Rationality and the Psychology of Abstraction

Kenneth R. Livingston

This monograph is part of The Atlas Society’s “Objectivist Studies” series.
Editor: William R Thomas
Rationality
and the
Psychology of
Abstraction

By Kenneth R. Livingston

Kenneth Livingston is professor of psychology and director of the program in cognitive science at Vassar College. He has done original research and published extensively in developmental psychology, concept-formation, and robotics. He is currently writing *Integrating the Sciences of Mind*, an overview and synthesis of the multiple perspectives that make up cognitive science.
Objectivist Studies is a monograph series published by The Atlas Society. Founded in 1990 as the Institute for Objectivist Studies, The Atlas Society is a not-for-profit research organization for the study, development, and dissemination of Objectivism, the philosophy originated by Ayn Rand. Objectivism is a secular world view that stresses reason, individualism, respect for achievement, and liberty.

Individual monographs deal with topics relevant to Objectivism: exposition of Objectivist themes, advances in the theoretical development of the philosophy, or critical studies of the philosophy. Manuscripts are selected for publication on the basis of their value for these ends. Manuscripts submitted for publication should:

- Be double-spaced throughout, including notes.
- Conform with *The Chicago Manual of Style*.
- Be submitted in duplicate and also in electronic form.
- Include an abstract and author’s biography of no more than 250 words each.

Opinions expressed in these works are those of the authors and do not necessarily reflect the views of The Atlas Society's staff or trustees. The Society is grateful to Frank Bubb for his generous support of *Objectivist Studies*.

Objectivist Studies Number 1, Second Printing, 2010

Copyright © 1998 by Kenneth Livingston. All rights reserved.
Library of Congress Cataloging-in-Publication Data
Livingston, Kenneth, 1949–
Rationality and the psychology of abstraction/Kenneth Livingston
Includes bibliographic references (pp. 26–29)
ISBN 1-57724-018-9
Printed in United States of America

The Atlas Society
1001 Connecticut Avenue, NW, Suite 830
Washington, DC 20036
Telephone: (202) 296-7263 (AYN-RAND)
Fax: (202) 296-0771
On the web: www.atlassociety.org
Email: tas@atlassociety.org

Cover design by Douglas Hesseltine
Contents

Rationality and the Psychology of Abstraction

Kenneth R. Livingston

What Needs Explaining? ............... 6
Alternative Explanations of the Data ......................... 11
Toward a Theory of Abstraction .. 15
Abstraction and Rational Thought ............................... 19
Reconciling Competing Theories of Rationality ............. 22
How Rational an Animal? ............... 24
Notes ........................................... 26
Rationality and the Psychology of Abstraction

“Animals lead for the most part a life of nature, although in lesser particulars some are influenced by habit as well. Man has reason, in addition, and man only.”

*Aristotle, Politics, Book VII, Chapter 13.*

Since Aristotle first proposed it, the idea that human beings are distinguished from all other animals by their possession of the capacity for rational thought has had a long history in both the popular and the philosophical imaginations. The view has not always been in the ascendancy, but it has largely defined the modern period. The celebration of reason was a hallmark of the Enlightenment and it remains at the core of the respect still accorded the sciences.

But many Western intellectuals declare that we have entered a postmodern period, central to which is the rejection of reason as a defining human activity and the best means to knowledge.¹ This is a broad and sweeping movement, one that even postmodernists themselves cannot agree to define, but there is no doubt about the considerable influence it has had in the academy.² From art history to zoology, historicist discourse and narrative exposition have replaced the scientific method and rational analysis as the preferred methods for many academics.

Because the claims being made by the postmodernists concern the nature of human thought and action, it would seem particularly important to examine how psychologists have weighed into this discussion over the course of the last century, particularly since the coming of “the cognitive revolution.”³ Although excellent research on the nature of human thinking was done during the first half of the century,⁴ the problem of how to characterize the human intellect received relatively little attention in America after William James’s influence was eclipsed by the rise of behaviorism and the various psychoanalytic theories. Since the Second World War, however, the development of new research technologies, new experimental methods, and new theories of mind have reinvigorated the search for good theories of mental life.

This application of the tools of science to the study of thought has itself provoked a heated debate about just how rational we are. Some defend the Aristotelian view that we are inherently rational and could not be otherwise;⁵ some take a postmodernist stance and argue that the traditional concept of rationality as adherence to a normative
standard of good reasoning is bankrupt. In between lie efforts to make sense of the data that motivate these two positions without accepting either conclusion in its extreme form.

There is, of course, something ironic, if not perverse, about the use of reason and science as the means to the criticism of reason and science, and therefore a true postmodernist is unlikely to be drawn into this debate. But empirical demonstrations of how people appear to depart from classical norms of good reasoning certainly contribute to a devaluing of reason and rationality in our concepts of ourselves, and thus further the postmodernist agenda. Those who are not ready to accept the postmodernist implication of these data owe some explanation of them that is at least not inconsistent with a contrary view of human thought and activity.

My goal in this paper is to question whether the data do in fact require the conclusion that we are not rational animals. I have no illusion about reconciling all of the data in one short paper. Indeed, the data are so voluminous that even a comprehensive review is out of the question. What I hope to offer is an account of the nature of rationality, and of its psychological origins, that suggests a strategy for integrating several diverse theoretical accounts of the data on human reasoning. I begin by considering the desiderata for a theory of rationality, with some attention to definitions and an exceedingly brief characterization of a few empirical findings that need explaining. The major categories of theoretical explanations for those findings will then be reviewed, again quite briefly. In the next section, I propose an account of the acquisition of the ability to think rationally based on an extension of a theory of abstraction, and then conclude by outlining how this perspective on reasoning reveals several of the major theories to be complements of one another rather than competitors.

What Needs Explaining?

To what does the concept of rationality refer? Standard dictionary definitions describe rationality as the ability to reason, which is itself defined as the ability to think systematically and coherently, according to the rules of logic. That is the meaning of rationality traditionally assumed by psychologists and cognitive scientists. Economists and game theorists, however, often use the term differently and refer instead to the process of making decisions so as to maximize one's expected utility. Note that a definition of rationality as the maximization of utility does not imply any particular description of the mental processes that lead to maximization. In fact, however, two assumptions are now widespread within the community of decision theorists: (1) that decision making is bounded by certain psychological constraints (for example, the capacity of short-term memory, limitations on attention, the nature of knowledge organization in long-term memory); and (2) that these constraints render decision making quite different from a straightforward application of the laws of logic.

This perspective has influenced discussions of rationality in the psychological literature, which also generally treat rationality as a bounded or constrained affair. Many
cognitive scientists have come to believe that people are perhaps rational in the economist’s sense (they consistently maximize their utilities, or at least attempt to do so), but that people are not rational in the classical sense of reasoning according to normative rules of inference.\textsuperscript{11}

The empirical evidence that encourages this conclusion comes from many different research models, far too many to review thoroughly here. Therefore, I will confine my discussion to a small number of cases that seem to make the strongest prima facie case against the view that human thinking amounts to a process of systematic reasoning.\textsuperscript{12}

\textit{Wason’s selection task.} Among the earliest tasks used to demonstrate bias effects was the Wason four-card task.\textsuperscript{13} In this test, the subject is presented with four cards, each of which is known to have a letter on one side and a number on the other. Only one side of each card is visible to the subject; two of the cards are presented showing letters, while the other two show numbers (see figure 1).

