COGNITIVE SCIENCE **16**, 1–40 (1992)

The Case for Rules in Reasoning

EDWARD E. SMITH, CHRISTOPHER LANGSTON
AND RICHARD E. NISBETT
University of Michigan

A number of theoretical positions in psychology—including variants of case-based reasoning, instance-based analogy, and connectionist models—maintain that abstract rules are not involved in human reasoning, or at best play a minor role. Other views hold that the use of abstract rules is a core aspect of human reasoning. We propose eight criteria for determining whether or not people use abstract rules in reasoning, and examine evidence relevant to each criterion for several rule systems. We argue that there is substantial evidence that several different inferential rules, including modus ponens, contractual rules, causal rules, and the law of large numbers, are used in solving everyday problems. We discuss the implications for various theoretical positions and consider hybrid mechanisms that combine aspects of instance and rule models.

One of the oldest views about the nature of thought is that reasoning is guided by abstract rules of inference. This view has its origins in Plato's theories of reasoning and education, and was the rationale behind "formal discipline" approaches to education ranging from the medieval scholastics' teaching of the syllogism to the English "public school" curriculum of Latin and mathematics. In modern times, abstract inferential rules have played important roles in some of the most influential theories of cognition, including those of Newell and Simon (e.g., 1972) and Piaget and Inhelder (e.g., 1958). This blue-blood intellectual history notwithstanding, the role of abstract rules has recently come under attack from a variety of sources.

Part of the attack stems from the development of alternatives to rule-based models of thought. One class of alternatives is *instance* models, which assume that solving a problem involves the retrieval of specific instances

The research reported here was supported by a grant from the National Science Foundation (BNS-8705444) to Edward E. Smith and by grants from the National Science Foundation (BNS-8709892 and SES-8507342), the Sloan and Russell Sage Foundations, and the Army Research Institute (MDA 903-89-C-022) to Richard E. Nisbett.

We are indebted to Keith Holyoak, Michael Morris, Steven Pinker, Zenon Pylyshyn, Lance Rips, Steven Sloman, an anonymous reviewer, members of the Wednesday night seminars at the University of Michigan, and especially Douglas Medin for critiques of earlier versions of this article.

Correspondence and requests for reprints should be sent to Edward E. Smith, Department of Psychology, University of Michigan, 330 Packard Rd., Ann Arbor, MI 48104.

from memory, one or more of which is then used as an analog for the current problem. Sophisticated instance models were first developed in the study of categorization, and the key ideas of the approach have been extended to reasoning (see Medin & Ross, 1989, for a review). Thus, numerous researchers contend that deductive reasoning is more a matter of retrieving examples than of applying rules (e.g., Griggs & Cox, 1982; Manktelow & Evans, 1979; Reich & Ruth, 1982). A related development in artificial intelligence is the emergence of *case-based* reasoning models. These models assume that knowledge about a topic is partly represented by particular cases, which are stored with a relevant generalization, and which figure centrally in reasoning processes (e.g., Kolodner, 1983; Schank, 1982). Still another theoretical development that eschews rules is *connectionism*. Connectionist models contain only simple processing units, each of which sends excitatory and inhibitory signals to other units, with nothing like a rule in sight. Yet these neural-like models can often produce the same behavior as rule models (e.g., McClelland & Rumelhart, 1986).

In addition to the challenge of rival models, the rule-based approach to reasoning is at odds with certain broad intellectual movements that affect psychology. One is the evolutionary approach to behavior, which holds that much of cognition may be attributable to specific mechanisms rather than to general purpose ones like applying abstract inferential rules (e.g., Buss, 1991). Along different iines, the heuristic approach to choice and decision making that is gaining strength in decision theory and economics contends that people lack the rules necessary for normatively correct reasoning, such as the base-rate and regression principles (e.g., Kahneman & Tversky, 1973; Nisbett & Ross, 1980; Tversky & Kahneman, 1971, 1986). Work in this tradition shows, for example, that people often substitute judgments about similarity for normatively required rule-based reasoning. Although these broad trends lack the "bite" of alternative models, they have contributed to the tarnishing of the rule-based approach to reasoning.

The case for abstract rules, then, appears debatable. In this article, we try to give some direction to the debate. We propose eight criteria for deciding whether a given abstract rule is applied, where each criterion essentially embodies a phenomenon that is more readily explained by a rule-based approach than by an alternative model. We argue that use of these criteria indicates there is substantial evidence for people's use of several deductive and inductive inferential rules, all of which have in common that they are widely considered to be normatively required for correct reasoning.

TWO CONTROVERSIAL ISSUES ABOUT RULE FOLLOWING

Abstraction and Application

To appreciate what is involved in the debate about rules, we need to say what it means to claim that a person is following a rule. Note first that our

interest is in a person *following* a rule, not in a person's behavior merely *conforming* to a rule. When we fall down, for example, our behavior conforms to certain rules of physics, but no one would want to claim that we are actually following these rules. For rule following to occur, there must be a correspondence between the rule and a mental event; indeed, there should be a one-to-one correspondence between the symbols of the rule and the components of a mental event (Pylyshyn, 1990).

As a paradigm case of following an abstract rule, consider the situation where a reasoner is presented with the statements, "If Abner is over 18, then he can vote; Abner is over 18," and tries to determine what follows from these statements by using the propositional logic rule *modus ponens—If p then q; p; therefore q* (quotation marks indicate specific statements, italics indicate rules.) To say that the reasoner "follows" or "uses" modus ponens requires that the reasoner:

1. Recognize that the input is of a certain abstract kind (the input is of the form *If p then q;*), and as a consequence it is subsumed by a certain rule (modus ponens); and
2. Applies the rule to the input (instantiates *If p then q* with "If Abner is over 18 then he can vote;", *p* with "Abner is over 18," and concludes *q,* that "Abner can vote").

Step 1 establishes that *the input can be coded as an instantiation of an abstraction.* Step 2 establishes that *the rule itself is applied;* that is, variables stated in the rule (*p* and *q*) are instantiated with constants from the input (such as, "Abner is over 18"), and then another process inspects this instantiated representation and draws the appropriate conclusions. Some opponents of rules have taken issue with the claim about abstraction, whereas others are troubled by the claims about applying a rule.

Consider first the abstraction issue. A code or representation can be abstract in several different senses. It can

- contain relatively few meaning components (this is the sense in which *color* is more abstract than *red*),
- contain variables (such as *p* and *q* in modus ponens),
- have a high degree of generality,
- be relatively nonperceptual.

The four meanings are clearly interrelated. In particular, a rule that contains variables must contain relatively few meaning components (because the variables have replaced some components), and must have some degree of generality (because the variables range over certain values). In this article, we generally use the term *abstract* to mean *contains variables,* with the other three meanings typically being implied as well (exceptions to this usage will be explicitly noted).

Given this interpretation of abstraction, we note that many of those who favor instance models, and some who champion case-based models (e.g., Lewis, 1988), object to the claim that inputs are coded and processed as instantiations of abstract structures. They would not, for example, believe that our miniproblem about Abner is ever coded in terms of anything as abstract as modus ponens. And if the problem is not coded abstractly, it cannot be assimilated to an abstract rule. Hence, the contrast between rules on the one hand versus instances and cases on the other, comes down, in large part, to the question of how abstractly we represent problems (see Barsalou, 1990).

We can further illustrate this contrast with a task that has been widely used in reasoning research and that will figure prominently in this article, Wason's (1966) four-card problem (also known as the "selection" task). In the standard version of the problem, four cards are laid out displaying the symbols "E," "K," "4," and "7." Each card has a letter on one side and a number on the other. The task is to determine which of these cards needs to be turned over to determine the truth or falsity of the hypothesis: "If a card has a vowel on one side, then it has an even number on the other." In another version of the problem, the four cards read "beer," "coke," "22," and "16," and the hypothesis to be tested is: "If a person is drinking beer, s/he must be over 18." Though the two versions are formally identical, people do much better on the drinking version than on the standard version. (The correct answers are "E" and "7" in the standard version, and "beer" and "16"' in the drinking version.) According to proponents of abstract rules, people solve the four-card problem by applying rules that, though less general than propositional logic rules, still are general enough to cover various kinds of relations; in this case, the rules concern the relations involved in *permission*. Because rules concerned with permission are likely to apply to drinking but not to alphanumeric symbols, people do better on the drinking version than on the standard version of the problem (Cheng & Holyoak, 1985). In contrast, according to the proponents of instance models, people solve the four-card problem by retrieving from memory either specific episodes or domain-specific rules (like rules about drinking in particular bars) that are applicable to the problem. The drinking version of the problem is likely to retrieve either a domain-specific rule that is applicable to the current problem or a specific episode that can be analogized to the current problem, whereas the standard version is not likely to do so. That is why the drinking version leads to better performance (Griggs, 1983; Manktelow & Evans, 1979).

Unlike instance models, connectionist models are not hostile to the notion of stored abstractions per se. Some connectionist models (e.g., Hinton & Sejnowski, 1984) include units that represent entities like *animate* and *metallic,* which are abstract in the senses of containing few meaning components, being very general, and being relatively nonperceptual. Other connectionist models embrace abstractions in that they deal with the representation

of variables and variable binding (e.g., Smolensky, 1988). However, connectionist models are incompatible with the claims that a rule can be represented *explicitly* as a separate structure, and that this structure is inspected by distinct processes. This seems to be the most widely held interpretation of rule following, and it is the one we will pursue for most of this article.

Thus, the two major issues that fuel the antirule movement concern how abstractly we represent problems, and whether we process explicitly represented rules. At this time, there is comparatively little empirical data on reasoning that can be brought to bear on the rule-application issue. The abstraction issue is a different story; in this case there is a large body of relevant data on reasoning. The bulk of this article is concerned with these data, in particular with determining how well the data line up with a set of proposed criteria that can be used to distinguish abstract representations from concrete, specific ones. In the final section we return to the rule-application issue.

Need for Criteria in Dealing With the Debate

Many researchers would agree that people can reason both ways: by applying abstract rules and by analogizing to stored instances. But to go beyond this bland and uninformative generalization we need to know how to determine when people reason in each way. That is, we need agreement about what counts as evidence for abstract-rule use and what counts as evidence for instance use. In this article we will propose eight such criteria and apply them to proposals about rule systems in deductive and inductive reasoning. We claim that the criteria taken together will often suffice to resolve controversy about a given case of reasoning concerning whether abstract rules are or are not being used. Furthermore, we will argue that the existing evidence concerning these criteria establishes that people often use abstract rules when reasoning about everyday problems. Before considering the criteria, however, several constraints on the scope of the discussion and several ground rules need to be spelled out.

