CHAPTER 4

Reasoning

Kenneth J. Kurtz
Dedre Gentner
Virginia Gunn

I. REASONING

Human reasoning can be broadly described as a set of cognitive processes by which people take an initial set of information and generate inferences that extend beyond the original data. In this sense, the expectations, generalizations, and assertions people reach in interpreting events and situations can all be considered the result of reasoning. The inferences people produce can range from conclusions justified by formal procedures to sketchy hunches supported by varying degrees of evidence. Reasoning ranges from the really “hard” thinking it takes to formulate an answer to a difficult question or resolve a complex situation to the nearly automatic inferences and predictions that occur in the planning and execution of everyday activities. Reasoning is multifaceted, ubiquitous, and fundamental to human cognition.

A. An Introductory Framework

Reasoning can be broken down into three basic components: the available information, the cognitive processes brought to bear, and generated inferences. These are naturally expressed in terms of a simple, abstract description of the processing of information underlying human reasoning.
\[ y = F(x) \]  

(1)

In this formulation, \( x \) is the initial available information; the function \( F \) is a summary of the set of computational tools used to manipulate, recombine, or transform the input information; and \( y \) is the inferential product of the reasoning process, which might take the form of a judgment, conclusion, or prediction. In certain kinds of reasoning, people draw not only on the presented information \( x \), but also on represented information—memory traces from personal experience or general conceptual knowledge. To reflect this, we can introduce a component \( k \), which refers to the stored knowledge used in reasoning:

\[ y = F(x,k) \]  

(2)

The component \( k \) lends additional richness and flexibility to the construction of inferences by allowing reasoning processes to access domain knowledge and particular cases. By way of clarifying the nature of the \( k \) component, Winiewski (1995) draws a useful distinction between knowledge and experience. Knowledge is the synthesis of abstracted principles, constraints, or organizations that comprise understanding of the world. Such knowledge might be organized as theory-like explanatory frameworks that can guide understanding and inference (Murphy & Medin, 1985). Experience is stored information collected in a bottom-up fashion from empirical observation. It is relatively unstructured and may be represented by statistical summaries or memory traces themselves in the form of exemplars or cases.

Accounts of human reasoning can be characterized according to either Equation (1) or Equation (2) as well as by their particular instantiations of the variables. To be clear, an explanation of reasoning must specify:

1. How the available information \( x \) is represented
2. How the information \( x \) is related to computational tools \( F \) (and stored knowledge \( k \))
3. How inferences \( y \) beyond the available information are generated

B. Kinds of Reasoning

Human reasoning is a well-trodden area, and there are many excellent reviews (e.g., Evans, Newstead, & Byrne, 1993; Rips, 1990). Rather than following the divisions authors have found convenient in the past, we have tried to select "cuts" through the conceptual space that builds on earlier progress and offer new perspectives. Before presenting our framework, we briefly visit the traditional distinction between deductive and inductive forms of reasoning. In deductive reasoning, conclusions are entailed or follow directly from the application of logical forms to the premises. An additional usage of the term refers to a top-down direction of inferencing from abstractions to specific cases. Two corresponding meanings are linked to the notion of induction. Inductive reasoning refers to the generation of inferences that are not guaranteed within a formal system. These inferences are essentially guesses made probable by a set of evidence. Induction also refers to the bottom-up construction of high-level abstractions derived from observation of specific cases. This duality of reasoning can lead to confusions in usage. Furthermore, the very issue of directionality between general and specific can become slippery. For example, consider the generalization of a concept (e.g., dog) to a novel instance "Spot." From the bottom-up, the abstract concept is used to interpret and make inferences about "Spot"—from the bottom-up, the concept representation is updated to reflect properties of the new member.

Rips (1990) presents an alternative general framework in his treatment of human reasoning. He structures his discussion around a contrast between what he terms the strict and loose views of reasoning. The strict view calls upon algorithmic processes involving the ordered application of abstract procedures to produce definitive conclusions. The loose view calls upon specific associations, stored instances, statistical summaries, and heuristics that generate continuous-valued predictions or best guesses. The strict–loose distinction concerns both the way in which processing occurs and the inferential products that result.