![Figure 1](image)

Figure 1. An example of four cards that might be presented to the subject in a Wason four-card experiment.

The subject is then presented with a rule: If a card has a vowel on one side, it has an even number on the other side. The subject’s task is to say which of the cards must be turned over in order to determine whether or not the rule is violated for this particular set of cards. Note that the subject does not have to say whether or not the set of cards fits the rule; he has only to state what the conditions are that would allow one to make a valid inference about whether the rule applies.

The rule at issue in this task is of the form “if \( p \) (a vowel on one side), then \( q \) (an even number on the reverse side).” The rule is violated if the flip side of a \( p \) (a vowel) is a not-\( q \) (an odd number), or if the reverse of a not-\( q \) (an odd number) is a \( p \) (a vowel). This means that one must examine every card that displays either a \( p \) (a vowel) or a not-\( q \) (an odd number) in order to determine whether the rule is violated.

Overwhelmingly, subjects fail to note the necessity to examine cases in which the number on the card is odd (the not-\( q \) cases), even though examination of the not-\( q \) card is essential in order to test the hypothesis that the rule is correct. Of the university students who served as subjects in a series of three investigations, 79 percent made this error.\textsuperscript{14} Almost half chose both the \( p \) (the vowel) and the \( q \) (the even number) cards, and another third chose the \( p \) (vowel) card only.

\textit{The 2 4 6 problem.} Wason is also responsible for bringing to our attention another surprisingly simple situation in which subjects respond in a way that seems illogical.\textsuperscript{15} This problem is designed to resemble the process of testing a hypothesis, a process in
which scientists are supposed to engage when they explore alternative explanations for a phenomenon. The subject is told that the experimenter has in mind a rule for categorizing triplets of numbers; he is further told that the triplet “2 4 6” obeys the rule. The subject’s task is to discover, by proposing other triplets, what rule the experimenter has in mind. The experimenter responds to each proposed triplet by indicating whether it fits the rule. The subject keeps track of his hypotheses and whether or not the feedback from the experimenter is consistent with those hypotheses. When the subject is confident that he has identified the rule, he announces what the rule is. If he is correct, the game is over; otherwise, the subject is told to continue guessing.

The correct rule is very simple: “any ascending sequence of numbers.” Yet a majority of subjects propose as an explicit guess at least one rule other than the one the experimenter has in mind, and many (a large minority) never get the correct rule at all. Upon examination of the written records produced by the subjects, Wason found evidence that subjects were producing test triplets that could not possibly disconfirm the hypotheses that they were considering at the moment. For example, suppose that a subject hypothesizes that the rule is “sequences of numbers produced by adding two.” The only way to falsify this hypothesis is to propose sequences that do not increase by two (for example, 2 5 8). Yet subjects propose sequences like “4 6 8” and “11 13 15,” finding in each case, of course, that the sequence fits the experimenter’s rule. This leads them to conclude that their hypothesis is correct when in fact it is not. Wason, and many others, have concluded from these results that people have a “confirmation bias” in hypothesis testing. That is, people seek evidence to confirm rather than disconfirm their hypotheses, a pattern Wason judged to be exactly the opposite of good reasoning.

Problems in statistical reasoning. A variety of problems have been given to subjects in studies of reasoning in order to determine whether or not they understand the logical principles of probability theory. These are not mere laboratory problems, but have real-world implications.

For example, we make choices every day in situations where there is no certainty about which of several alternatives will prove most beneficial in the long run. Many laboratory demonstrations have shown that people fail to reason about these alternatives according to well-established statistical principles such as the law of large numbers or the importance of taking account of base rates in estimating probabilities. Such failures are easy to demonstrate for judgments about the characteristics of other people, for example. One widely reported task asks subjects to read the following brief description of a young woman named Linda.

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 antinuclear demonstrations.
After reading the description, the subject is asked to rank the likelihood that each of a set of eight descriptions of Linda applies to her. Most of the descriptions function as distractions from the two critical items, which are (a) Linda is a bank teller, and (b) Linda is a bank teller and is active in the feminist movement. The critical finding, as explained by Amos Tversky and Daniel Kahneman, is that a majority of subjects (roughly 85 percent) judge (b) a more likely description of Linda than (a). Yet the product of two events, each with a probability less than one, must be smaller than the probability of either of them taken alone. Tversky and Kahneman even found that a majority of Ph.D. students at the Stanford Business School, all of whom had extensive training in decision theory and statistics, made the same error.

The logic of relationships among sets. Interestingly, people generally find statistical reasoning problems relatively easy to address; they do not provoke feelings of great effort. Some problems used in the study of reasoning are experienced quite differently. Take the THOG problem, which is presented to subjects as follows. Consider a black diamond, a white diamond, a black circle, and a white circle. Suppose I have written down on a piece of paper the name of one of the colors and the name of one of the shapes. If and only if a design includes either the color I have written down or the shape I have written down, but not both, then it is a THOG. The black diamond is a THOG in this case. What can you conclude about each of the remaining three shapes?

This task requires correct understanding of the exclusive “or” (either A or B, but not both), a nonstandard interpretation of “or.” Most subjects respond incorrectly to this problem by denying that the white circle is a THOG and also claim (erroneously) that black circles and white diamonds are THOGs, or that they may be.

Effects of context and content on subject performance. Each of the results just described has been replicated many times, but some of the most informative follow-up investigations have varied in interesting ways the conditions under which the data are collected. As it turns out, success on these tasks can be improved dramatically by simply changing the content of their premises.

Consider, for example, the Wason four-card task. Presenting the problem as one of understanding implicative relationships among letters and numbers results in very poor performance, as already described. But Patricia Cheng and Keith Holyoak have shown that when the same logical structure is presented, but the content is designed to suggest that issues of permission are involved, performance increases dramatically.

For example, suppose that there are four cards and that the subject knows that the cards in question contain on one side either the word “TRANSIT” or the word “ENTERING,” and on the other side a list of diseases. The subject is told that if the word “ENTERING” appears on one side of a card, then the other side must include cholera among the diseases listed. The problem thus has exactly the same formal structure as the original Wason four-card task. Even in this form, performance is significantly better than on the original Wason task (approximately 60 percent of the subjects give the correct answer), but performance leaps to 90 percent when subjects are told that the name of a disease on the card indicates that the person described on the card has been inoculated.
against that disease. A card labeled “ENTERING” must therefore show that a person who is entering the country has been protected against cholera.

Leda Cosmides has shown similar effects using different content in a four-card task. In one study, for example, subjects are asked to evaluate a set of cards for violation of the rule that says: If you are drinking beer, then you must be over eighteen. Again, subjects show a marked increase in the degree to which they respond correctly, compared with the original Wason task.

Mode of presentation also affects performance on the THOG task. For example, consider the following list of four women and their parents:

- Tiffany (my father and Sarah)
- Karen (Bill and Sarah)
- Louisa (my father and my mother)
- Kate (Bill and my mother)

The subject is told that my father and Bill are two different men and my mother and Sarah are two different women; the subject is then asked, “If Tiffany is my half-sister, which of the others is also my half-sister?” Very few subjects fail to answer this version of the exclusive “or” problem correctly, in stark contrast to results using the THOG version.