First, we are concerned with the use of rules in *reasoning* (i.e., evaluating a hypothesis in light of evidence). Many arguments have already been advanced for the use of abstract rules in language comprehension and production (e.g., Chomsky, 1985; Pinker & Prince, 1988), but in view of the possibility that language may be a special skill, we cannot generalize this evidence to the case of reasoning.

Another constraint is that, within the realm of reasoning, we are concerned with *inferential* rules, which, by definition, apply to multiple content domains (where content domains are different areas of knowledge that have specific properties, areas like chess or physics or adult social relations). Rules at this level include logical rules, rules for causal deduction dealing with necessity and sufficiency, contractual rules including rules for permission and obligation, statistical rules such as the law of large numbers, and decision rules such as cost-benefit rules. They are to be distinguished from

empirical rules, no matter how general that describe events in some content domain. Inferential rules are also to be distinguished from *operating principles* (Holland, Holyoak, Nisbett, & Thagard, 1986), which are immutable principles that work automatically in running the cognitive system. An example is the similarity principle, which holds that objects sharing known properties tend to share unknown ones as well. This principle plays a substantial role in reasoning but it is not clear that use of the principle involves following an explicit rule (for contrasting views on this, though, see Collins & Michalski, 1989; Smith, Lopez, & Osherson, in press). Throughout the rest of this article, when we refer to "abstract rules" we mean "abstract inferential rules."

A related constraint is that most of the abstract rules of interest are, in some sense, *natural* ones. We have in mind the kind of rule that *could* be induced by any cognitively mature human given normal experience with the environment. That is, exemplars of the rule are plentiful in everyday experience, and inducing the rule from these exemplars would require neither excessive demands on any relevant processing mechanism (e.g., short-term memory) nor coding of events in ways that are uncongenial (e.g., disjunctions, as in "a red circle or a loud tone"). Furthermore, natural rules are such that they lead to many pragmatically useful inferences. We realize that all this does not amount to a definition, but we take comfort in the fact that the notion of a natural rule, or the related notion of a natural concept, has proven exceptionally difficult to characterize formally (see, e.g., Goodman, 1955; Murphy & Medin, 1985).

Another ground rule is that we do not assume that there is always conscious awareness of the use of inferential rules. Some inferential rules may be applied only unconsciously (Nisbett & Wilson, 1977). Others may be applied some of the time with a concomitant recognition that the rule is being used.

In our discussion, we will make no attempt to distinguish a specific instance ("Abner being told that he must be over 21 to drink at Joe's Bar") from an instance-specific rule (*If Abner wants to drink at Joe's bar, he must be over 21*). We neglect this distinction in part because our concern for most of this article is with the abstraction issue, not the rule-application issue, and in part because it is not clear what empirical evidence could be brought to bear on the distinction.

Our final ground rule concerns the criteria themselves. We do not believe that any single criterion provides iron-clad evidence for the use of an abstract rule (nor does negative evidence for a single criterion establish that the rule does not operate). Rather, it is the use of multiple criteria in converging operations that can make a strong case for or against the use of a particular rule. We also make no claim that the criteria are exhaustive of those that would provide evidence for or against the assertion that an abstract rule is used for some task. They exhaust only our knowledge of criteria that actually have been examined.

EVIDENCE FOR CRITERIA OF RULE FOLLOWING

In this section we will defend the use of eight different criteria for establishing whether reasoning makes use of abstract rules and apply each criterion to relevant evidence. Three of the criteria derive from psycholinguistics, where more than in any other area an effort has been made to establish that behavior is based on rule following. Three of the other criteria involve the performance measures of speed, accuracy, and verbal report that are routinely used by experimental psychologists to examine cognitive processes. The remaining two criteria make use of training procedures to establish that highly general rules can be "inserted" by abstract training methods. We list the criteria, and then present a rationale for each of them and a discussion of its use to date.

Eight Criteria for Rule Use

Criteria Stemming from Linguistics

1. Performance on rule-governed items is as accurate with unfamiliar as with familiar material.
2. Performance on rule-governed items is as accurate with abstract as with concrete material.
3. Early in acquisition, a rule may be applied to an exception (the rule is overextended).

Performance Criteria

4. Performance on a rule-governed item or problem deteriorates as a function of the number of rules that are required for solving the problem.
5. Performance on a rule-governed item is facilitated when preceded by another item based on the same rule (application of a rule primes its subsequent use).
6. A rule, or components of it, may be mentioned in a verbal protocol.

Training Criteria

7. Performance on a specific rule-governed problem is improved by training on abstract versions of the rule.
8. Performance on problems in a particular domain is improved as much by training on problems outside the domain as on problems within it, as long as the problems are based on the same rule.

We note in advance that the criteria vary among themselves with respect to the strength of evidence they provide for rule use. This variation will become evident as we discuss the criteria. We note further that the criteria also vary with respect to how many different abstract rules they have been applied to. Consequently, for some criteria, such as Criterion 1, we will consider numerous rules, whereas for other criteria we will discuss but a single rule.

Criterion 1: Performance on Rule-Governed Items Is as Accurate with Unfamiliar as with Familiar Items

Rationale. The logic behind Criterion 1 stems from the idea that an abstract rule is applicable to a specific item because the item can be represented by some *special abstract structure* that also defines the rule (the special structure is the antecedent part of the rule). Because even novel items can possess this special structure, they can be assimilated to the rule (see Rips, 1990). Consider the phonological rules for forming plurals of English nouns. One of the rules is (roughly) of the form, *If the final phoneme of a singular noun is voiced, then add the phoneme* |z| *to it.* This rule identifies the special structure, *singular noun whose final phoneme is voiced,* and any noun—familiar, unfamiliar, or nonsense—that can be represented by this structure can be assimilated to the rule. This is why the fact that any English speaker can tell you the plural of the nonsense item "zig" is "zigz" (as in "cows") has been taken by many psycholinguists as evidence that people do indeed possess the phonological rule in question (e.g., Berko, 1958).

To see how this criterion can be applied to reasoning rules, consider again modus ponens (*If p then q; p; therefore q*). Clearly, this rule can be applied to novel items, even nonsense ones. If someone tells you that "If gork then flum, and gork is the case," you no doubt will conclude that "flum" follows. To the extent you can draw this conclusion as readily as you can with familiar material, the rule should be attributed to your repertoire.

To make the argument for rule following even stronger, it is useful to consider a sketch of a prototypical rule model (which is just an amplification of our previous comments about rule following):

> When a test item or problem is presented, it is coded in a form that is *sufficiently abstract* to lead to access of an abstract rule: Once accessed, if need be, the rule can be used for further abstract coding of the test item. The next stage is to instantiate, or bind, the variables in the rule with entities from the input. Finally, the rule is applied to yield the desired answer; that is, inspection of the instantiated representation reveals that the antecedent of the rule has been satisfied, thereby licensing the conclusion. There are therefore four stages: coding, access, instantiation (variable binding), and application.

We can illustrate the model with our "If gork then flum; gork; ?" example. When presented with this item, you might code it, in part, as an "If X, then Y" type item. This would suffice to access modus ponens. Next, you would instantiate *p* with "gork" and *q* with "flum." Then you would apply the rule and derive "flum" as an answer. Note that had you initially coded the item more superficially—say, as an "If-then claim"—this might still have sufficed to activate modus ponens, which could then have been used to elaborate the abstract coding. Though this is merely a sketch of a model, it is compatible with the general structure of rule-based models of deductive and inductive reasoning (e.g., Collins & Michalski, 1989; Rips, 1983).

With this sketch in hand we can be more explicit about how our criterion of equivalent accuracy for familiar and unfamiliar items fits with rule-following. If we assume that there is no effect of familiarity on the likelihood of coding an item sufficiently abstractly, then there will be no effect of familiarity on the likelihood of accessing an abstract rule. Similarly, if we assume there is no effect of familiarity on instantiating a rule or inspecting an instantiated representation, there will be no effect of familiarity on applying a rule. Both assumptions seem plausible, which makes the criterion plausible (i.e., familiar items should not lead to greater accuracy). Indeed, if anything, the more familiar an item is, the *less* likely it is to be coded abstractly. This is because familiarity often rests on frequency, and frequent presentations of an item might lead one to represent it in terms of its specific content.

For a criterion to be truly useful, of course, the phenomenon it describes must also be difficult to account for by a nonrule-based explanation. The major alternatives to rule models are instance models, and Criterion 1 is indeed hard to explain in terms of instances. To appreciate this point, consider a rough sketch of a prototypical instance model:

> When a test item or problem is presented, it is first coded, and this representation serves to activate stored instances from memory. The basis for access is the similarity of the test item and stored instances. One or more of the stored instances then serve as an analog for the test item. More specifically, a mapping is made between certain aspects of the retrieved instance and known aspects of the test item; this mapping then licenses the transfer of other aspects of the retrieved instance to unknown aspects of the test item. There are, therefore, three major stages: coding, access, and mapping.

This sketch of a model captures the general structure of current analogy models (e.g., Gentner, 1983; Holyoak, 1984; Holyoak & Thagard, 1989). In applying the sketch to the phenomenon captured by Criterion 1, two critical questions arise. The first is whether the representation of an instance codes the special structure of the rule, or is instead restricted to more concrete information. To illustrate, suppose you have stored an instance of the statement, "If you drive a motorcycle in Michigan, then you must be over 17"; the question of interest amounts to whether your stored instance includes information equivalent to *If p implies q; p; therefore q*. If an instance representation does include such information, then it essentially includes the rule. This strikes us not only as implausible, but also as contrary to the intended meaning of "instance." In particular, one does not think of an instance as containing variables. In what follows, then, we will assume that instances do not encode the abstraction they instantiate, though often they may encode features that are correlated with the abstraction. Thus, instance models differ from rule models not just in whether the test item accesses an instance or a rule, but also in how abstractly the test item is coded to begin

with. (A possible exception to this principle arises when people are explicitly encouraged to process the instances deeply. In experimental situations like this, there is evidence that abstractions are indeed coded, though the abstractions that have been studied are different from the inferential rules that we discuss; see Hammond, Seifert, & Gray, 1991).

The second critical question for an instance model is how to compute the similarity between the test item and the stored instance. If the similarity is computed over all features, then the model cannot explain the phenomenon of equal accuracy for familiar and unfamiliar items, because there is no guarantee that the stored instances most similar to "gork implies flum" will be useful in dealing with the test item. Perhaps "glory and fame" will be retrieved, and this conjunction is of no use in dealing with the test item. A comparable story holds for our phonological example. If overall similarity is what matters, "zig" may retrieve "zip" from memory, and the latter's plural will not work for the test item.