Equations (1) and (2) above suggest a related but slightly more precise way to differentiate among approaches to reasoning: the degree to which content or domain knowledge, the \( k \) component, is used in the reasoning process. Reasoning methods based on Equation (1) involve domain-general computational procedures applied to the available information to generate inferences. Reasoning methods based on Equation (2) rely on bringing to bear stored knowledge or experience of the world that can be used to draw inferences beyond the available data. In considering a related set of issues, Gentner and Medina (1998) made use of a distinction between strong and weak methods of reasoning originally proposed by Newell and Simon (1972). Weak methods are general strategies that can operate without specific knowledge of a domain. Examples include methods such as means-ends analysis or an inference rule like modus ponens. Strong methods make intensive use of specific or abstract represented knowledge. Weak methods are valuable because of their generality; they provide a means of operating on novel or knowledge-poor domains. However, as Newell and Simon noted, strong methods are often superior when the appropriate knowledge is present. We will classify forms of reasoning such as logical deduction that fit Equation (1) as weak methods and forms of reasoning based on knowledge of a concept or stored analog consistent with Equation (2) as strong methods. Note that while the process of mapping a target to a base or classifying an instance as a category member may be domain-general, it is the role of specific knowledge content that makes these methods strong.

A second issue that follows from the weak–strong distinction is how the reasoning system connects to the rest of higher-order cognition. Traditionally, deductive reasoning is largely modularized in that there is a lack of explicit dependencies or interactions with other cognitive processes. Typically, other cognitive processes are referred to primarily in order to explain poor reasoning performance in term
much like that of Rips (1990), but takes the view that loose reasoning based on similarity and strict reasoning based on rules are two separate, and essentially competing, systems. One is associative and largely automatic while the other is logical and effortful. Sloman argues from cases of simultaneous contradictory belief that two separate systems operate independently to produce different solutions to the same problem at the same time. Another version of the separate systems view takes one set of methods to function as a backup in case of failure of the primary system. Rips (1995) notes the possibility that heuristics or probabilistic procedures may be called into play when logical deductions exceed some threshold of difficulty for the natural logic system. Rumelhart (1989) takes the view that the dominant system is reasoning by similarity instantiated as pattern matching in a parallel processing system. Formal reasoning is a special case for complex or novel problems. Rumelhart also points toward a possible unifying basis—formal reasoning may be derived from the internalization of sensorimotor interactions with external representations.

The third possible “big picture” of reasoning involves unifying strong and weak methods within a common framework. In explicating the strict-loose distinction, Rips (1990) pays particular attention to special structure. He argues that strict reasoning depends upon special components of representations, such as, “implies” and “or.” Rips speculates that analogy (often considered a loose approach) may in fact be a matter of special structure because of the central importance assigned to certain higher-order relational structures (such as causality) in analogical processing (Gennari, 1983, 1989). In the framework we have presented, analogical processing stands out as well: it is a strong form of reasoning with high process specificity. This powerful combination leads us to consider a unifying perspective on reasoning that emphasizes the explanatory potential of structured comparison processes.

II. REASONING BY FORMAL SYSTEMS: WEAK METHODS

There is a substantial body of research on formal reasoning—what we describe as weak methods with highly specified processing functioning in a generally modularized reasoning system. The longest standing view of reasoning is essentially a syntactic view with an emphasis on explicit prescriptions for the lawful combination and transformation of propositions. This view has produced accounts of reasoning based on weak methods (e.g., natural logic) that specify the conclusions that follow from a set of premises without regard for the particular content of the represented information being operated upon.

A. Reasoning with Propositions: Basic Phenomena

Although the traditional domain of “deductive” reasoning includes phenomena involving quantificational and relational forms of deduction, we will concentrate
on reasoning with propositions (statements using "not," "if," "and," and "or"). For example, consider the following conditional rule (from Evans et al., 1993):

If the ignition key is turned then the engine runs.