Finally, in a broad array of tasks involving statistical reasoning, Jepson, Krantz, and Nisbett have shown that when the relevance of statistical principles is obvious, as when the problem concerns a lottery, most people reason statistically. But when the problem invites taking a personal perspective, as when people are asked whether to decide which of two courses to take based on a single visit to each class or on the basis of evaluations of many other students who took the course previously, most people do not prefer the statistical alternative.

Taken together, these results, and a great many others like them, present a problem for any simple assumption that thinking is always a straightforward application of mental logic. Certainly, these results add to the list of phenomena that any account of thinking will have to explain. In particular, they add the requirement that there be some explanation for the systematic errors that occur in human reasoning. What this research shows is not merely that people make mistakes when they think about problems. That was never in doubt. What it shows is that our mistakes are not always attributable to irregular slips in the machinery of the mind, as John Stuart Mill would have it, or random perturbations of an otherwise flawless logic machine. Many of our errors in reasoning seem to be quite consistent, suggesting the operation of systematic biasing mechanisms. On the other hand, this research does not show that all of our efforts at reasoning are illogical (quite the contrary, in fact), and any good theory of human thinking will have to account for our successes as well as our shortcomings. Lastly, that a problem of given formal structure can be made easier or harder by changing the way in which it is presented is itself an observation for which some account is due.
Alternative Explanations of the Data

Attempts to respond effectively to all of these constraints on theory building have led to a remarkable variety of models. Some theorists attempt to defend the view that there is a mental logic by showing that the evidence of bias in reasoning does not go to the heart of what subjects are doing in these studies. Others accept the implication that whatever subjects are doing, it is not reasoning according to formal logical rules, and they then propose theories about what it is that the subjects are doing instead. A brief look at a few key examples of each approach will set the stage for the consideration of the major theoretical proposal of this paper.

Thinking about individual cases. Many of the findings in the research literature seem to encourage the view that we don’t reason based on abstract rules or principles but rather that we reason based on comparisons of the current problems we confront with specific cases encountered in the past. The salient features of the current problem are compared with stored memories and the information that best aligns with the current predicament is recalled into working memory, where clues to an effective solution are considered on analogy to what worked and what didn’t in the remembered situation. This reasoning strategy does sometimes require that one find patterns of similarity between recalled and current situations at the level of analogy, a reasoning process that is not itself easily explained in a case-based way.

Thinking as heuristic. Several theorists have suggested that the various patterns of bias in reasoning, and the ways in which they are sensitive to variation in presentation of the problem, support the view that people think heuristically rather than in a rule-governed or algorithmic fashion. For example, Jonathan Evans has endorsed the view that subjects use the simple tactic of matching their responses to the structure of the language used in presenting the task. When the four-card problem is presented as “If there is a vowel on one side, then there is not an odd number on the other side,” subjects tend to select the vowel and the odd number, which is the correct response. The problem is formally equivalent to the one given above, but in this case not-q (not an odd number) is mentioned in the statement of the problem, and so subjects tend to select it.

Norman Wetherick describes a related strategy for responding to syllogisms. He observed that many subjects evaluate syllogisms correctly only when the conclusion has the same logical form as the less general premise. A subject who uses this strategy will correctly conclude from “All M are P” and “Some S are M” that “Some S are P.” But he will reach the same conclusion, erroneously, if the first two premises are “Some M are P” and “All S are M.” This is perhaps easier to see given concrete examples. Consider the following:

a. All motors have moving parts
   Some vehicles have motors
   Some vehicles have moving parts

b. Some motors have moving parts
   All vehicles have motors
   Some vehicles have moving parts
Subjects who endorse the conclusion on the right (which does not follow) also tend to draw mistaken conclusions for other syllogistic forms where using the rule of thumb that the conclusion should match the less general premise is in conflict with the correct conclusion. Not all subjects seem to use this heuristic, however, which raises the interesting problem of how to account for individual differences on tasks like these. That is a seriously neglected problem lurking in the background of most research on human reasoning and rationality, and one to which I will return later.

Other heuristic rules have been suggested to account for different kinds of errors. For example, Tversky and Kahneman have proposed that many problems involving statistical reasoning are solved by applying a heuristic of representativeness. In the problem that asks subjects to evaluate descriptions of Linda, for example, Tversky and Kahneman explain the tendency to rank the conjunct of two descriptions as more likely than one of them alone as follows: one of those descriptions (that Linda is active in the feminist movement) is highly representative of the kind of person Linda is described as being in the paragraph that introduces her. Other heuristics, like judging the likelihood of an event by how easily a response may be retrieved from memory (in other words, by its availability), have been proposed to deal with different kinds of findings of apparent irrationality.

Theories proposing that people solve problems by the application of specific heuristics often account very simply for the evidence from particular studies. And in some cases, they also have allowed researchers to make very accurate predictions of subject performance in variations on the standard reasoning tasks, at least for some subset of subjects. But this approach has the disadvantage of seeming ad hoc: if people fail to reason as formal logic prescribes, with regard to a certain type of problem, simply propose a heuristic specific to that type (that “problem domain”) and consider the phenomenon explained. Such ad hocery is risky, but the evidence is quite good that many subjects do respond to problems in domain-specific fashion, at least some of the time. What is missing from these accounts is a motivated explanation for the origins of these specific reasoning strategies, and for the reason that some people seem to employ the strategies in situations where others do not.

Reasoning by means of pragmatic schemas. Yet another proposal for explaining the patterns of success and failure in reasoning is based on the idea that thinking depends on a set of pragmatic reasoning schemas. It is similar to the heuristics strategy in that it calls upon multiple procedures to explain the range of human problem-solving abilities. But these procedures are defined with respect to types of problems in a way that is better linked to more general theories of knowledge, and thus the procedures have a somewhat less ad hoc character.

The basic idea is that problems can be classified according to the kinds of knowledge structures or schemas that they evoke. Particular problem domains are addressed by schemas for pragmatic reasons. The goal is to achieve good ends, and the best reasoning for doing so is better described as pragmatic than formally logical. In that respect, these theories are easily reconciled with the view that human beings are rational maximizers of utility, even if they are not particularly logical about it.
Two different accounts of such schemas are discussed in the literature. Cheng and Holyoak, already cited for their work on the importance of how questions are framed in four-card problems, suggest that human beings acquire knowledge structures related to certain classes of situations.\textsuperscript{30} For example, they propose the existence of a schema for reasoning about “permissions,” as in the drinking-age problems or the inoculation versions of the four-card task, and they have focused primarily on such situations in their empirical research. But they suggest alternative schemas for handling problems about causation, evidence, and obligation, to mention just a few other domains. The original Wason four-card task is difficult on this view because it lacks the right kind of content to activate one of these knowledge structures, and so it does not engage the skills that people have developed for reasoning about problems with this particular structure.