To salvage an instance model we must assume that the similarity between the test item and stored instance is computed over very restricted features, namely, those correlated with the special structure of the rule. Consider again a stored instance of the regulation, "If you drive a motorcycle in Michigan, then you must be over 17." The representation of this instance may well contain features corresponding to the concepts *if* and *then,* where these features are correlated with modus ponens. If such features were given great weight in the similarity calculation, a useful analog might be retrieved. There are, however, three problems with the assumption of differential weighting. First, it is ad hoc. Second, it may be wrong, as a growing body of evidence indicates that the retrieval of analogs is influenced more by concrete features, like appearance and taxonomic category, than abstract ones (e.g., Gentner & Toupin, 1986; Holyoak & Koh, 1987; Ross, 1987). Third, for some rules there may be no obvious features correlated with the rule's special structure (a good example is the law of large numbers, as we will see later). In short, when it comes to explaining the phenomenon that accuracy is as high for novel rule-based items as for familiar ones, an instance model seems to be either wrong or ad hoc. As we will see, the same conclusion holds for many of the other phenomena we consider.

Evidence About Modus Ponens

Criterion 1 supports the hypothesis that people use modus ponens. Our "if gork then flum" example suggests that we can perform extremely well on unfamiliar rule-based items.

Surprisingly, we have had difficulty locating a published experimental report that permits a comparison between performance with familiar and unfamiliar instances of modus ponens. Perhaps the closest to the mark is a study by Byrne (1989, Experiment 1). In this study, subjects were given

statements of the form, *If p then q* and *p,* and had to decide which of three possible conclusions was correct, one of them of course being *q*. Subjects' performance— which was extremely close to perfect—showed no difference between the very familiar item, "If it is raining, then we'll get wet. It is raining. ?", and the seemingly less familiar item, "If she meets her friend, then she will go to a play. She meets her friend. ?" For these data, modus ponens passes Criterion 1.

Evidence About Modus Tollens. Modus tollens is a rule in propositional logic that states, *If p then q; not q; therefore not p.* Unlike modus ponens, subjects seem to have more difficulty in applying modus tollens to unfamiliar than to familiar items. Some critical evidence comes from a study by Cheng and Holyoak (1985), which used the four-card problem described earlier. Recall that in this paradigm subjects decide which of four cases must be checked to determine the truth or falsity of a hypothesis. Cheng and Holyoak used the hypothesis, "If a letter is sealed, then it must carry a 20-cent stamp," along with four cards corresponding to "sealed," "unsealed," "20-cent," and "10-cent." Note that the hypothesis has (part of) the special structure of modus tollens with the "10-cent" card instantiating the role of *not q*. Cheng and Holyoak presented the hypothesis and choices to two groups of subjects, with one group being familiar with the hypothesized regulation and the other group not being familiar with the regulation. There were more choices of the *not q* card in the group familiar with the hypothesized regulation than in the group that was not. Hence, modus tollens fails Criterion 1, suggesting that it is not a rule that most people naturally follow.

Evidence About Contractual Rules. Cheng and Holyoak (1985) and their colleagues (Cheng, Holyoak, Nisbett, & Oliver, 1986) proposed that people have sets of abstract rules (often referred to as "schemas") that characterize contractual relations of various types. Thus, people have a set of abstract *permission* rules, which they use to understand that a certain action may be carried out only when a precondition of some kind is established. The permission rules include:

1. If action A is taken, precondition P must be satisfied.
2. If action A is not taken, precondition P need not be satisfied.
3. If precondition P is satisfied, action A can be taken.
4. If precondition P is not satisfied, action A must not be taken.

Note that this set of rules carries with it an indication of the checking procedures necessary to establish whether a permission contract has been violated: Examine cases where an action has been carried out (to establish that the precondition obtained, Rule 1), and cases where the precondition does not obtain (to establish that the action was not carried out, Rule 4). Presumably

people also have a set of abstract *obligation* rules, which they use to understand that when a certain precondition obtains, a particular action must be carried out. The rules include:

1.′ If precondition P is satisfied, action A must be taken.
2.′ If precondition P is not satisfied, action A may be taken.
3.′ If action A is taken, precondition P may or may not be satisfied.
4.′ If action A is not taken, precondition P must not be satisfied.

Again, the rules specify checking procedures to establish whether violations of an obligation contract has occurred (see Rules 1′ and 4′).

The major line of evidence establishing that people use such abstract rules comes from studies using the four-card problem. One important finding is that as long as the hypothesis being tested can be assimilated to the permission rules, the familiarity of the hypothesis has no effect on performance. For example, Cheng et al. (1986) presented subjects with the relatively unfamiliar hypothesis, "If a passenger wishes to enter the country, then he or she must have had an inoculation against cholera," along with the choices "entering," "not entering," "inoculated," and "not inoculated"; subjects were also presented with the relatively familiar hypothesis, "If a customer is drinking an alcoholic beverage, then he or she must be over 21," along with the choices "drinking," "not drinking," "over 21," and "under 21." Subjects performed as well with the unfamiliar hypothesis as the familiar one. Subjects correctly identified which cases must be checked ("entering," "not inoculated," "drinking," "under 21") and avoided checking the other cases that could not establish a violation of the hypothesis, and did so to the same extent whether the rule was familiar or not. (Note that selecting "not inoculated" or "under 18" counts as evidence for a permission rule but not for modus tollens, because other items that fit modus tollens but not the permission rule were handled poorly.)

There is similar evidence for the use of obligation rules. Again using the four-card problem, Cheng et al. (1986) presented subjects with relatively unfamiliar hypothesis that could be assimiliated to the obligation rules, such as, "If one works for the armed forces, then one must vote in the elections," along with choices like "armed forces," "not armed forces," "vote," and "not vote." They also presented subjects with somewhat more familiar hypotheses that could be assimilated to obligation, such as "If any miner gets lung cancer, then the company will pay the miner a sickness pension," along with choices like "lung cancer," "not lung cancer," "pension," and "no pension." Again, subjects performed as well with the unfamiliar hypothesis as with the familiar one.

The evidence just cited has some weaknesses. There was no independent check on the variation in familiarity, and very few items were used. Still, the evidence is suggestive. Furthermore, as noted earlier, it is difficult to con-

struct an account of these results in terms of an instance model. Such an account has to explain why it is that whatever instances are dredged up from memory by the intersection of events like "entering a country" and "having an inoculation" are just as likely to key the appropriate checking procedures as the direct memory of actual cases of drinking though less than 21 years old, not being able to drink because of being less than 21 years old, and so on.

Evidence About Causal Rules

Morris, Cheng, and Nisbett (1991) have investigated a version of Kelley's (1972) causal schema theory. Kelley's theory assumes that people have different rule sets (often referred to as "schemas") for causal situations that differ with respect to the necessity and sufficiency of the causes involved. For example, people understand that some types of causes are both necessary and sufficient (e.g., 100°C temperature causes water to boil); some types are necessary but not sufficient (exposure to the Hong Kong flu virus, together with other preconditions, causes Hong Kong flu); some types are sufficient but not necessary (lack of fuel, among other factors, causes a car to be inoperable); and some types are neither necessary nor sufficient (smoking, together with other preconditions, and among other factors, causes lung cancer).

Using the four-card problem and related paradigm, Morris et al. (1991) provided evidence that people follow such causal rules. They showed that subjects usually performed appropriate checking procedures to establish whether a given case could overturn a particular causal hypothesis. Moreover, this was true even when the hypothesis was an unfamiliar one, involving entities never encountered before by the subjects. For example, for the hypothesis, "Temperature above 1500°C causes the element Floridium to turn into a gas," most subjects understood that all four possible events ("temperature above 1500°C," "temperature below 1500°C," "Floridium in gaseous form," "Floridium in liquid form") should be checked in order to see whether the hypothesis was overturned. If we focus on the data from the more sophisticated subjects (advanced graduate students), more than 70% of their tests of unfamiliar hypotheses were completely correct, whereas only 6% of the tests would be expected to be completely correct by chance alone. Although the study lacks a comparison with the testing of familiar hypotheses, the obtained level of performance is sufficiently high to suggest that subjects (particularly sophisticated ones) were using the rules.

Evidence About the Law of Large Numbers. Nisbett, Krantz, Jepson, and Kunda (1983) argued that people have an intuitive appreciation of the law of large numbers and an ability to apply it to real-world situations. The central notion in the law of large numbers is that sample parameters approach

population parameters as a direct function of the number of cases in the sample, and as an inverse function of the degree of variability associated with the parameter. In the limiting case of no variability for the parameter, a sample of one case is adequate for an induction to the population value.

To show that people appreciate these notions as an abstract rule, Nisbett et al. (1983) asked subjects to imagine that they were visitors to a South Pacific island who were being introduced to a range of local phenomena they had never seen before. They were to imagine that they saw an unusual bird called a "shreeble," which was blue in color, and asked to estimate what percent of shreebles on the island were blue. Other subjects were asked the same question after being told to imagine they had seen either 3 or 20 shreebles, all of which were blue. Subjects' estimates were systematically affected by the number of cases. They believed that a higher fraction of shreebles were blue when examining 20 cases than when examining 3 cases, and believed that a higher fraction were blue when examining 3 cases than 1 case. In contrast, the number of cases did not affect the percentage estimates when the entities in question were members of the "Barratos" tribe and the parameter was skin color. (The modal estimated percentage to the skin-color question was 100%, even with only one case.) This pattern of findings is consistent with subjects' reports of their assumptions about variability, as they generally assumed that bird kinds are variable with respect to color, whereas isolated tribes are uniform with respect to color. Again, we see a high level of performance with relatively unfamiliar material, so high as to suggest the use of rules even though the study lacks an explicit comparison with familiar material.

Finally, it should be noted that it is difficult to explain the high level of performance by direct application of an instance model. Presumably, such a model would assume that, when told about shreebles, subjects retrieve similar instances, some particular tropical birds, for example, examine their variability with respect to color, and qualify their generalizations as a function of the presumed variability. This still leaves unexplained, however, why subjects recognize that they have to qualify generalizations more for small samples than for large ones. And it is extremely unlikely that subjects retrieve a prior problem that they had solved by applying the law of large numbers, because there are no obvious features of the shreeble problem that are correlated with that rule.

Criterion 2: Performance with Rule-Governed Items is as Accurate with Abstract as With Concrete Materials

Rationale. This criterion is similar to our first one. However, whereas Criterion 1 was concerned with unfamiliar or nonsensical items, Criterion 2 is concerned with abstract items that may in fact be very familiar. To appreciate Criterion 2, note that intuition suggests that the rule modus ponens

can readily be applied to a totally abstract item, such as "If A then B; A; therefore B." (This item is abstract in the sense of containing few features, and, possibly, in the sense of containing variables.) Good performance on this item fits with the sketch of a rule model we presented earlier, because there is no reason to expect that abstract items are less likely than concrete ones to access the modus ponens rule, and no reason to expect abstract items to fare less well than concrete ones in instantiating the rule or inspecting an instantiated representation. If anything, we might expect abstract items to be both more likely to access the rule and easier to instantiate, because abstract items are more similar to the rule than are concrete items. Note further that good performance on abstract items is quite difficult to explain in terms of an instance model, because the only thing that an abstract item and a retrieved instance can possibly have in common is the special structure of the rule. That is, the use of abstract items allows one to strip away all content but the special structure, and consequently, performance must be based on the special structure alone (Rips, 1990). For these reasons, Criterion 2 is among the most diagnostic ones we will consider.