In the abstract case, this would be:

\[ \text{if } p \text{ then } q \]

What conclusions can be drawn from such a rule? Given this rule as a premise and an additional premise \( (p, q, \neg p, \text{ or } \neg q) \) there are four classical types of inferences:

- **Valid
  - modus ponens**: The key is turned; therefore the engine is running.
  \[ (p \text{ then } q, \text{ therefore, } q) \]
  - modus tollens**: The engine is not running; therefore the key is not turned.
  \[ (p \text{ then } q, \neg q, \text{ therefore, } \neg p) \]

- **Invalid
  - denial of the antecedent**: The key is not turned; therefore the engine is not running.
  \[ (p \text{ then } q, \neg p, \text{ therefore, } \neg q) \]
  - affirmation of the consequent**: The engine is running; therefore the key is turned.
  \[ (p \neg q, q, \text{ therefore, } p) \]

Evans et al. (1993) summarize behavioral studies of deductive performance on conditional rules. Adult participants are almost universally correct with modus ponens, but not with modus tollens (41–81%). Incorrect inference rates for denial of the antecedent and affirmation of the consequent are highly variable across studies (25–75%). This range might reflect differences in the materials used and the presentation of the task (such as whether or not forced choice).

Another important pattern of results stems from Watson’s (1966) selection task involving an open-ended conditional syllogism. In the standard abstract form of the task (Wason & Johnson-Laird, 1972), participants see four cards that they are told have a letter on one side and a number on the other. A conditional statement is given, such as, "If there is an A on one side of the card, then there is a 3 on the other side." The values on the sides of the cards facing the participant include two that match the elements in the rule (A and 3) and two differing values (e.g., D and 7). Participants are told to turn over the necessary cards to test the rule. The rule is an instance of the classic conditional "if \( p \) then \( q \)," and the cards correspond to "\( p \)," "\( q \)," "\( \neg p \)," and "\( \neg q \)."

Although the logically correct response to this task is to turn over the A and 7 cards ("\( p \)" and "\( \neg q \") only a minority of participants (6–33%) choose the correct two cards (Evans, 1982; Wason, 1968, 1969; Wason & Johnson-Laird, 1970). The majority of participants choose only the A (\( p \)) card or the A and D cards (\( p \) and \( q \)). This may be due to selection based on a match between the rule and the cards. When conditional statements contain negated antecedents and/or consequents participants seem to ignore the negation and simply match their selections to the cards mentioned in the premises (Evans, 1989). Additionally, the potential falsifying \( 7 (\text{not} \neg q) \) is rarely selected. In a reduced array version (only \( p \) and \text{not} \( q \) as choices) performance improves; possibly because participants’ attention directed to the need to test \( \neg q \) (Johnson-Laird & Wason, 1970).

The question arises whether performance would be facilitated when the selection task is taken out of the abstract. Wason and Shapiro (1971) obtained 62% correct selection using a specific rule grounded in a thematic context ("Every time go to Manchester, I go by car") and a set of cards representing various destination and modes of transport. However, this result failed to replicate. In a different study the rule involved an arbitrary relation between specific objects ("Every time I eat haddock, then I drink gin") and resulted in no facilitation relative to the low level of performance with the abstract, arbitrary conditions (Mankelow & Evan 1979).

Rather than specificity, it might be that familiarity and prior experience are the heart of these content effects. Griggs and Cox (1982) found strong facilitative using a highly familiar rule such as, "If a person is drinking beer, then the person must be over 20 years of age." This result has been replicated, but may require explicitly establishing a police officer scenario involving “looking for violators” (Grig & Cox, 1982; Pollard & Evans, 1987). Improvement up to 81% correct was found using a postal rule ("If a letter is sealed, then it has a 50 lirc stamp on it") among a set of actual envelopes (sealed and unsealed, with a 50 lirc or a 40 lirc stamp) materials; critically, the facilitation is dependent on participant familiarity with the rule (Cheng & Holyoak, 1985; Johnson-Laird, Legrenzi, & Legrenzi, 1972). The studies suggest that the source of facilitation is prior experience with the rule at counter-examples to the rule, but facilitation was also found by D’Andrade (cit in Evans et al., 1993) in a case where there is no direct real-world experience ("a purchase exceeds $30 then the receipt must be approved by the departmental manager"). However, it is possible that participants in this study were able to generalize from personal experience of a highly similar nature. We return to the issue of transfer effects with the selection task later in our discussion.