Cosmides also proposes domain-specific reasoning systems,\textsuperscript{31} but she argues that the best examples of such reasoning structures are genetically prepared, as in the case of an algorithm for detecting cheating on “the social contract.” People who are younger than the legal drinking age but nevertheless consume alcohol are covered by such Darwinian algorithms, but nothing in our evolutionary history to date has led to the development of an algorithm for answering questions reliably about letters and numbers on cards. This is supposed to explain why variation in the content of the four-card task affects subject performance as it does.

There is much to recommend the view that reasoning proceeds according to domain-specific knowledge structures. But, as with more specific heuristics, one risks the ad hoc proliferation of hypothetical reasoning schemas, although thus far the proponents of such approaches have been relatively conservative on this score. Nevertheless, as will become clear, a case can be made that domain-specific reasoning schemas are at best only a part of the story when it comes to reasoning. Before turning to the rest of that story, it is worth examining two additional ways of explaining rationality. They have in common the view that a more domain-general procedure can explain the facts about thinking; on these accounts, domain-specific schemas and multiple heuristics are not required.

**Mental models.** Philip Johnson-Laird and his colleagues propose that whatever the form of the problem one confronts, the problem solver reasons by building a cognitive model of the facts of the problem.\textsuperscript{32} The thinker invokes a distinctive token to represent each element in the situation, and invokes it in such a way as to capture the (relevant) properties of the elements represented. The relationships among the elements in the real situation are similarly captured by the relationships among the tokens in the mental model. If, for example, the problem specifies that Xs occur only when Ys are present, then no X-token will occur unless it is associated with a Y.

The purpose of building a mental model is to discover previously unrepresented meaningful implications of the facts that are currently known. Human intelligence did not evolve to manipulate contentless symbols according to their shapes. The tokens in the model, therefore, always have content, though the theory makes no claims about the exact nature of these representations. They may be language-like or they may be depictive, although some elements, like negation, are difficult to represent imaginistically. Whatever the
form, however, the structure of the model is not intended to correspond to sentences used to present a problem. Rather, the model is built in the mind to correspond to the situation it represents, regardless of how that situation is presented. But because the information-processing capacities of the mind are limited, the thinker is always seeking ways to represent the system as parsimoniously as possible.

Once the model is constructed, the thinker scans for conclusions that are not explicit in the model. If such conclusions are found, then the thinker attempts to find alternative models of the premises that falsify the conclusion. If falsifying alternatives can be found, then the conclusion is rejected; otherwise, it is accepted. If there is insufficient time, or insufficient processing capacity, to fully explore alternative models (as there may well be if the situation is complex enough), then the thinker may reach only tentative conclusions or may simply decide that nothing can be concluded.

The theory clearly predicts that items not represented in a mental model will not be used to evaluate conclusions that might otherwise follow. The normal mode of presentation of the Wason four-card task, for example, does not mention the negated consequent (the odd number), which is therefore not represented and so is not considered as relevant to an evaluation of the rule. Naturally, any presentation of the problem that highlights this element will result in improved performance. Presenting the problem as “If there is a vowel, then there is not an odd number” has exactly that effect, because the item that is negated is brought to mind in its unnegated form, and so winds up as part of the model. It is then available for inspection during both the conclusion-discovery and conclusion-evaluation phases of problem solving.

The mental models theory has been developed to a high degree of sophistication in recent years, resulting in formal computer models that make predictions of human performance, and giving rise to a complex formal notation to describe the models themselves. Indeed, at least one observer has noted that mental models have been developed to the point of being much more similar to theories of reasoning by mental logic than they are different from them. Certainly, they share in common the view that a single set of procedures will account for thinking without the need for domain-specific heuristics or schemas, and the system has been highly formalized, to the point of being mechanically implemented in a computer program. But differences do remain, as a brief review of recent attempts to defend the theory of mental logics will show.

In defense of mental logics. The view that thinking is equivalent to the logical manipulation of symbols is not new in psychology. It is, for example, at the heart of the Piagetian account of thought as a logic of classes and formal operations on those classes. That particular view of mental logic, however, is fraught with difficulties, and most defenders of mental logics no longer look to Piagetian theory for a defense of their position.

Instead, the general strategy for defenders of mental logics is to (a) propose the existence of a set of core inference procedures or routines that amount to ineliminable rules of logic, but to (b) emphasize that these core procedures operate in the context of a larger, and constraint-filled, cognitive system. Such constraints are not only unavoidable but necessary on these accounts, because unconstrained formal logics generate vast num-
bers of trivial conclusions from even the simplest of premises. For example, given the rule that A entails (A or B) for any B, A entails (A or B1), which entails (A or B1 or B2), which entails (A or B1 or B2 or B3), and so on, ad infinitum. Constraints come in the form of goals or plans that limit the range of desirable conclusions, and in the form of a limited-capacity working memory that precludes the use of exhaustive algorithmic procedures. There are several different proposals for how to defend the idea that thinking is formal-logical. These differ in such details as which procedures are taken to be essential for rational thought and what specific heuristic constraints are required. But the essential strategy is the same for all formal logical models. As evidence for their theories, proponents of this view point to the fact that computer simulations of their models do a good job of matching the performance of human subjects on a variety of reasoning tasks.

**Toward a Theory of Abstraction**

The theoretical perspectives just reviewed have two points in common, however implicitly. The first has already been discussed: people are assumed to be basically rational at least in the sense that they seek to maximize their utilities. However, only those theorists who offer a defense of mental logics take the position that this rationality is achieved by employing something like traditional logic. The second point of commonality is: whatever the process by which this goal of maximum utility is sought, it is an abstract process. Case-based models may be an exception here, at least when the current problem is very similar to ones encountered before, but even they invoke abstract processes of comparative reasoning when the situation requires the use of analogical thinking. Thus, whether the process is supposed to be heuristics, mental models, schemas, or mental logics that are deployed in the service of solving problems, the systems can be characterized as highly abstract. Both the nature and, even more important, the level of the abstraction differ among theories, but that they assume abstract representation and processing is a constant.

I want to claim, first, that what is missing in all of these theories is an explanation of what it is that makes the process of thinking abstract; and, secondly, that many of the disputes about rationality can be traced to this gap. Indeed, psychologists have paid remarkably little attention to the importance of abstraction as one of the fundamental capabilities of minds, even taking it for granted. Certainly, they have had little encouragement from philosophers of mind, many of whom agree (even while agreeing about almost nothing else) that a noncircular empiricist theory of abstraction is not possible.

One important exception is found in the work of David Kelley, who offers a specific theory of abstraction, based on prior work by Rand as part of an effort to explain the process of concept formation. Most of the literature on concepts is characterized by a focus on what a concept is and what functions it serves rather than on how it is that concepts come to be in the first place. Kelley’s work, in contrast, directly confronts the essential problem, which is to explain how it is that cognitive systems abstract away from particulars in order to establish classes or concepts. This is relevant to the question of how
thinking works, and whether it is rational, in two respects. First, given that thinking consists in operations on abstract elements in all of the theoretical accounts reviewed above, it is useful to know how those elements came to be and what properties they have in virtue of how they are acquired. Nearly all of these theories include presuppositions about how the nature of the represented elements affects the quality of thinking. Secondly, these mental processes themselves, whether heuristics, schemas, model-construction procedures, or formal rules, are also concepts, not of object-like entities but of procedures or events. A good theory of abstraction should be extendible to these sorts of concepts as well. And extending such a theory to these concepts, I believe, uncovers interesting possibilities for reconciling many features of the various models presented thus far, showing how they complement rather than compete with one another as explanations for human thinking. To fully understand the logic of the argument as applied to models of reasoning requires a closer look at what the process of abstraction is like, beginning with a theory of how concepts of concrete objects are formed.