Evidence About Modus Ponens

As for Criterion 1, intuitive evidence alone makes it plausible that modus ponens passes Criterion 2. But it is worth considering some experimental results. Evans (1977) presented each of 16 subjects four modus ponens problems of the following abstract sort:

If the letter is L, then the number is 5
The letter is L

Therefore the number is 5.

The task was to decide whether the conclusion (the statement below the line) was valid or invalid. Performance was perfect: all 16 subjects got all four questions right. Modus ponens passes Criterion 2 with flying colors.

Evidence About Modus Tollens. Comparable research shows poor performance on modus tollens using abstract material. This was the striking finding of the classic Wason (1966) article that introduced the four-card problem. Given cards labeled "A," "B," "4," and "7," and asked to turn over enough cards to test the hypothesis, "If there is an A on the front, then there is a 4 on the back," even highly intelligence subjects rarely turn over the "7" card (finding an "A" on the other side would establish the falsity of the hypothesis). This is the chief evidence against people using modus tollens.

We note in passing that this sort of negative evidence was overgeneralized by many to become evidence against formal rule systems in general, and is another component in the current popularity of instance models. The studies

on contractual and causal schemas by Cheng and her colleagues amount to a demonstration that there has been such an overgeneralization. Subjects solve problems that are syntactically identical to the Wason four-card problem so long as the content of the problem suggests a contractual or causal interpretation allowing an appropriate, abstract rule to be applied.

Evidence About Contractual Rules. Some of Cheng's work just alluded to shows good performance on the permission rule using abstract materials. In the four-card problem, Cheng and Holyoak (1985) presented subjects with the hypothesis, "If one is to take action A, then one must first satisfy precondition P," along with choices like "A," "not A," "P," and "not P." Performance on this abstract problem (61% correct) far exceeded performance on a control problem ("If a card has an A on one side, it must have a 4 on the other") that could not be assimilated to the permission rule (19% correct). Although the study lacked a direct comparison with concrete materials, the level of performance was sufficiently high to suggest the use of a rule. (See, also, Cheng & Holyoak's 1989 study of an abstract precaution rule.)

Evidence About Causal Rules. Morris et al. (1991) provided some evidence that people can accurately apply causal rules to purely abstract material. They presented subjects with causal hypotheses that were qualified with respect to necessity and sufficiency, and asked subjects to indicate whether particular states of affairs could overturn the hypotheses. For example, subjects were told that a scientist believes that "Event A causes event B," and further believes that "The occurrence of event A is the only cause of event B, and that event A only sometimes causes event B." When presented with possible patterns of events—namely, "A and B," "A and not B," "not A and B," and "not A and not B"—subjects were highly accurate in selecting those patterns that could refute the hypotheses ("not A and B"). Furthermore, a change in the causal hypothesis—say, "A causes B, but A is not the only cause of B, and A always causes B"—led to marked changes in the subjects' choice of a refuting pattern ("A and not B"). Although the study again lacked an explicit comparison to concrete materials, the high level of performance seems difficult to account for by an instance model: Over 60% of the tests of abstract hypotheses were completely correct, whereas the percentage expected by chance is only 6%.

Criterion 3: Early in Acquisition, A Rule May be Applied to an Exception (A Rule is Overextended)

Rationale. In psycholinguistics, this criterion has figured prominently in studies of how children master the regular past-tense form of English verbs.

The relevant rule is to add "ed" to the stem of verbs to form the past tense, such as "cook–cooked." A finding that has been taken as evidence for following this rule is the tendency of young children to overextend the rule to irregular forms, such as "go–goed," even though they had previously used the irregular form correctly (Ervin, 1964). The rule specifies a special structure—the stem of a verb—and the phenomenon arises because children apply the rule to items containing the special structure even though the items should have been marked as exceptions. In terms of our sketch of a rule model, early in acquisition, exceptional verbs are likely to be represented in a way that accesses the relevant rule, and once the rule is accessed it is instantiated and applied.

Perhaps for more than any other criterion, there has been a concerted effort to formulate nonrule-based accounts of overextension. Thus, Rumelhart and McClelland (1987) offered a connectionist account of the overextension of the past-tense "rule," and others offered instance-based accounts of apparent overextensions of classification rules (see, e.g., Medin & Smith, 1981). In general, then, this criterion seems less diagnostic than the previous two we considered. We include it, though, because it may prove to be diagnostic in specific cases. Indeed, with regard to overextension of the past-tense rule, critiques of the Rumelhart and McClelland proposal by Pinker and Prince (1988) and Marcus et al. (1990) suggest that a rule-based theory still provides the fullest account of the data. The critics noted, for example, that children are no more likely to overgeneralize an irregular verb that is similar to many regular ones than to overgeneralize an irregular verb that is similar to few regular ones. Yet, in most connectionist models, as in instance models, generalization is based on similarity. The lack of similarity effects fits perfectly with a rule-based account, of course. Thus, in situations where the likelihood of overgeneralizing an exception does not depend on the similarity of the exception to the regular cases, the criterion is indeed diagnostic.

Evidence About the Law of Large Numbers. The overextension criterion has rarely been applied to abstract rules. An exception is the law of large numbers. Fong, Krantz, and Nisbett (1986) trained subjects on this rule, and found they sometimes applied it to cases where it was inappropriate. For example, in one problem presented after training, subjects were told about a basketball talent scout who watched a particular prospect through two games and concluded that he had excellent skills but a tendency to misplay under extreme pressure. The former inference was based on nearly 2 hours of play, the latter on a single episode. Trained subjects were more likely than controls to assert correctly that the "pressure" doagnisis was based on too little evidence, but were also more likely to assert incorrectly that the global judgment of excellent skills was similarly based on too little

evidence. Thus, trained subjects sometimes overextended the rule to cases where it was not appropriate. The fact that this kind of overextension occurred lends credibility to the claim that a rule corresponding to the law of large numbers was indeed being followed, especially because virtually no control (untrained) subjects expressed the view that a larger sample would have been helpful in assessing the prospect's skills.

An instance model has difficulty explaining this specific phenomenon. According to such a model, an overextension would occur whenever the basketball problem retrieves a stored problem that just happened to utilize the law of large numbers. But such problems might be very diverse, with few if any sharing content with the basketball problem. Therefore, the only way to insure that the basketball problem retrieves a useful analog is to make the problematic assumption that retrieval is heavily based on the features correlated with the special structure of the rule.

Criterion 4: Performance on a Rule-Governed Problem Deteriorates as a Function of the Number of Rules that are Required to Solve the Problem

Rationale. Criterion 4 essentially holds that rules provide the appropriate unit for measuring the complexity of a problem. We can illustrate the criterion by considering problems that vary in the number of times they require application of the rule modus ponens. Even after equating for reading time, deciding that Argument 2 is valid presumably would take longer and be more error prone than deciding that Argument 1 is valid, because Argument 2 requires one more application of modus ponens:

1. If it's raining, I'll take an umbrella
   It's raining

   I'll take an umbrella

2. If it's raining, I'll take an umbrella
   If I take an umbrella, I'll lose it.
   It's raining

   I'll lose an umbrella

(Our example might suggest that the phenomenon is an artifact of the premises being more complex in Argument 2 than in Argument 1; however, using correlational techniques, Rips, 1983, found no evidence that premise complexity per se affects the accuracy of reasoning.)

The phenomenon of interest follows from our sketch of a rule model as long as one or more of the stages involved—coding, access, instantiation, and application—is executed less efficiently when it has to do $n + 1$ things than just n things. As many theorists have pointed out, this vulnerability to

sheer number may disappear with extended practice. In Anderson's (1982) rule-based model of cognitive skills, for example, rules that are frequently applied in succession come to be "compiled" or chunked into a single rule; in such a case, performance would be rule-based yet fail to meet Criterion 4. The diagnosticity of this criterion is further reduced by the fact that the basic phenomenon involved seems roughly compatible with an instance model: What needs to be assumed is that problems that supposedly require more rules are really just problems that generally have fewer or less accessible analogues in memory. Again, though, we include the criterion because it may prove very diagnostic in certain cases, for example, in cases where there is a *linear* relation between the number of rules that a problem requires and the reaction time needed to solve the problem. Also, the criterion has a history of use in evaluating rule-based hypotheses. For example, in psycholinguistics, it figured centrally in testing the hypothesis that the complexity of a sentence was an increasing function of the number of transformational rules needed to derive the surface form of the sentence (Miller, 1962).

Evidence About Modus Ponens. We know of no direct application of Criterion 4 like our double modus ponens example. Rather than being applied to a single rule used a varying number of times, the criterion has been applied to a set of rules. Osherson (1975), Rips (1983), and Braine, Reiser, and Rumain (1984) all applied the criterion to proposed sets of logical rules that include modus ponens along with a dozen or so other rules from propositional logic (such sets are capable of determining the validity of most arguments in propositional logic and hence, constitute relatively complete theories of people's logical capabilities). The work of these investigators shows that there is a monotonic relation—and sometimes a linear one—between the number of rules needed to determine whether an argument is valid and the reaction time and accuracy of the final response. Insofar as modus ponens is a rule in the systems of all three investigators, there is indirect evidence for the use of modus ponens.

Criterion 5: Performance on a Rule-Based Item is Facilitated When Preceded by Another Item Based on the Same Rule (Application of a Rule Primes Its Subsequent Use)

Rationale. The rationale for this criterion is that, once used, a mental structure remains active for a brief time period and during this period the structure is more accessible than usual. In terms of our rule model, the access stage has been facilitated. (Anderson, 1982, made a similar assumption relating recency of rule use to ease of subsequent access.) Our sketch of an instance model would be able to account for the phenomenon to the extent that successively presented rule-based items are also similar in content;

but as we will see, the plausibility of this account depends on the specific findings involved.

Evidence About Contractual Rules. As far as we know, the priming criterion has been applied only to contractual rules. In a study we performed recently (Langston, Nisbett, & Smith, 1991), subjects were presented on each trial with a different version of the four-card problem. Sometimes the version conformed to a permission rule (*If precondition P is satisfied, action A can be taken*), whereas other times it conformed to an obligation rule (*If precondition P is satisfied, action A must be taken*). It was therefore possible to have successive trials in which the permission rule would be used twice, as illustrated in Argument 3 below, as well as successive trials in which the permission rule is used only once, as in Argument 4:

3a. If a journalist has a press pass, she can cross a police line
  b. If a journalist gets a statement on the record, she can quote her source
4a. If a journalist is a member of the union, she must pay dues
  b. If a journalist gets a statement on the record, she can quote her source

Subjects made more correct responses in testing the rule in 3b than in 4b. The same permission rule was involved in both cases, but was primed only in 3b. (Repetition of the rule was confounded with repetition of the word "can," but as we will see later, repetition of "can" alone has no effect.)