With these basic findings in mind, we now move to consider the most influential attempts to account for the major patterns of correct versus fallacious reasoning: rule-based models (also called mental logic or natural logic) and the mental model theory of reasoning. Additionally, we will discuss a third class of theories known as the dynamic reasoning models, including pragmatic reasoning schemas and social contract theories.

### B. Rule-Based Theories

Although the focus and details vary somewhat across different rule-based theories (e.g. Braine, 1978; Braine & O’Brien, 1991; Ohsherson, 1975; Rips, 1983), there is a basic consensus. Rule-based accounts fit the framework of weak methods of reasoning represented by Equation (1), \( y = F(x) \); where \( x \) is the available evidence
is the set of rules, and \( \gamma \) is the set of logical inferences. The rule-based account proposes that people possess explicit mental inference rules that operate on and transform propositions in working memory. These mental rules are similar, but not identical, to the deductive laws of formal systems of logic. The goal of rule theories of reasoning is to provide a psychological version of traditional logic that accounts for systematic patterns of successful and unsuccessful inferencing. Such a "natural logic" (Braine, 1978) is thus constrained by the demands of normative deductive steps as well as the vagaries of human performance.

1. Reasoning by Rules

Rule-based systems are characteristically discrete and syntactic. The assertions within a logical system are all-or-none, with no distinctions among degrees of validity or likelihood. This makes rule-based systems clear and consistent in their use and interpretation. Logical rule systems are weak methods of reasoning in that the processing is syntactic; blind to content. The rules specify the allowable ways in which information, any information, can be put together, related, and structured. The rules instantiate a mental logic with specifications of (a) the conditions under which they are applicable and (b) their entailments when brought to bear. For example, recall modus ponens, which is fundamental to every rule-based account:

\[
\begin{align*}
\text{if } & A, \text{ then } B \\
A & \quad \text{ therefore, } B
\end{align*}
\]

The rule applies universally without regard to the semantic reference of \( A \) and \( B \). There are two key advantages to this domain-generality: (a) the reasoning procedure is performed objectively and reliably, and (b) the reasoning procedure is robust and can be applied to any situation.

Reasoning by rules has been applied most often to the realm of propositional logic. First, the premises are encoded to extract the underlying logical form. This translation draws a predicate calculus language of propositions composed of symbols and connectives. As characterized by Rips (1994a), the premises are then stored in working memory. Second, the application of rules is coordinated by a reasoning program that searches for and produces appropriate inferences by constructing and linking steps in a mental proof. This chaining process may proceed forward or backward in a constrained way and may involve building assertion trees in working memory (Rips, 1983, 1994a). To limit working memory load, inferences are drawn only when needed for a direct answer to a problem or a subgoal to an answer. The set of deductive inferences comprises an implicit mental proof that validates the conclusions that are finally translated back into the content of the premises. Some accounts allow for a role of nonlogical processes (such as heuristic approaches or strong methods of reasoning) if logical processes fail to produce a straightforward conclusion (Braine, 1978; Rips, 1995).

2. Support for Rule-Based Theories

E. E. Smith, Langston, and Nisbett (1992) propose a set of criteria by which to specify and evaluate rule-based accounts. The authors emphasize that the criteria should distinguish between reasoning performance that is rule-governed versus that which is rule-described. Rules are often convenient ways to characterize a general pattern of performance or processing steps. Rule-based theories require that the rules are explicitly represented and activated during processing; that they be the causal force in the reasoning process. We focus here on two key predictions from the E. E. Smith et al. (1992) framework. In accord with the content-blind, universal quality of formal logical rules, rule-driven reasoning should exhibit no variation in performance due to the degree of familiarity or level of abstraction of the information. Second, following from the process-level description of rule-driven reasoning, the quality and ease of performance should be predicted by the number of rules required in the implicit deductive proof. In addition, some rules may be more easily retrieved than others (Rips, 1983).

