infer a biconditional where someone's language, strictly speaking, indicates a conditional.
That's a judgment about how to translate the words of others.
We naturally have to make such judgments.
But, differences in interpretation need not be taken as indicating some kind of fundamental human irrationality.

While fairly recent by philosophical standards, Cohen's article is thirty years old.
It appeared in the wake of the early studies by Kahneman and Tversky and others, in the 1970s.
Indeed, Kahneman, Tversky, Fischhoff, and Nisbett are among the invited respondents; you can read their responses among the many, many following the article.
There has been lots more recent work on irrationality, including this new article which makes me want to cancel the rest of the syllabus and just read that. (I'm kidding, but only sort of.)
It is an open question, and one which you may find worthy of pursuit, whether there are substantial improvements to either the argument for irrationality or the response in recent work.

Many of the responses to Cohen's article are worth reading.
There are even more published in the following issue of *Behavioral and Brain Sciences*; I haven't read those.
Most of the responses which appeared with the original article are quite negative.
Many are acerbic.
Some of the criticisms attack Cohen's uses of probability theory.
I've chosen three responses for you to read.
Evans and Pollard capture the view of many psychologists that Cohen's argument is unscientific.
Kahneman agrees, and I thought it would be good to see his response.
Stich is the godfather of the experimental philosophy movement, and his response foreshadows much of what is to come in the third part of our course.

IV. Response: Evans and Pollard

Cohen's central claim is that experimental evidence of poor inferential performance does not, and can not, show that human beings lack rational competence.
Evans and Pollard argue that Cohen's reliance on the competence/performance distinction renders his argument empty of interest.
Their objection relies in part on Popperian grounds.

> He adopts a philosophical stance that is impossible to refute and is of little practical relevance to the scientific study of human reasoning (Evans and Pollard 335b).

Evans and Pollard claim that Cohen's argument renders his view immune from disproof.
Even if we agree that some errors in performance don't denigrate our reasoning abilities, we might believe that at some point, some kinds of experimental evidence should weigh against our rationality.
Evans and Pollard complain that Cohen doesn't sketch what kinds of evidence would do so.
For example, Cohen rejects the claim that performance errors undermine our rationality when we are experiencing cognitive illusions because, under ideal conditions, we ignore the illusions and reason properly.

> No criteria are specified for the definition of "ideal conditions," so we can never actually know when the subject is being "rational" or when he is exhibiting an illusion (Evans and Pollard 335b).

Evans and Pollard are correct that Cohen provides no such criteria.
That's because Cohen believes that there are none.
Cohen is arguing that rationality is the backdrop against which we measure errors.

Performance errors can not be evidence for irrationality since we calibrate our scales for rational inference on our intuitions which are expressions of our performance.
Cohen is not saying that the evidence is weak.
He is claiming that no evidence could support the irrationality thesis, in principle.
Cohen's point is conceptual, not experimental.
Evans and Pollard can complain that he has presented an argument that does not admit of experimental refutation, but they can not do so without begging the question against Cohen.
If you assume that the result is experimental, then there must be ways to refute it experimentally.
If the argument is conceptual, then you can not expect any experimental refutation.

Still, Evans and Pollard's complaints that Cohen's argument seems to miss the point of the evidence seems right.
The kinds of evidence presented by the defenders of the irrationality thesis does seem troubling.
We make all sorts of errors, systematically, in identifiable conditions.
Whatever we call our propensity to err in such inferences, it's worth noting.

> It is of no practical value to consider whether such behavior can, by some philosophical device, be deemed to be "rational." It is evidence that such behavior is *undesirable*, in the sense that it is likely to produce inefficient decisions and costly errors. The point is that performance factors are not a theoretical complication to be removed in a search for rational competence. Understanding performance is the very essence of the scientific study of human inference (Evans and Pollard 336b).

Evans and Pollard's complaint, then, seems to show that they and Cohen are talking past each other.
Cohen is making a conceptual point about the amount of weight one can put on such results in pronouncing on human cognitive capacities.
Evans and Pollard are insisting that the experimental results about performance are important to gather.
There need be no inconsistency between the positions.

V. Response: Kahneman

Kahneman makes the distinction between the kind of question in which Evans and Pollard are interested and the kind of question in which Cohen is interested.

> What is the relation between the psychological study of human reasoning and the normative study of inductive inference? (Kahneman 339b).

But Kahneman believes that Cohen is fighting a straw man.
Cohen is arguing against the claim that human beings are fundamentally irrational.
Kahneman points out that no one actually makes that claim.

> The statement that "it is not the case that people are always rational" merely rejects an extreme thesis that would attribute rationality to every belief and act. This statement, however, is easily misunderstood as claiming that people are never rational. Cohen apparently misread the psychological literature in just this way, and was prompted to a superfluous defense of the human race against accusations of "deep-level irrationality" that had not in fact been made (Kahneman 340a).

Kahneman and Tversky, and others, raise profound questions about the reliability of our inferences, especially statistical, probabilistic, and inductive ones.
They argue that the unreliability of our inferences raises important worries for studies of human cognition.
Our intuitions are often misleading and insecure.
These results are, at least *prima facie*, worrisome.

> Cohen has nothing of substance to offer on these difficult issues beyond a vague message of faith, charity, and authority. Faith that all inconsistencies are apparent, none real; charity in finding an interpretation of any person's judgments that will eliminate all inconsistency; the authority of experts in competence to arbitrate remaining difficulties (Kahneman 340a).

Kahneman's criticism seems, at least in part, to repeat that of Evans and Pollard: Cohen fails to provide room for experimental evidence to weigh against claims that humans are rational.
Again, Cohen's argument is conceptual: we can't make claims of irrationality since the way in which we evaluate rationality defines rationality.
Here's Cohen:

> Certain regulative principles for theory construction, such as ideals of comprehensiveness, consistency, and simplicity, have in any case to be granted a priori status, so that in the defense of this status at least some principles of reasoning may have to be conceded an intuitive warranty... The same logical principles have to be applied within each piece of scientific reasoning about the relative merits of two or more hypotheses, so that if ever any hypothesis has to be given up in the face of adverse experience it is always a factual, rather than a logical one (Cohen 318b).

In particular, if we are faced with giving up the claim that humans are rational or the claim that some evidence shows that we are irrational, we should cede the latter, empirical, factual claim rather than the former, conceptual one.
Don't forget to check out the amusing attacks on Cohen in the last three paragraphs of Kahneman's response.

VI. Response: Stich

Where Evans and Pollard and Kahneman complain that Cohen rule out empirical refutation of his claim, Stich gets closer to Cohen's central conceptual point.

> He has simply told the wrong story about the normative theory of inference. His narrow reflective equilibrium account of the way normative theories of inference are supported or justified cannot be maintained (Stich 353-4).

Cohen takes the evidence described by the defenders of the irrationality thesis to show that our inferential performance varies, but our inferential competence, an idealized, normative epistemic theory of inference, is left alone.
Stich complains that Cohen has to show that competence is as universal as he maintains.
Stich believes that competence varies culturally and any claim for an underlying universal theory of epistemic competence has to be defended.

> It is hard to see how he is in any position to insist on this a priori. The view has about as much to recommend it as the parallel suggestion that in teaching Eliza Doolittle to speak the English of the aristocracy, Henry Higgins was simply eliminating performance errors and enabling an underlying linguistic competence to shine through (Stich 354a).

We will return to Stich's concerns about cultural variation in epistemic norms through the term, in much of the experimental philosophy that traces back to Stich's work, and directly in Stich's paper with Nisbett in our next class.