It might seem that an instance model can readily explain these results. All that need be hypothesized is that subjects use the previous item as an analog for the current problem they are working on. This would lead to a correct response for 3b and an incorrect one for 4b. This predicts, however, that errors on permission (obligation) problems would be correct responses, had the problems in fact been obligation (permission) problems. This prediction was not supported. There is, however, another aspect of the Langston et al. results that does suggest a role for instances. Langston et al. found priming effects only for items similar in content (as in 3). If the priming item shared little content with the target item, there was no improvement either in accuracy or latency (even though the word "can" was repeated). Nisbett (1991) found a similar failure of semantically unrelated items to produce priming of the law of large numbers in an untimed problem-solving situation. The fact that rule-priming depends on the similarity of the prime and target items suggests that both rules and instances may be involved in these tasks (hence, the criterion is not very diagnostic). We return to this issue in the final section.

Criterion 6: A Verbal Protocol May Mention a Rule or Its Components

Rationale. The rationale for this criterion is based on the standard interpretation of protocol analysis. Presumably, the protocol is a direct reflection

of what is active in the subject's short-term or working memory (Ericsson & Simon, 1984), and if a particular rule has been in working memory, then it may have been recently used. Or, to put it in terms of our sketch of a rule model, the products of the access, instantiation, or application stages may reside (perhaps only briefly) in working memory, which makes them accessible to report. There is no reason to expect an instance model to yield such reports. However, the protocol criterion is still of limited diagnosticity, given that there are cases of apparent rule following in which the rules cannot be reported (namely, in language), as well as cases of reported rules for tasks for which there is independent evidence that the rules were not followed (Nisbett & Wilson, 1977).

Evidence About Modus Ponens. In Rips' (1983) studies of deductive reasoning, he had subjects talk aloud while solving some problems. Rips found some clear parallels between the successive statements in a protocol and the sequence of propositional rules needed to solve the problem. Because one of these rules is modus ponens, these findings provide some indirect evidence for modus ponens meeting our protocol criterion. Similarly, Galotti, Baron, and Sabini (1986) collected verbal protocols while subjects tried to generate conclusions to syllogistic arguments. They concluded that the protocols "provide direct evidence of the existence of deduction rules" (p. 19; the protocols also provide evidence of the existence of nonrule-like entities).

Evidence About the Law of Large Numbers. In Piaget and Inhelder's (1951/1975) classic study of the child's conception of chance, they found surprisingly clear paraphrases of the law of large numbers even from children aged 10 to 12. For example, in one situation a child is presented with a pointer that could stop on one of eight different colored locations, and is asked if there is more likely to be an equal number of stops on each color if the pointer is spun 15 times or 800 times. One child replied:

> It will be more regular with 800 because that's larger. For a small number [of chances] [the outcome] changes each time and it depends on the number of times, but with a larger number of tries it has more chances of being more regular. (p. 89)

Although this protocol provides some prima facie evidence for the use of the law of large numbers, a skeptic could easily claim that the reasoning revealed in the protocol is not what is actually mediating the problem solving, and that people are merely inventing plausible stories to explain their behavior. What is needed to strengthen protocol evidence is a linking of it to performance measures. This is exactly what Nisbett and his colleagues have done. They found evidence that some people can articulate an abstract version of the law of large numbers, and that those who invoke it in justification of their answers to problems covered by the rule are in fact more likely to give correct answers. For example, in the isolated-island problem discussed

earlier, Nisbett et al. (1983) found that subjects often justified their willingness to make strong generalizations from a single case on the basis of assumptions about low variability and the resulting generalizability even from small samples. Subjects who explicitly gave such justifications were more likely to reason in accordance with the law of large numbers in general. Similarly, Jepson, Krantz and Nisbett (1983), and Fong et al. (1986) found that some subjects often articulated quite general versions of the rule in justifying answers to specific problems. For example, it was common for subjects to say things like, "The more examples you have, the better the conclusion you can draw." Subjects who provided such articulations of the rule gave answers in accordance with the rule on a higher proportion of problems than did other subjects.

Criterion 7: Performance on a Specific Rule-Based Problem Is Improved by Training on an Abstract Version of the Rule

Rationale. The idea behind this criterion is that, because rule following is presumably what underlies performance on specific problems, practice on an abstract version of the rule (abstract in all senses we have considered) can improve performance on specific problems. In part, this should be true because training improves the rule—clarifies it, renders it more precise, and even changes its nature so as to make it more valid. From the perspective of our sketch of a rule model, practice on the rule in the abstract could also benefit performance by increasing the accessibility of the rule and perhaps also by facilitating the application of the rule. (To the extent that there were *any* examples in the training, there could be a facilitation of the instantiation stage as well.) From the perspective of an instance model, there is no obvious reason why such abstract training should have any effect on performance. Criterion 7 is therefore quite diagnostic.

Evidence About Modus Tollens. Cheng et al. (1986) showed that training on rules from propositional logic, particularly modus tollens, did not lead to any improvement in performance on the four-card problem, specifically on selection of the choice corresponding to *not q*. Training was of two forms. One form was an extensive laboratory session describing the rule and its application in Venn diagrams, truth tables, and an illustrative conditional statement. The other was an entire course in introductory logic that was centered on conditional logic, including the modus tollens rule. Criterion 7, therefore, speaks against the use of modus tollens. (Abstract instruction also did not improve performance on the component of the four-card problem that could be solved by application of modus ponens—selection of the *p* choice—but errors were sufficiently infrequent for ponens as to raise the possibility that there was a ceiling effect.)

Evidence About Contractual Rules. In another study, Cheng et al. (1986) showed that comparable training on an abstract statement of the obligation

rule ("If precondition P is satisfied, action A must be taken") did improve performance on the four-card problem. Training included drill in the checking procedures required to establish whether an obligation had been violated. Subjects were then asked to solve various versions of the four-card problem, including versions to which an obligation interpretation could be applied relatively easily, and arbitrary versions, such as the original Wason (1966) letter-and-number problem. We performed a reanalysis of the Cheng et al. results and found that the abstract training improved performance on those versions of the problem that could possibly be interpreted in obligation terms ("If a house was built before 1979, then it has a fireplace"), and did not improve performance on problems for which an obligation interpretation seemed out of the question (such as the original Wason problem).

We mentioned earlier that there is no obvious way in which an instance model can handle these results, but a nonobvious way might proceed as follows: Although the training involved only abstractions, subjects may have generated their own examples and subsequently retrieved those examples during the four-card problem. What is wrong with this account is the usual set of difficulties. It seems most unlikely that the examples generated during training would have anything in common with the test items in the four-card problem, other than that they involved the notion of obligation. Again, the account rests on the ad hoc assumption that retrieval is primarily based on whatever is correlated with the special structure of the rule.

Evidence About the Law of Large Numbers. Fong et al. (1986) showed that training on the law of large numbers affects the way people reason about a wide range of problems involving variability and uncertainty. They taught their subjects about the law of large numbers using purely abstract concepts and procedures. They defined for them the notions of *sample, population, parameter,* and *variability,* and showed by urn-problem demonstrations that larger samples are more likely to capture population parameters than smaller samples. (These demonstrations, according to our sketch of a rule model, might have influenced the instantiation stage). Subjects were then asked to solve problems involving random generating devices, such as slot machines and lotteries, problems dealing with objective, quantifiable behavior, such as athletic and academic performances, and problems dealing with subjective judgments or social behaviors that are not normally coded in quantifiable terms. For example, one objective problem referred to earlier, required subjects to recognize that a basketball talent scout's assessment of a potential player was based on a relatively small sample of behavior and might be mistaken. A subjective problem described a head nurse's assertion that the most compassionate nurses, as judged from the first few days on the job, generally turn out to be no more concerned than the others, together with her attribution that this was probably the case because the most caring nurses build up a shell to protect themselves. A statistical answer to this problem recognized that a few days' observation of nurses' behavior

might not be a large enough sample for a stable estimate of an attribute like compassion. In line with a rule model, the abstract rule training produced a substantial increase in the number and quality of statistical answers, and did so to about the same degree for all three problem types.

Further Evidence on the Law of Large Numbers and Other Rules. An extensive set of studies by Nisbett and his colleagues on the effects of undergraduate and graduate training on reasoning is relevant to Criterion 7. They found that undergraduate training in psychology and the social sciences (Lehman & Nisbett, 1990) and graduate training in psychology (Lehman, Lempert, & Nisbett, 1988) markedly increased the degree to which students call on statistical principles (like the law of large numbers) in reasoning about everyday events involving uncertainty. Fong et al. (1986) found that a single course in statistics had a marked effect on the way students reason about sports. These results speak to Criterion 7 to the extent that statistics is typically taught as a highly abstract set of rules.

Similarly, Morris et al. (1991) found that graduate training in psychology improved students' abilities to apply causal rules to both unfamiliar and purely abstract material. In contrast, training in philosophy or chemistry had no effect on students' causal reasoning, presumably because neither of these fields emphasizes the reasoning required for inferences about various types of causality. Again, the work is relevant to Criterion 7 to the extent that instruction about causality in psychology is quite formal and owes little to detailed work with concrete examples.

Finally, work by Larrick, Morgan, & Nisbett (1990; Larrick, Nisbett, & Morgan, in press) shows that formal training in cost-benefit rules affects people's reasoning about an indefinitely large number of problems involving choice in everyday life.

Criterion 8: Performance on Problems in a Particular Domain is Improved as Much by Training on Problems Outside the Domain as on Problems Within it, as Long as the Problems are Based on the Same Rule

Rationale. If a major product of training is an abstract rule that is as applicable to problems from one domain as to those from another, then subjects taught how to use the rule in a given content domain should readily transfer what they have learned to other domains. To put it in terms of our sketch of a rule model: The major products of training are increases in the accessibility of the rule and in the consequent ease with which the rule can be instantiated and applied, and all of these benefits should readily transfer to domains other than those of the training problems. The upshot is that domain-specificity effects of training might be relatively slight. To the ex-

tent such effects are slight, instance models are embarrassed because they naturally predict better performance for test problems that resemble training ones. Hence, Criterion 8 is very diagnostic of rule following.