Abstract concepts. Most of the theories of reasoning outlined earlier are in some important respect in the tradition of British empiricism. Cheng and Holyoak, Wetherick, and Johnson-Laird and Byrne are all explicit in their acknowledgment of the role of induction from experience in the acquisition of the capacity to reason. If such processes are the source of the abstract ideas that enable rational thought, then our theory of abstraction will also have to be (at least generally) within this same empiricist tradition. And that means that the process of abstraction begins with the process of perception. Indeed, the problem of abstraction could be characterized as the problem of how to progress from percepts to concepts. How does this occur?

It is in the nature of abstract ideas that they treat the set of instances that they subsume as if they were qualitatively identical, although in fact they are not. One standard way of trying to account for this conundrum is to note that similarity is comparative. That is, a given object is an instance of class A if it resembles As more than it resembles Bs. Many psychological models of concepts use algorithms based on just this sort of principle to predict how subjects will assign instances to categories. This represents an improvement over attempts to define similarity noncomparatively, but it fails to address a deeper problem, which is that any given object will be similar to A in some respect, that is, along some dimension. This leaves us with the problem of determining which dimensions are the relevant dimensions of comparison. The standard solutions to this problem, in research on concepts, are to (1) construct artificial stimuli that restrict to a very small number the dimensions along which items vary, and then tell the subject explicitly what they are; and (2) estimate the salience of each dimension after the fact using a variety of curve-fitting procedures, thereby acknowledging that one has no theory for predicting in what respect people will judge similarity.

The alternative suggested by Kelley is to substitute the notion of comparative difference for the idea of comparative similarity. In other words, assume that differentiation is prior to integration in the cognitive analysis of the perceptual field. The process of perception certainly begins with a process of discriminating objects from their backgrounds.
Indeed, developmentally, the acquisition of the capacity to distinguish objects as entities independent of the contexts in which they are encountered is among the first of the young child’s cognitive accomplishments. When multiple objects are so distinguished in the visual field, how do we treat the relationships among them? Any awareness of similarity relations among them must overcome the prior awareness of their difference. Given three objects, say, two chairs and a table, it is the greater difference between either of the chairs and the table than between the chairs themselves that is the first step in abstracting the concept of chair. It is the awareness of the lesser difference between the chairs, as compared to their differences from the table, that gives rise to the awareness of them as similar. Comparative difference, not comparative similarity, is the key to the first stage of abstraction.

It is important on this account that the difference is noticed with respect to some dimension. The differences that divide chairs from tables are experienced as qualitative relative to the degree of difference between the chairs on the dimension of shape. For the chairs, the differences seem more quantitative. Note that this makes the difference between qualitative and quantitative differences “itself quantitative—a difference in the degree of difference.” Given a chair, a table, and a lamp in the perceptual field, the more natural grouping is now the chair and table because their comparative difference is the lesser difference.

It is also important to emphasize that differences are noticeable only because the objects are commensurable, that is, they share at least one dimension in common. In the foregoing example, shape is such a dimension. For concept formation to occur, the objects in the comparison set have to be commensurable in this sense. That is, the same dimensions must be employed in all of the comparative differences if any stable sorting is to emerge. Otherwise, no distinction could be drawn between the lesser, more quantitative differences between the two chairs and the greater, more qualitative differences between the table and either of the chairs.

The first stage in the process of abstraction thus involves the ability to distinguish the degrees of difference that separate at least a trio of objects along some common dimension of perceived variation. This actually alters awareness of objects that are grouped together as a result of the process. In other words, their integration into a cognitive unit (their comparative similarity, as it were) arises out of their initial comparative differences with other items in the visual field. Note that the similarity and difference that are discussed here are perceptual in nature, not conceptual. That is, one does not have to possess an abstract concept of “similar” or “different” to get this procedure off the ground, and thus there is no circularity of argument. Relative similarity and difference are specified in the sensory array and thus reflect relationships among objects and events themselves. The sensory surfaces of a human being (or any other animal) have evolved to detect these relationships by themselves responding similarly or differently to two sensory events in a way that tracks the actual similarities or differences in the world. The fact that we understand this process by deploying the concepts of similar and different does not mean that the process itself requires these concepts to operate. This is a very brief rendering of what
I realize to be a complex theoretical argument in its own right. Because it diverges from the major point of the present paper, however, I defer elaboration to another occasion.

But this does not get us all the way to a fully abstract concept. So far, we have described how the process begins in the perceptual field, but at that level, attention is focused upon a finite number of determinate objects. Even though this set of concrete objects has been assembled as a set by a process of learning, the very fact that the process is limited to sets that can be assembled in the perceptual field limits the usefulness of the resulting mental state. It is still neither universal (that is, applicable to any instance of the category, even if previously unseen) nor abstract (that is, applicable to instances that differ qualitatively). Clearly, the heart of the process of abstraction has not yet been reached.

The key to the final stage of abstraction is the new state of mind that emerges from the perception of comparative difference in the first stage. One can see the shape of an object in isolation, without reference to the shapes of other objects. But once shape has been identified as a dimension of difference, and of comparative difference at that, we are in a position to ignore or omit from consideration the specific value that an object takes on that dimension. This is the core of the process of abstraction. It acknowledges that a pair of objects (as compared to a third) are both similar and different in the same respect. The objects of the set are identical in the sense of differing merely quantitatively, in context of their (qualitative) comparison with the contrast object.

Recent empirical evidence provides support for some version of this account. For example, note that one of the corollaries of the theory is that the degree of variation in the initial contrast set will be a determining factor in establishing the range of values over which measurements will be omitted, and thus the range of particular instances that will be subsumed under the concept. Lisbeth Fried and Holyoak have shown that subjects who learn a category distinction from instances that vary more widely in their measurements along determinate dimensions tend to take longer to learn the distinction, but generalize to a wider range of novel instances than subjects for whom the training set is more circumscribed. And this author and Janet Andrews have shown that variations in the initial contrast set in a concept learning task can alter the dimensions selected as determinate.

The theory also suggests that there should be change in the psychological similarity of a pair of items as a result of inducing that they belong to the same category. In particular, the quantitative differences between them should be ignored or at least minimized, an effect that should result in a systematic change in ratings of their similarity following category induction. This author, Andrews, and Stevan Harnad, among others, have found such an effect. People who have learned to categorize a set of objects find two examples drawn from the same category to be more similar than do people who see the same objects but have not learned to categorize them. Two items drawn from different categories do not show this effect. Indeed, two items drawn from different categories may sometimes be judged to be more different by people who have learned the categories than by people who judge the same pair of items without having learned to categorize them.