Marcus and Rips (1979) found that participants confirm the validity of modus ponens arguments regardless of the content of the problems. Furthermore, most participants perform logically on modus ponens problems across levels of abstractness (Evans, 1977) and familiarity of items (Byrne, 1991). However, these findings do not extend reliably beyond modus ponens. Rule-based theories have received some support for predictions involving perceived difficulty. Problems are rated as more difficult when more applications of a rule are needed to formulate a proof (Braine, Reiser, & Ruman, 1984; Rips, 1983, 1989). Greater variety of rules and greater complexity of rules may likewise increase difficulty, Rips (1994a) reports longer reaction times for participants to evaluate conclusions for two-rule arguments than one-rule arguments. On higher-level deduction problems (so-called knight–knaves puzzles), larger-step problems took longer to solve than did smaller-step problems (Rips, 1989).

Rule theories must also be evaluated according to their account of reasoning failures. Errors can occur if a logical rule does not exist, if there is a failure to apply rules appropriately, or if the complexity of the task exceeds processing or capacity limitations (Rips, 1995). The actual circumstances of such failures are not well-specified, although limits on working memory and problems with the control mechanism responsible for applying the rules are potential contributing factors. Several researchers have attempted to argue that "actual reasoning" is logical, but errors may arise in the initial comprehension/encoding stage or in the final translation stage (Braine & O'Brien, 1991; Henle, 1962; Marcus & Rips, 1979; Ruman, Connell & Braine, 1983). There is some evidence that comprehension errors may occur separately from the reasoning process per se (see Braine & O'Brien, 1991; Evans et al., 1993; Henle, 1962; Ruman et al., 1983). For example, the term if might be interpreted as a biconditional, as opposed to a conditional due to an everyday usage of the English word if that is compatible with Greicean maxims of informativity and
relevance (Rips, 1988). Logical reasoning may also be hampered at the outset if people are not encoding the logically pertinent aspects of the input. Researchers have posited that an availability bias (Pollard, 1982) or relevance heuristic (Evans, 1989) may lead people to focus on the salience of certain features and associated memories that can lead them astray.

Even if reasoning proceeds logically from encoding through the application of inference procedures, participants may still fall prey to errors in the statement or interpretation of conclusions. The most common of these errors is the belief bias (Evans, 1989; Evans, Barston, & Pollard, 1983; Henle, 1962). Belief bias occurs when participants evaluate the validity of an argument based on whether the conclusion is compatible with prior beliefs, rather than on logical entailment. Because the effects of belief bias are more pronounced with arguments that are not deductive, Rips (1995) suggests that this may be a strategy that participants use when they are unable to find a proof. In addition, participants may misinterpret their task as one of evaluating confidence in a conclusion, rather than evaluating strict entailment.

3. Learning and Transfer of Rules

Where do the rules of natural logic come from? Rips (1994a) argues for innateness of basic logical abilities. He notes, however, that innateness does not necessitate perfection in application of rules, nor does it preclude the possibility of learning and improvement. E. E. Smith et al.'s (1992) criteria for rule-based accounts include two that are particularly relevant to learning and transfer: (a) successive uses of the same rule should show priming effects or positive transfer, and (b) performance should improve with training on relevant rules. E. E. Smith et al. (1992) found more correct tests (using the Wason paradigm) of a permission rule following another permission rule than following an obligation rule. However, the transfer may have been due to shared surface similarities in the problem, which only helped when the rules were of the same type. Johnson-Laird et al. (1972) collected think-aloud protocols in a study of transfer in the selection task and found that only 2 of 24 participants recognized the underlying similarity between specific and abstract problems. This evidence contradicts the rule priming prediction as well as the general claim that the level of abstraction of problem content should not matter.