Evidence About the Law of Large Numbers. This criterion has thus far been applied only to the law of large numbers. Fong et al. (1986) trained subjects in one of three domains: random generating devices, objectively measurable abilities and achievements, or subjective judgments. Then subjects worked on test problems from all three domains. Performance on the test problems—as measured by the frequency of mention of statistical concepts and laws, and by the quality of the answers—was improved by the training. Most importantly, the degree of improvement for problems in the untrained domains was as great as for problems in the trained domain. For example, training on probabilistic device problems improved performance on objective and subjective test problems as much as it did for probabilistic device test problems.

The domains employed by Fong et al. (1986) are very broad ones, leaving open the possibility that two problems from the same domain shared very little in the way of content, perhaps little more, in fact, than two problems from different domains. But this possibility is ruled out by a more recent study. Fong and Nisbett (1991) examined two different objective attribute domains: athletic contests and ability tests. They taught some of their subjects to apply the law of large numbers to one domain, and some to apply this rule to the other domain. When subjects were tested immediately, again, there was no effect at all of training domain on performance. This is strong evidence for rule following. When subjects were tested after 2 weeks, however, there was some effect of domain on performance, although there was still a significant training effect across domains as well. The domain-specificity effect after a delay should probably not be attributed too quickly to retrieval of examples from memory. Performance at the later testing time was unrelated to the ability to recall details of examples, but was related to the ability to recall the abstract rule. The latter findings suggest that, during training, subjects may have learned how to code the elements of a given domain in terms of the rule, which could result in domain-specific coding and access processes. Such processes would lead to an advantage for problems in the trained domain after a delay when access was more problematic.

It is worth emphasizing that the utter lack of domain-specificity effects, when testing takes place immediately, is particularly problematic for an instance model. Such a model requires that the more similar the content of the test and training problems, the more likely a test problem will retrieve a training problem, which will culminate in better performance when the test and training problems are from the same domain. The only way to salvage the model is to posit that retrieval is heavily based on only those features correlated with the special structure of the rule. Yet, it is not even clear that

there are any content features of a problem that are correlated with the law of large numbers. As usual, then, the assumption in question seems ad hoc, and likely wrong.

A Possible Ninth Criterion

Criterion 8 says that after training, performance on a rule-governed item is unaffected by its similarity to items encountered during training. A generalization of this phenomenon yields a new criterion: *Performance on a rule-governed item is unaffected by its degree of similarity to previously encountered items.* This is a very diagnostic criterion, because the hallmark of instance models is their sensitivity to similar items stored in memory.

We have not included the preceding as one of our criteria because we have not been able to find a study in which it has been successfully used to bolster the case for abstract rules in reasoning. Perhaps one reason the proposed similarity criterion has not been used is that it is exceptionally sensitive to any use of instances whatsoever. But we may be being too pessimistic here, because there are psycholinguistic studies where the proposed similarity criterion has been met, thereby providing very strong evidence for rule use. Consider again research on phonological rules showing that people can supply the plurals of nonsense nouns. The fact that people can as readily supply the plural for "zamph" as for "zig"—even though "zamph" does not rhyme with any English word and hence is not very similar to any known instance—is an indication that performance is unaffected by the similarity of the test item to previous instances (Pinker & Prince, 1988). A comparable story holds for the rule for forming the past tense of regular verbs. Young children are no more likely to produce the correct past tense for regular verbs that are similar to many other regular verbs, than they are to produce the correct past tense for regular verbs that are similar to few other regular verbs (Marcus et al., 1990). Perhaps this kind of evidence can be obtained with abstract reasoning rules.

GENERAL DISCUSSION

In this final section, we begin by summarizing our results, and then take up a number of outstanding issues. One such issue concerns reasoning mechanisms that involve both rules and instances; a second issue concerns the possibility of a type of rule following other than the explicit sort we have considered thus far; the final issue deals with the implications of our findings and arguments for connectionist models of reasoning.

Summary

Throughout most of this article we have been concerned with two interrelated matters: possible criteria for rule following and possible rules that

are followed. Let us first summarize our progress regarding the possible criteria, then turn to what we have found out about rules.

Criteria. We have presented and defended a set of criteria for establishing whether or not a rule is used for solving a given problem. Satisfaction of the less diagnostic of these criteria—those concerned with overextension, number of rules, priming, and protocols—adds something to the case that a given rule is used for solving a given problem. Satisfaction of the more diagnostic criteria—those concerned with familiarity, abstractness, abstract training effects, and domain independence in training—adds even more to the case for rule following. And satisfaction of most or all of these criteria adds greatly to the case for rule following. These criteria can serve to put the debate between abstraction-based and instance-based reasoning into clearer perspective.

Table 1 presents each of the eight criteria crossed with the five different rule systems we have examined in detail; broken lines indicate that the rule system failed the criterion of interest. Table 1 makes it easy to see a pair of points concerning the criteria. One is that most of the criteria have been underused. It is clear that application of the criteria has been relatively haphazard, with many tests of a particular criterion for some rules and only one or two tests of a smattering of the other criteria. We suspect that the criteria used have been chosen relatively arbitrarily, and that investigators often have tested less powerful criteria than they might have, simply because they were not aware of the existence of other, more powerful ones. Our overview of criteria and the rationales behind them should help to organize and direct research on the use of rules.

The other point about the criteria that is readily apparent from Table 1 is that the criteria converge. That is, if a rule passes one criterion it generally passes any other criterion that has been applied. Conversely, if a rule fails one criterion it generally fails other criteria that have been applied. We have only one case of this convergence of failures—modus tollens—because our main concern has been with abstract rules that are likely to be in people's repertoires. If we turn our attention to unnatural rules, which are unlikely to be in people's repertoires, we should see other failures to satisfy the criteria. Consider, for example, work by Ross (1987), in which people are taught relatively unnatural rules from probability theory, such as the rule that specifies the expected number of trials to wait for a particular probabilistic event to occur (the "waiting time" rule). Ross observed a strong violation of our domain-independence-of-training criterion, that is, performance on a test problem markedly depended on its similarity to a training problem. Recent results by Allen and Brooks (1991), who taught subjects artificial rules, makes exactly the same point. These failures of unnatural rules to pass the criterion attest to the validity of the criteria.

Three qualifications of the criteria are also worth mentioning. First, for purposes of clarity we have stated some of our criteria in an absolute or all-

TABLE 1

Criteria for Use of Abstract Rules for Reasoning and Evidence Base Relating to Them

Criteria	Rule Types				
	Modus Ponens	Modus Tollens	Contractual (Permissions & Obligations)	Causal	Law of Large Numbers
1. Good performance on unfamiliar items	Bryne (1989)	Cheng & Holyoak (1985) Numerous others	Cheng et al. (1986) Cheng & Holyoak (1985)	Morris et al. (1991)	Nisbett et al. (1983)
2. Good performance on abstract items	Evans (1977)	Wason (1966) Numerous others	Cheng & Holyoak (1985)		
3. Overextension early in training					Fong et al. (1986)
4. Number of rules and performance	Osherson (1975) Rips (1983) Braine et al. (1984)				
5. Priming effects			Langston et al. (1991)		
6. Protocols identify rules	Rips (1983)				Piaget & Inhelder (1951/1975) Jepson et al. (1983) Nisbett et al. (1983) Fong et al. (1986)
7. Abstract training effects	Cheng et al. (1986)	Cheng et al. (1986)	Cheng et al. (1986)	Morris, Cheng, & Nisbett (1991)	Fong et al. (1986) Lehman & Nisbett (1990) Lehman et al. (1988)
8. Domain independence of training					Fong et al. (1986) Fong & Nisbett (1991)

Note. Broken lines indicate rule system filed the criterion of interest.

or-none fashion, but probably it would be more useful to treat each criterion in a relative fashion. We can illustrate this point with Criterion 1, *performance on rule-governed items is as accurate with unfamiliar as familiar items.* Taking the criterion literally, there is evidence for rule-following only when there is absolutely *no* difference between unfamiliar and familiar items. But surely the phenomenon that underlies the criterion admits of degrees, perhaps because of moment-to-moment variations in whether an individual uses a rule. Given this, Criterion 1 is better stated as *the less the difference in performance between unfamiliar and familiar rule-governed items, the greater the use of rules.* Similar remarks apply to Criterion 2 (good performance on abstract items), Criterion 7 (abstract training effects), and Criterion 8 (domain independence of training). It is noteworthy that actual uses of these criteria tend to employ the relative interpretation (see, e.g., the Allen and Brooks, 1991, use of domain-independence-of-training effects).

A second qualification of the criteria stems from the fact that their diagnosticity has been measured in terms of how difficult they are to explain by models based on *stored* instances. But Johnson-Laird (1983) has championed a theoretical approach which holds that people reason by generating *novel* instances (in his terms, "reasoning by means of mental models"). To illustrate, suppose someone is told, "If gork then flum." They would represent this conditional in terms of the following sort of mental model:

gork1 = flum 1
gork2 = flum 2
(flum 3).

The equal sign indicates that the same instance is involved, and the parentheses indicates that the instance is optional. If now told there exists a gork, one can use this mental model to conclude there also exists a flum, and in this way implement modus ponens. What is important about this for our purposes is that a theory based on such novel instances seems more compatible with our criteria than theories based on stored instances. For example, there is no obvious reason why one cannot construct a mental model as readily for an unfamiliar item as for a familiar one, or as readily for an abstract item as a concrete one.

The final qualification is simply that the application of our criteria does not provide as definitive data on the rule-versus-instance issue as does a contrast of detailed models. Our criteria are needed mainly in situations where detailed reasoning models have not been developed: the usual case as far as we can tell. (An exception is Nosofsky, Clark, and Shin, 1989, who did contrast detailed rule and instance models, but who considered rules that are not abstract by our definitions.) Our criteria also provide useful constraints in developing detailed rule models, for example, any rule model that is concerned only with abstract rules ought to produce comparable per-

formance for unfamiliar and familiar items, for abstract and concrete items, and so on.

Rules. Table 1 also tells us about what rules are followed. We believe that the applications of the criteria to date serve to establish that people make use of a number of abstract rules in solving problems of a sort that occur frequently in everyday life. In particular, there is substantial evidence for at least three sorts of rule systems.

For modus ponens, there is evidence that people: (a) perform as well—that is, make inferences in accordance with the rule—on unfamiliar as on familiar material; (b) perform as well on abstract as on concrete material; (c) perform better if they must invoke the rule fewer rather than more times; and (d) sometimes provide protocols suggesting that they have used the rule. (On the other hand, there is some evidence that the rule cannot be trained by abstract techniques, but this evidence may merely indicate that the rule is already asymptotic.)

For contractual rules, namely permission and obligation rules, there is evidence that people: (a) perform as well on unfamiliar as on familiar material; (b) perform as well on abstract as on concrete material; (c) show priming effects of the rule, at least within a content domain; and (d) benefit from training in their ability to apply the rule to any material that can plausibly be interpreted in terms of it. There is also some evidence of a comparable kind for formally similar causal rules.