There is thus some evidence for the plausibility of the view that abstraction is a process of omitting from consideration or minimizing lesser differences among items of
the same kind while preserving awareness of the differences that distinguish between classes. The result is a psychological unit that has the properties of both abstraction (it refers identically to objects that are in fact different) and universality (it establishes a space of psychological similarities into which never-before-encountered items can be mapped and recognized as of the same type). The formation of abstract concepts is a vital part of the story of rationality, because so much of our thinking involves the manipulation of abstract concepts. But the procedures for manipulating those concepts can also be compared and categorized, and those procedural concepts have an abstract quality in the same sense that object concepts do. That is, the procedures, once abstracted, can be applied to situations other than the particular ones to which they initially applied. The rest of this paper focuses on the implications of this relatively simple insight for a theory of rationality.

Abstraction and Rational Thought

Parents and researchers alike are quite familiar with the following very illuminating scenario. A young child, between the ages of twelve and eighteen months, is seated in a highchair, ostensibly to eat a meal. For the child, however, the situation offers other opportunities as well. The child’s spoon, for example, makes a fine tool for drumming on the tray, and the spoon’s trajectory when dropped onto the floor seems to be a phenomenon of great fascination. The latter in particular can be maddening for the parent, because the child seems willing to repeat this action endlessly, with no sign of boredom.

Jean Piaget called stereotypically repetitive behaviors of this kind “tertiary circular reactions,” and saw them as a penultimate step in the progression toward truly symbolic thought. But Piaget was never able to give an account of why or how these kinds of actions become symbolic (he proposed what proved to be a purely stipulative explanation based on an unanalyzed theory of interiorization or internalization of action), perhaps because he lacked a theory of the mechanism of abstraction. Nevertheless, I want to claim, first, that Piaget was right to notice the relevance of such behavioral patterns to the emergence of certain abstract capacities; and, secondly, that the theory of abstraction just reviewed gives a way of understanding the relationship between these sorts of repetitive actions and the early emergence of the capacity to think rationally.

Briefly, the argument is as follows. Each act of dropping the spoon is a particular event, in the same way that each of two chairs is a particular object. And the trajectory of the spoon on each occasion is determinate or concrete, in just the same way that, for example, the shape of a specific chair or table is determinate. Eventual collision with the floor or some other solid object, the sounds made thereby, and so forth are also particular and concrete. Among the things that vary across episodes is the identity of the object dropped. In fact, children often seem to be willfully varying the identities of things they toss over the side of the chair. (This occurs especially when parents cease to cooperate in retrieving dropped objects.) The identity of the object dropped thus becomes a dimension of difference across episodes of dropping, and as such object-identity becomes a candidate
for omission from consideration. The result of ignoring variations in object identity is a mildly abstracted concept of dropping.

A further variation, easily observed in this situation, is also typically introduced. This often occurs in the same session in which different objects are dropped, though it is most salient, for both child and observer, when the same object is retrieved and returned to the child. This variation is in the manner of the letting go itself. The child grasps the object, palm down, and then opens the fingers. On a subsequent occasion, the palm is turned upward and then the fingers opened and the hand tilted, or the arm is used to toss the object into the air. Other variations include pushing or pulling the object to the edge of the tray, sometimes with the hand and sometimes using yet another object. These variations identify the mode of removing support for the object as a dimension of comparative difference across episodes, thereby enabling its abstraction out of the rule. Other variations are introduced by changing the location in which the experiments are performed. Cribs provide very useful drop zones, for example, and more than one infant has been known to climb intentionally to a precarious spot for the express purpose of dropping things from it. These variations allow omission from awareness of differences among places from which the dropping occurs, and thus set the stage for a still more abstract rule that might be expressed by the proposition, “If an object isn’t supported, then it falls.”

This is a very abstract rule compared to “If I open my fingers with the palm of my hand pointing down and held out over the side of my highchair, then the spoon will fall.” But it is still specific to a particular domain of events. Nevertheless, it is sufficiently abstract to allow reasoning about an uncountably large number of novel events that were not part of the training set. The child who has abstracted this rule from his experiences has acquired a limited, domain-specific reasoning strategy. Nor is this an isolated capacity. Dropping objects is but one of many kinds of episodes with reliable causal properties about which the child is learning, and in each case, a similar account can be given of how the rule is abstracted from particular episodes. For example, the process of understanding chains of causation begins even in infancy, when the child discovers that pulling on one object (for example, a string) can cause the movement of another (an attached toy). Once enough of these domain-specific causal rules are in place as abstractions, the ground is prepared for movement to still higher levels of abstraction.

And here is one of the key advantages of the theory of abstraction outlined here. Previously abstracted concepts can themselves serve as inputs to subsequent bouts of comparison and abstraction. The process is fully recursive, as abstraction itself appears to be. This is because the abstractness of a mental representation is a function of how wide the range is over which differences are disregarded. Thus, lower-level abstractions themselves, once they are sufficiently numerous and varied, become candidates for further abstraction. An active event, for example, typically precedes a predictable consequence within some relatively constrained period of time, a dimension of comparative difference that applies across causal domains and thus can be omitted from a more abstract notion of causation.
Other domains with slightly different properties give rise over the same period of time to other hierarchies of abstraction, and I propose that by adulthood these are abstracted to the level of the domain-specific strategies and heuristics described by researchers such as Evans, Wetherick, and Cheng and Holyoak. By adulthood, most people have acquired sufficiently numerous and varied domain-specific strategies at a sufficiently high level of abstraction to permit successful reasoning about most of the problems that they encounter in the course of an ordinary day. But given the right range of experiences and a suitable focus of attention, it is possible to move to still higher-level abstractions. This is because across many different domains where temporal structure of events is a relevant dimension, the rules for occurrences and actions in those domains have both an antecedent and a consequent. If you reach the age of 21, then you may drink alcohol; I am over 21, so I may drink alcohol. If an object is released, then it falls; I release this pen, then it falls. If the temperature of this paper is raised above 451°F, then it combusts; I raise the temperature of this paper above 451°F, and it bursts into flames. Omitting from consideration the differences in characteristics of antecedents and consequent, one is left with the rule of implicate for any content: “If p, then q; p, therefore q.”

Full appreciation of this relationship at the highest levels of abstraction requires still further examination of comparative differences across an even wider range of cases before one arrives at the fully abstracted truth table for implication. And historical evidence strongly supports the conclusion that the initial achievement of such high levels of abstraction is a rare and transformative event in human affairs. Greek philosophy was already centuries old before Aristotle’s genius gave us syllogistic logic. And symbolic logic, a further abstraction away from earlier logical systems, required another two millennia to emerge in modern form.

Even once invented and recorded, the most abstract of human systems of reasoning are notoriously difficult to master. Just ask any student (or teacher) of logic. Indeed, the theory of abstraction as applied to the development of the skills of abstract reasoning strongly implies that this should be so. The further one abstracts away from determinate features and particulars, the broader the range of cases that must be examined in order to discover the dimensions on which measurements will be omitted. At some point, the particulars that are being considered are no longer perceptual objects, but are themselves concepts. Their roots are still to be found in a chain that leads back to the perceptual process, but at the moment when attention must be focused for the purpose of discovering commensurability and comparative difference among a set of concepts (as opposed to a set of perceived objects), the limitations of human working memory and attention become formidable. That is no doubt why literacy, which expands the capacity and reliability of memory, precedes the development of the capacity to master formal systems of reasoning, both across the history of a culture and across the lifetime of an individual. Notational systems for abstract concepts bring the concepts back into the perceptual sphere, thus reducing memory and other cognitive processing demands.
Reconciling Competing Theories of Rationality

A theory of reasoning based in a theory of abstraction carries with it the obvious implication that one's reasoning skills will vary with training and practice. In particular, the acquisition of the kinds of formal logics that are assessed in many psychological studies of human reasoning is far from a certain thing. On the other hand, learning the somewhat more abstract rules that apply to large but not all-encompassing domains may be somewhat more likely, and still more specialized case-based models for special areas in which one has a great deal of experience are more probable still. Hence the sensitivity to the mode of presentation exhibited by most (but not all) subjects in reasoning studies. Variations in the mode of presentation usually involve variations in the level of abstraction at which the subject is asked to reason.