For the system of statistical rules under the rubric of the law of large numbers, it has been shown that people: (a) perform well with unfamiliar material; (b) overextend the rule early in training; (c) often mention the rule in relatively abstract form in justification of their answers for particular problems; (d) improve in their ability to apply the rule across a wide number of domains by purely abstract training on the rule; and (e) improve their performance on problems outside the domain of training as much as on problems within it.

The demonstrations that people follow modus ponens and the law of large numbers are of particular interest in view of the fact that these two rules are normative and promote optimal inferential performance. Evidence for people following certain abstract inferential rules thus amounts to evidence for people manifesting aspects of rationality. Although there is less data about causal rules, what evidence there is suggests that people also follow these rules (see Table 1), which again are normative. And there is some recent evidence for the use of still another set of normative rules, those governing economic choices (Larrick et al., 1990; in press).

In contrast to the positive evidence summarized before, there are three lines of negative evidence on the question of whether people use modus

tollens. It has been shown that people perform poorly: (a) with unfamiliar items; (b) with abstract items; and (c) even after formal training in the rule. We therefore believe that the consensus among students of the problem that most people do not use modus tollens is justified in terms of the criteria studied to date. This demonstration indicates that application of our criteria can cut both ways: Negative evidence relating to the criteria can cast substantial doubt on the use of a rule, just as positive evidence can buttress the case for its use.

Of course modus ponens, modus tollens, contractual rules, and the law of large numbers are just a handful of the many possible seemingly natural rules that people may follow in reasoning about everyday problems. There are, for example, numerous rules in propositional logic other than ponens and tollens that have been proposed as psychologically real (see, for instance, Braine et al., 1984). One such rule is *and-introduction,* which states *If p is the case and if q is the case then p and q is the case.* The obvious question is: How does and-introduction stack up against our eight criteria? The same question applies to other rules from propositional logic, and to rules that have figured in Piagetian-type research (including transitivity, commutativity, and associativity), as well as to rules that come from other bodies of work. The point is that all we have done in this article is sample a rule or two from a few major branches of reasoning—deduction, statistics, and causality— and there are other rules of interest in these and other branches of reasoning.

A final point to note about the evidence for rules is that the work to date shows not merely that people *can* follow rules when instructed to do so in artificial problem-solving situations, but that they *do* follow quite abstract inferential rules when solving ordinary, everyday problems. For example, in their studies of the law of large numbers, Fong et al. (1986) performed not merely laboratory experiments, but field studies in which subjects did not even know they were being tested. In one study, male subjects were called in the context of an alleged "survey on sports opinions." Subjects were enrolled in introductory statistics courses and were called either at the beginning of the course or at the end. After being asked a few questions about NBA salaries and NCAA rules, it was pointed out to them that although many batters often finish the first 2 weeks of the baseball season with averages of .450 or higher, no one has ever finished the season with such an average. They were asked why they thought this was the case. Most subjects responded with causal hypotheses such as, "the pitchers make the necessary adjustments." Some, however, responded with statistical answers such as, "there are not many at-bats in 2 weeks, so unusually high (or low) averages would be more likely; over the long haul nobody is really that good." There were twice as many statistical answers from subjects tested at the end of the term as from subjects tested at the beginning.

Similarly, Larrick et al. (1991) found that subjects who were taught cost-benefit rules came to apply them in all sorts of life contexts, from consumer decisions about whether to finish a bad meal or bad movie, to professional decisions about whether to pursue a line of work that was turning out to be disappointing, to hypothetical questions about institutional policy and international relations.

Thus, the work reviewed here establishes not merely that people can follow abstract rules self-consciously in appropriate educational, experimental, or professional settings, but that such rules play at least a limited role in ordinary inference.

Combining Rule and Instance Mechanisms

Our review indicates that pure instance models of reasoning and problem solving are not viable. There is too much evidence, stemming from the application of too many criteria, indicating that people use abstract rules in reasoning. On the other hand, there is also abundant evidence that reasoning and problem solving often proceed via the retrieval of instances (e.g., Allen & Brooks, 1991; Kaiser, Jonides, & Alexander, 1986; Medin & Ross, 1989; Ross, 1987). At a minimum, then, we need to posit two qualitatively different mechanisms of reasoning. Whereas some situations may involve only one of the mechanisms, others may involve both.

In addition to *pure-rule* and *pure-instance* mechanisms, hybrid mechanisms may be needed as well. In particular, hybrid mechanisms may be needed to account for the situations noted earlier in which people process instances deeply enough to encode some information about the relevant abstraction as well as about the concrete aspects of the instance. These are the situations that are the concern of most case-based reasoning models (e.g., Hammond et al., 1991; Kolodner, 1983; Schank, 1982). In such situations, people have essentially encoded both an instance and a rule, so a hybrid mechanism must specify how the two representational aspects are connected. We consider two possibilities.

One possibility is that a retrieved instance provides access to a rule. That is, when an item is presented, it first accesses similar instances from memory that the reasoner can use to access a rule. Then, the final stages of rule processing—instantiation and application—ensue, though the instance may serve as a guide for these two stages. We can illustrate this mechanism with the drinking version of the four-card problem. When presented the problem, presumably a subject uses this item to retrieve from memory an episode of a drinking event; this representation may contain the information that people below the drinking age are in violation of the law, and the concept of *violation* may be used to access the permission rule; from here on, processing would continue as specified in our sketch of a rule model except that the retrieved instance can be used to guide the instantiation and application

stages. This hybrid process, which we will refer to as *instance-rule mechanism,* captures the intuition that we often understand an abstract rule in turns of a specific example.

The other possibility is that a rule provides access to a relevant instance (a *rule-instance* mechanism). That is, when an item is presented it is coded abstractly, and this abstraction accesses the appropriate rule (these are the first two stages of our sketch of a rule model). The rule then provides access to some typical examples, and these instances control further processing. Again, we can illustrate with the drinking version of the four-card problem. When presented the problem, a subject codes the item in terms of *permission,* and uses this code to access the permission rules. Associated with these rules are typical examples of *permission* situations, and one or more of these instances is used as an analogue for the present problem (that is, it is used for the mapping stage).

A few comments are in order about these mechanisms. Note that we are not proposing the two hybrid mechanisms as alternatives to the two pure mechanisms (rule and instance). Rather, we suspect that all four mechanisms can be used, albeit with different situations recruiting different mechanisms. (The experimental situations we reviewed in this article likely involved either the *pure-rule* or the *rule-instance* mechamisn.) In situations where more than one mechanism is involved, presumably the processes operate simultaneously and independently of one another. Thus, the final answer may be determined by a kind of "horse race" between the operative mechanisms, with the mechanism that finishes first determining the final judgment.

Note further that our hybrid mechanisms allow room for instance-type effects should they occur. Consider again Criterion 1, that novel rule-based items are treated as accurately as familiar ones. The available evidence is consistent with this criterion, but the criterion deals only with accuracy. Perhaps if one were to measure reaction times, familiar rule-based items might be processed faster than novel ones. Such a result could be handled easily by our *instance-rule* mechanism. Familiar items should be faster in accessing a relevant instance because familiar items are themselves likely to be instances. In addition, we have already seen an indication of instance effects even for accuracy. Such an effect appeared in connection with Criterion 5, that application of a rule primes its subsequent use. Recall that in the four-card problem, Langston et al. (1991) found evidence for priming of contractual rules only when the prime and target were similar in content. This pattern of results also fits nicely with the *instance-rule* mechanism. Only when the target and prime are similar in content does the target retrieve the prime instance, and only when the prime is retrieved does one gain access to the relevant rule. Thus, instance-type effects do not imply that rules were not involved.

Finally, another case of instance-type effects during rule use is provided by Ross (1987). Ross trained subjects on the waiting-time rule of probability

theory and then had them solve new test problems with the rule present. Even though the rule was present, subjects appeared to rely on training problems when determining how to instantiate the rule. These results indicate that instances are used not just to access a rule but also to help instantiate it, as in the instance-rule mechanism. (These results, however, may depend in part on the fact that the rule involved was not a natural one).

In short, the dichotomy between pure rules and pure instances is too simple. Hybrid mechanisms seem plausible, particularly in light of the role they play in current versions of case-based reasoning.

Two Kinds of Rule Following

Until now we have acted as if explicit rule following is the only kind of rule following. But a critical observation suggests the need to consider a second-kind. The observation (due to Douglas Medin, personal communication, April 1991) is that, when *linguistic* rules are stacked up against our eight criteria they seem to consistently fail three of them, namely verbal protocols, abstract training effects, and context independence in training. That is, people are notoriously unable to verbalize the linguistic rules they purportedly use, and they fail to benefit much from explicit (school) instructions on these rules. If linguistic rules meet only five of our criteria whereas reasoning rules (generally) meet all eight, perhaps the kind of rule followed involved in language is different from that involved in reasoning.

Presumably there is a kind of rule following that is *implicit* rather than *explicit;* that is, the rule is never explicitly represented, which accounts for why it can neither be reported nor affected by explicit instruction. The rule might be implemented in the hardware, and is essentially a description of how some built-in processor works (see Pinker & Prince, 1988, Section 8.2). Implicit rules are close to what we earlier characterized as operating principles of a system, and rules like this may be part of our basic cognitive architecture. Such notions fit nicely with Pylyshyn's (1984) concept of *cognitive penetrability*. His basic idea is that anything that is part of the fixed cognitive architecture cannot be altered (penetrated) by goals, context, or instruction. If some linguistic rules are part of our basic architecture, they should not be affected by instruction, which means that our two instructional criteria should fail, as they in fact do. (The seeming imperviousness of modus ponens to instruction leaves open the possibility that this rule too may be represented implicitly.)

Implications for Connectionist Models

Although we know of no limit, in principle, on the ability of connectionist models to code abstractions, the evidence we have presented for abstract rules does not fit well with the connectionist program.

For one thing, what seems to be the most straightforward account of much of the evidence involves concepts that are anathema to connectionism. The account we have in mind is that of explicit rule following: The rule and input are mentally represented explicitly, and application of the rule to the input involves an inspection of the input to determine whether the antecedent of the rule has been satisfied. Notions of *explicit data structures* and *inspection of explicit structures* simply lie outside the ontology of connectionism. Of course, connectionists may be able to develop alternative accounts of the data, but there is no reason to believe the resulting connectionist models will be as parsimonious as the sort of rule-based model we advocate. This is particularly the case given that the abstract rules that have to be modeled all involve variable bindings, which remains a difficult issue in connectionist work (for discussion, see Holyoak, 1991). In short, rule-based models provide a simple account of the data, and no comparable connectionist alternatives are thus far in sight.