For example, all human beings accumulate experience with their fellow human beings, and come to understand that a person's beliefs and motives generally operate as a system. Having information about even a small subset of the system seems to increase one's ability to predict other features of a person's character. That is what people do when presented, for instance, with the Linda example. The heuristic used in this case takes advantage of some quite reasonable classification hierarchies. Linda is concerned with issues of discrimination generally, and has shown herself willing to become involved in movements. Feminism is a movement built around issues of discrimination. Prior information suggests that people are at least moderately consistent in such matters; that is, the general patterns of belief and action are relatively stable over short-to-intermediate intervals. Moreover, people are never unidimensional, defined only by their occupations, for example. So the real contrast for the reader of the Linda problem is not between “Linda is a bank teller” and “Linda is a bank teller and active in the feminist movement,” but instead between “Linda is a bank teller and not active in the feminist movement” and “Linda is a bank teller and active in the feminist movement.” One takes the conjoint probability in either case, but because the probability is higher that Linda is a feminist than that she is not, given the earlier description and the assumptions about human character, the probability is higher that she is both a bank teller and feminist than that she is a bank teller and not feminist.

Note that the procedure for assessing Linda's personality is less abstract than the conjoint probability rule from statistics, in that it is specific to the complex structure of human character. And, in fact, Tversky and Kahneman may well be correct that people typically use a representativeness heuristic as their approximation to the appropriate rule rather than the rule itself. The representativeness heuristic operates by recalling from memory information that is typical or characteristic of category members. Typicality is a rough or intuitive measure of the frequency with which a category instance has certain values on the dimensions of relevance to the category (i.e., dimensions along which differences are omitted in category induction). And because characteristic features come in sets, one highly frequent characteristic is likely to occur together with other highly frequent ones. So by attending to instance-typicality information one is actually approximating more sophisticated probability measures reasonably well in a wide variety of situations.
The rule is only an approximation, however, so it will fail on occasion, giving rise to the kinds of shift in attention that lead eventually to the discovery of the formal rules of probability theory.

Similar arguments could be made for other patterns in the data described earlier. The THOG vs. half-sisters case obviously reflects a difference in level of abstraction. And Hillary Farris and Russell Revlin have provided evidence to show that when reasoning about problems like the Wason 2 4 6 problem, subjects overwhelmingly pursue a disconfirmation strategy when given problems with familiar content (for example, factors affecting performance on a job interview), even though they do not use such strategies consistently in Wason’s more abstract task.62

To summarize, the patterns of reasoning that people acquire vary in the degree of their abstraction. Case-based strategies can be abstracted to establish domain-specific rules of thumb, which themselves may vary in their level of abstraction. Common patterns across domains allow abstraction away from narrow domains to wider ones, culminating in formal, symbolic systems. Formal logics are maximally abstract in that they reduce the elements to be entered into inferences to contentless symbols, and so may cover any discrete, formalizable problem.

One of the things that follows from looking at reasoning as a set of more or less abstract procedures is that one should expect to see variation between individuals in the mode of reasoning adopted for a given problem. After all, individuals differ in their experiences, and in the ways they attended to those experiences. Most researchers who study rationality give remarkably little attention to the problem of individual differences in performing their tasks, but when they do, the results are easily interpreted as reflecting differences in the level to which reasoning skills have been abstracted.63 One can even find the occasional offhand report in the literature indicating the operation of the abstraction process over the course of one of these research studies. For example, Kathleen Galotti, Jonathon Baron, and John Sabini observed that some (but not all) of their subjects seemed to be attending to the forms of the syllogisms they were asked to judge and, given feedback about their successes and failures, had begun to induce a simple heuristic based on the linguistic features of the syllogisms, just the kind of heuristic that Evans says would explain performance on linguistically presented problems of a certain type.64

The theory also predicts that one should expect to see variation in the level of abstraction of a single individual’s reasoning strategies, as these are employed both across time and across situations. The subject who has mastered more highly abstracted rule systems may not employ them if the domain of reasoning is a highly familiar one and a more direct reasoning strategy, not requiring translation into more abstracted form, is available and known to be reliable. Thus, one may be able to explain many failures of even statistically sophisticated subjects to use Bayes’s rule about base rates when making judgments about personality or courses of action in social situations. Note that in addition to familiarity of the domain of the problem, the time available for thinking emerges as a critical factor affecting the choice of a reasoning strategy.

Finally, since the process of induction is not infallible and is heavily dependent
on available examples and the quality of feedback from efforts to use the concepts induced, fallacious principles of reasoning may also become part of the repertoire of reasoning. The people in Wetherick’s studies, discussed earlier, who attempt to solve problems presented syllogistically by matching the form of conclusions to one of the premises have missed the essential point. Misled by a pattern that sometimes works, they will make errors of reasoning if they use this strategy consistently. Other fallacies may also be induced from experience. During childhood it is easy to induce the rule that experts and persons in authority have privileged access to good conclusions, or that what most people believe is in fact the case, but it is fallacious to conclude that an assertion is true because someone in authority or a majority says so. There is no guarantee that a more abstract reasoning schema is a good one.

Thus, one expects to see variation both across and within individuals in the kind of reasoning strategy employed to solve problems. We do not all reason at the same level of abstraction, and none of us has the luxury of always reasoning at the most sophisticated level of which he is capable. Reasoning schemas and heuristics will sometimes give the best account of how we reason, but more general and formal systems are clearly employed by many people at least some of the time. An understanding of reasoning as a set of hierarchically ordered procedures acquired as a result of cumulative abstractions, ultimately from perceptual information, paints a picture of human reasoning that is quite a bit more complex than any of those suggested by the theories reviewed earlier. It follows from the view proposed here that these theories do not have to be seen as a set from which we must choose the one correct theory. Instead, each of them describes a mode of reasoning that is characteristic of at least some acts of reasoning.

How Rational an Animal?

If this view is correct, then the goal for a comprehensive theory of reasoning is to understand what factors affect the choice of reasoning mode made by a particular person in response to a given problem. Differential experience, the capacity of working memory (which varies with factors such as age and general intelligence), the availability of external tools such as paper and pencil or computers, and constraints of time have already been mentioned as relevant factors. But it would be a mistake to overlook the fact that how (and even whether) to think about a problem is itself a matter of choice. The extent to which a person values rationality, and is concerned with reaching reasonable conclusions by procedures that can be used to justify the conclusions, may affect the choice of procedure to be used, and even whether the effort is made in the first place to master more abstract, and therefore more general, procedures.