In constructing alternative models of the evidence, connectionists face another difficulty. The evidence indicates that people can use two qualitatively different mechanisms in reasoning, which we have termed "rules" and "instances," whereas connectionist models endorse a uniform representation. Connectionist models can either blur the rule-instance distinction, in which case they are simply failing to capture a major generalization about human cognition, or they can somehow mark the distinction, in which case they may be merely implementing a rule-based model in a connectionist net. We say "merely" because it is not clear that such an implementation will yield any new important insights about reasoning.

The preceding points have been programmatic, but the remaining one is more substantive. According to rule models, the rationale for some rules hinges on a *constituency relation*—like that which holds between *If p then q* and *p*—but most current connectionist models lack true constituency relations. In discussing this issue, we need to keep separate *localist* connectionist models, in which a concept can be represented by a single node, and *distributed* models, in which a concept is represented by a set of nodes. We consider localist models first.

To understand the constituency issue, consider modus ponens. Given *If p then q* and *p*, the fact that the latter is a constituent of the former is part of why we can conclude *q*. To take an even simpler example, consider again and-introduction: *If p is the case and if q is the case then p and q is the case.* Here, it is clear that the basis of the rule is a constituency relation; the rule essentially states, *if each of its constituents is the case, then a conjunction is the case.* In contrast, localist connectionist models lack constituency relations, so such relations can never serve as the bases for rules.

The reason localist connectionist models lack constituency relations is that their nodes (their representations) lack any internal structure, including

a part–whole structure. In a localist model for and-introduction, for example, there might be separate nodes for *p*, *q*, and *p and q*, which are connected in such a way that whenever the nodes for *p* and *q* are both activated, the node for *p and q* is activated. Importantly, the node for *p and q* has no internal structure, and in no sense contains the node for *p* or that for *q*. Hence, the relation between the *p* and *q* nodes on the one hand, and the *p and q* node on the other, is strictly causal (as opposed to constituency). That is, activation of *p* and *q* causes activation of *p and q* in exactly the same way that activation of a node for *fire* might cause activation of a node for *smoke*. Although we know of no data on whether constituency relations are perceived as the bases of some rules, our intuitions suggest they are, which favors the rule account. (For a fuller discussion of these issues, see Fodor and Pylyshyn, 1988).

Distributed connectionist models seem better able to accommodate constituency relations because they at least have a part–whole structure. Thus, if *p and q* is represented by a set of nodes, then some part of that set can, in principle, represent *p* and another part *q*. However, current distributed connectionist models still have trouble capturing constituent structure, as Fodor and McLaughlin (1990) pointed out. The latter authors take up a proposal of Smolensky's (1988), in which a concept (rule) is represented in terms of a vector whose components represent the activity levels of the members of the relevant set of nodes. According to Smolensky, vector *a* is a constituent of vector *b* if there exists a third vector—call it *x*—such that $a + x = b$; *a* is a part of *b* because *b* is derivable from $a(+x)$. But this proposal permits the possibility that *b* may be activated without *a* being activated. In the case of and-introduction, this means that *p and q* could be activated without *p* being activated. Such a thing should be impossible if *p* is a true constituent of *p and q*. Again, to the extent some rules are based on constituent structure, the rule account is favored over current connectionist rivals.

None of this is to suggest that connectionist models do not have an important role to play—they have been very successful in capturing aspects of perception, memory, and categorization, for example—but rather to suggest that some aspects of reasoning may be inherently rule based, and hence, not naturally captured by connectionist models. Of course, a rule-based model, unlike a connectionist one, will not look like a biological model. Thus, to pursue rule-based models of reasoning is to give up the wish that all mental phenomena be expressive of biological phenomena rather than merely emergent on them. It has always been hard to make the leap from mere neural connections to abstract rules that seem metaphorically to sit astride the hustle and bustle of biological activity in the brain, altering and managing the results of such activity, and being modified by the mere words of outsiders and the ministrations of educators. We do not pretend to be able to make the leap from the known facts of the behavior of the nervous system to a

plausible, emergent set of highly modifiable abstract rules. We claim merely that a correct theory of mind may have to do so.

REFERENCES

Allen, S.W., & Brooks, L.R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 278–287.

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406.

Barsalou, L.W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T.K. Srull & R.S. Wyer (Eds.), *Advances in social cognition* (Vol. 3). Hillsdale, NJ: Erlbaum.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150–177.

Braine, M.D.S., Reiser, B.J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G.H. Brown (Ed.), *Psychology of learning and motivation.* Orlando, FL: Academic.

Buss, D. (1991). Evolutionary personality psychology. *Annual Review of Psychology, 42,* 459–491.

Byrne, R.M.J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 61–83.

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416.

Cheng, P.W., & Holyoak, K.J. (1989). On the natural selection of reasoning theories. *Cognition, 33,* 285–314.

Cheng, P.W., Holyoak, K.J., Nisbett, R.E., & Oliver, L.M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293–328.

Chomsky, N. (1985). *Knowledge of language.* New York: Praeger.

Collins, A.M., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science, 13,* 1–50.

Ericsson, K.A., & Simon, H.A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Ervin, S.M. (1964). Imitation and structural change in children's language. In E.H. Lenneberg (Ed.), *New directions in the study of language.* Cambridge, MA: MIT Press.

Evans, J. St. B.T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29,* 297–306.

Fodor, J.A. & McLaughlin, B.P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition, 35,* 183–204.

Fodor, J.A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3–72.

Fong, G.T., Krantz, D.H., & Nisbett, R.E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18,* 253–292.

Fong, G.T., & Nisbett, R.E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General, 120,* 34–45.

Galotti, K.M., Baron, J., & Sabini, J. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General, 115,* 16–25.

Gentner, D. (1983). Structure mapping; A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10,* 277–300.

Goodman, N. (1955). *Fact, fiction and forecast* (chap. 3). Cambridge, MA: Harvard University Press.

Griggs, R.A. (1983). The role of problem content in the selection task and in the THOG problem. In J. St. B.T. Evans (Ed.), *Thinking and reasoning*. London: Routledge & Kegan Paul.

Griggs, R.A., & Cox, J.R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407–420.

Hammond, K.J., Siefert, C.M., & Gray, K.C. (1991). Functionality in analogical transfer: A hard match is good to find. *The Journal of the Learning Sciences, 1,* 111–152.

Hinton, G.E., & Sejnowski, T.J. (1984). Learning semantic features. *Proceedings of the Sixth Annual Conference of the Cognitive Science Society* (pp. 63–70). Boulder, CO.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.T. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: Bradford Books/MIT Press.

Holyoak, K.J. (1984). Analogical thinking and human intelligence. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2). Hillsdale, NJ: Erlbaum.

Holyoak, K.J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*. Cambridge, MA: Cambridge University Press.

Holyoak, K.J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15,* 332–340.

Holyoak, K.J., & Thagard, P.T. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Jepson, C., Krantz, D.H., & Nisbett, R.E. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences, 6,* 494–501.

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kaiser, M.K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning on abstract and familiar physics problems. *Memory and Cognition, 14,* 308–312.

Kelley, H.H. (1972). Causal schemata and the attribution process. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

Kolodner, J.L. (1983). Reconstructive memory: A computer model. *Cognitive Science, 7,* 281–328.

Langston, C., Nisbett, R., & Smith, E.E. (1991). *Priming contractual rules*. Unpublished manuscript, University of Michigan, Department of Psychology, Ann Arbor.

Larrick, R.P., Morgan, J.N., & Nisbett, R.W. (1990). Teaching the normative rules of choice. *Psychological Science*.

Larrick, R.P., Nisbett, R.E., & Morgan, J.N. (in press). Who uses the normative rules of choice? *Organizational behavior and human decision-making*.

Lehman, D.R., Lempert, R.O., & Nisbett, R.E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday life events. *American Psychologist, 43,* 431–443.

Lehman, D., & Nisbett, R.E. (1990). A longitudinal study of the effects of undergraduate education on reasoning. *Developmental Psychology, 26,* 952–960.

Lewis, C. (1988). Why and how to learn: Analysis-based generalization of procedures. *Cognitive Science, 12,* 211–256.

Manktelow, K.I., & Evans, J. St. B.T. (1979). Facilitation of reasoning by realism: Effect or noneffect? *British Journal of Psychology, 70,* 477–488.

Marcus, G.F., Ullman, M., Pinker, S., Hollander, M., Rosen, T.J., & Xu, F. (1990). *Overextensions* (Occasional Paper No. 41). MIT, Center for Cognitive Science.

McClelland, J.L., & Rumelhart, D.E. (1986). *Parallel distributed processing* (Vol. 1). Cambridge, MA: MIT Press.

Medin, D.L., & Ross, B.H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5). Hillsdale, NJ: Erlbaum.

Medin, D.L., & Smith, E.E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 241–253.

Miller, G.A. (1962). Some psychological studies of grammar. *American Psychologist, 7,* 748–762.

Morris, M.W., Cheng, P., & Nisbett, R.E. (1991). *Causal reasoning schemas.* Unpublished manuscript, University of California, Department of Psychology, Los Angeles.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–317.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R.E., (1991). Priming the law of large numbers. Unpublished manuscript, University of Michigan, Department of Psychology, Ann Arbor.

Nisbett, R.E., Krantz, D.H., Jepson, D., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Nisbett, R.E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 8,* 231–259.

Nosofsky, R.M., Clark, S.E., & Shin, H.J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282–304.

Osherson, D. (1975). Logic and logical models of thinking. In R. Falmagne (Ed.), *Reasoning: Representation and process.* New York: Wiley.

Piaget, J., & Inhelder, B. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children.* New York: Norton. (Original work published 1951)

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73–194.

Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press.

Pylyshyn, Z. (1990). Rules and representations: Chomsky and representational realism. In A. Kasher (Ed.), *The Chomskyian turn.* Cambridge, MA: Blackwell.

Reich, S.S., & Ruth, P. (1982). Wason's selection task: Verification, falsification and matching. *British Journal of Psychology, 73,* 395–405.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90,* 38–71.

Rips, L.J. (1990). Reasoning. *Annual Review of Psychology, 41,* 321–353.

Ross, B.H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.

Rumelhart, D.E., & McClelland, J.L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

Schank, R.C. (1982). *Dynamic memory: A theory of learning in people and computers.* Cambridge, England: Cambridge University Press.

Smith, E.E., Lopez, A., & Osherson, D.N. (in press). Category membership, similarity, and native induction. In A. Healy, R. Shiffrin, & S.M. Kosslyn (Eds.), *Essays in honor of*

W.K. Estes, Hillsdale, NJ: Erlbaum.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1–23.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2,* 105–110.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* S251–S278.

Wason, P. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology.* Harmondsworth, England: Penguin.