Unfortunately, the volitional character of decisions about how best to solve a problem has not emerged as a feature of the major theories of reasoning reviewed here. Other work is beginning to appear that locates the process of reasoning firmly within the context of a person’s goals and plans, and such work may point the way to understanding
this difficult issue. But a great deal of research is still needed before it is clear how best to handle it.

Discussion of the volitional character of reasoning puts the age-old question of whether we are inherently rational in new perspective. It also raises anew the problem of whether there are normatively better and worse modes of reasoning. The process of abstraction described here occurs when the thinker attends to the world in a certain way. Apart from reflexive acts of orienting (to sudden intense events, to regions of high contrast in a particular sensory modality, and so forth), paying attention is a voluntary act. The level of abstraction in reasoning procedures ultimately achieved by a given learner will therefore be, to a very large extent, a matter of choice, a conclusion that is strongly reinforced by numerous recent demonstrations that even highly abstract rules of reasoning can be taught effectively and even rather easily. If this is correct, and I recognize the range of assumptions it makes, then it goes too far to suggest, as Aristotle did, that we are inherently rational creatures. Indeed, it suggests why those who still believe in the value of reason ought to be particularly worried about the efforts of some postmodernists to persuade us to abandon reason as our primary tool of thought. Because if the theory presented here is correct, then, all else being equal, human beings are rational just in the degree that they choose to be.
Notes


4. See, for example, R.S. Woodworth’s discussion in his *Experimental Psychology* (New York: Holt Publishers, 1938).

5. “Inherently” is not meant to imply “innately.” Many defenders of this view would argue that rationality is a set of capacities that any healthy human nervous system develops with experience. What is common to all of these theorists is the view that the nature of adult human cognition is such that it cannot help but be rational in certain core respects. See, for example, L.J. Cohen, “Can Human Irrationality be Experimentally Demonstrated?” *Behavioral and Brain Sciences*, Sept. 1981, pp. 317–70; J. Smedslund, “On the Circular Relation between Logic and Understanding,” *Scandinavian Journal of Psychology*, 11 (1970), pp. 217–19.


12. For a more detailed introduction to the literature on what have come to be known as “bias effects” in reasoning tasks, see Evans, “Bias and Rationality”; and Alvin Goldman, *Epistemology and Cognition* (Cambridge, Mass: Harvard University Press, 1986).
14. See ibid., table 1.
22. Evans, *Bias in Human Reasoning*.
25. Cognitive scientists use the term “heuristic” to refer to what are known colloquially as rule-of-thumb procedures, which generally lead to problem solutions but are not guaranteed to do so. An algorithm, on the other hand, specifies a finite series of steps that is guaranteed to lead to the solution of a problem. In the study of reasoning, heuristic procedures are usually contrasted with formal logical ones.
29. The term “pragmatic” is used here in the colloquial sense to mean “practical.” It is not intended to refer to any of the more technical ideas found in modern philosophy.


44. The following account of abstraction owes a great deal to the model presented in Kelley’s "A Theory of Abstraction."


51. Kenneth R. Livingston and Janet K. Andrews, “On the Interaction of Prior Knowledge and Stimulus Structure in Category Learning,” *Quarterly Journal of Experimental Psychology*, Feb. 1995, experiment 1. Briefly, our procedure was to select different subsets of a group of unfamiliar objects (represented in simple drawings) for initial presentation to different groups of subjects in a standard concept learning task. The subject views the objects one at a time with instructions to name the category to which it belongs. Feedback is given to indicate whether the categorization was correct. Subjects in the different experimental groups will see the same, complete set of objects eventually, but during the first few exposures, subjects in the different groups see subsets of items that vary to different degrees on the three major dimensions of...
variation. Both the pattern of errors and the subjects’ reports of their hypotheses about category structure show an effect of the comparison set on the course of concept formation. It should also be noted that in this and many other concept learning scenarios, the process of abstraction must rely heavily on memory, because the contrast objects are not all in the perceptual field at the same time. Virtually no direct research exists at present on the role of memory in abstraction.


55. Note that in contrast to Piaget’s explanation of these situations, the present theory of abstraction does not assume that the child has to engage in the action of dropping objects in order to discover the comparative differences and similarities among cases. Being actively engaged in the process might change the level of motivation and thus persistence on the task, and certainly it gives the child control over the range of variations and duration of experimentation, but these are not necessary conditions for learning the abstract rule. This gives the present theory a great advantage over the Piagetian version because the power of observational learning is well established for primates generally. In particular, the current theory explains how limbless infants born to mothers who took thalidomide during pregnancy succeeded at learning many of the same basic abstract concepts as the rest of us; internalization of actions is not required.


58. It is an empirical question whether the abstraction to the rule “If p, then q” occurs directly from examples of this kind, or whether there are intermediate stages of abstraction.


60. For example, assume a probability of .5 that Linda is a bank teller (it does not matter what value is chosen because it operates as a constant across the two cases in any event). Even if it is only slightly more likely that Linda is a feminist than that she is not (for example, .6 vs. .4), the conjoint probability of being a feminist and a bank teller is higher than that of being a bank teller and not a feminist (.3 vs. .2 using the numbers given here).


63. For two particularly interesting cases that fit the current theory nicely, see Johnson-Laird and Wason, “A Theoretical Analysis”; and Wetherick, “Human Rationality.”

64. Evans, Bias in Human Reasoning.


Objectivism, the philosophy originally developed by novelist-philosopher Ayn Rand, is a revolutionary body of ideas that challenges both the tired dogmas of the cultural Right and the empty relativism of the cultural Left. In their place it offers a new moral and social ideal. The essence of Objectivism is, in the words of Ayn Rand, “the concept of man as a heroic being, with his own happiness as the moral purpose of his life, with productive achievement as his noblest activity, and reason as his only absolute.”

Founded in early 1990, The ATLAS SOCIETY is now internationally recognized as the premier source for Objectivist research, analysis and development. Serving students, scholars and the public at large, The Atlas Society’s mission is to advance Objectivist ideas and ideals in our culture and society, and to create an open community of people who share those ideals. We are crusaders for rational individualism, and a home for rational individualists.

TAS offers a wide range of educational and social programs for the growing Objectivist community. These programs include research, training, public advocacy, courses, conferences, publications and social events. And because Objectivism is a philosophy of reason, not dogma, the Center conducts its work in a spirit of free inquiry and intellectual tolerance.

We at TAS believe that Objectivism offers the moral and philosophical basis for cultural regeneration. The ATLAS SOCIETY is laying the intellectual groundwork; finding principled solutions to the practical problems of modern living, and fostering a benevolent community rooted in respect for individual human life and liberty.

Our membership program allows individuals to invest in the future of Objectivism. A not-for-profit 501(c)(3) educational organization, TAS relies on the support of its participants, through tuition and attendance fees, and on tax-deductible donations from individuals, businesses and foundations. For more information, write to The ATLAS SOCIETY, 1001 Connecticut Avenue NW, Suite 830, Washington, DC, 20036, call 1–800–374–1776, or e-mail tas@atlassociety.org.

Explore our products and programs; participate in our events. Help us advance a revolutionary new moral ideal for the 21st century.

www.atlassociety.org