There are mixed results on the issue of transfer. In one study, extensive training on propositional logic rules (including modus tollens) failed to improve performance on the Wason selection task (Cheng, Holyoak, Nisbett, & Oliver, 1986). A semester course in logic was equally ineffective. However, others (e.g., Fong, Krantz, & Nisbett, 1986) have found that abstract rule training, such as a course in statistics, did improve reasoning on problems involving the Law of Large Numbers (larger samples are more representative of a population). Hoch and Tschirgi (1985; described in Evans, 1989) also report that master's students perform better on this task than do bachelor's students.

4. Criticisms of the Rule-Based Account

Although the rule-based view is appealing and advantageous in some ways, the basic tenets of the view have been strongly challenged. General objections to rule-based theories include the observations that people make invalid inferences and that errors do not tend to be the trivial, irrelevant, or useless inferences that syntactic logic might predict (Johnson-Laird, Byrne, & Schacken, 1992). Along these lines, Rips (1994a) incorporates goals and subgoals to constrain rule-driven inferences. The very idea that deductive reasoning is a matter of syntactic rule application has been questioned by researchers led by Johnson-Laird (1983), who suggest an alternative approach based on the notion of mental models. Evidence shows that it is not always the case that a greater number of rules leads to a more difficult problem—for example, modus tollens is easier from a biconditional than a conditional (Johnson-Laird & Byrne, 1991; Johnson-Laird et al., 1992; Johnson-Laird, Byrne, & Schacken, 1994).

The syntactic, context-blind character of logical rules as a weak method of reasoning has been challenged by researchers (including Cheng & Holyoak, 1985) in response to a number of findings of content effects in reasoning behavior. People perform quite poorly on the abstract version of the Wason selection task (Wason & Johnson-Laird, 1972) described above. The facilitatory effects of content on the task and the lack of transfer from content to abstract problems are not effectively explained. Furthermore, the matching bias (matching the terms of the conclusion to the terms of the premises) only occurs with abstract or arbitrary materials, not content problems (Evans, 1992; Griggs & Cox, 1983; Manktelow & Evans, 1979) (Reich & Rutte, 1982). Rips (1990, 1995) suggests that content effects may be a least partially accounted for by the use of modal and deontic logics (containing operators such as PERMISSIBLE, OBLIGATORY), as well as nonlogical operations, availability effects, and analogical reasoning.

The "flawed comprehension stage" explanations of errors have inspired a number of criticisms. Johnson-Laird (1983) notes that this comprehension component remains underspecified, and as such, may be a buffer against potentially falsifying results. Further, if prior knowledge interferes with correct abstract encoding, then people should do better on abstract tasks than content-specific tasks—but clearly they do not. Lacking a faulty comprehension stage does not lessen the need to prove that people reason by natural logic (for instance, Evans, 1989, describes a heuristic-analytic mode where faulty encoding is followed by manipulations of mental models). Such a criticism may also be applied to the "flawed end processes" explanation of errors, knowledge effects (e.g., belief bias) alter performance even at the end stage of reasoning, then content-free abstract rules seem less plausible (Evans et al., 1993).

C. Reasoning by Mental Models

We begin this section with a brief clarification: the term mental models has been used in different ways. The mental models approach to deductive reasoning of Johnson
Laird (1983) and his colleagues should be distinguished from the use of mental models for reasoning in knowledge-rich domains (Gentner & Stevens, 1983). The latter version focuses on the representation and application of long-term beliefs about domains and devices. Although the two theories share some similarities, they differ in their representational and processing assumptions as well as in the core phenomena they aim to explain. In this section the term mental models will refer to the kinds of analog mental models described by Johnson-Laird and his colleagues; in cases where we contrast the two kinds of mental models, we will use the terms analog mental models and causal mental models for the sake of clarity.

1. Reasoning by Models

According to the mental models theory of reasoning (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Johnson-Laird et al., 1994), human reasoning is more concerned with the truth conditions in the world (semantics) than about logical form (syntax). The argument is that people do not reason using abstract rules, but rather they construct and combine mental models and generate inferences consistent with those models. The models are based on the given premises and on general semantic knowledge including the meaning of quantifiers and connectives. Models consist of symbolic tokens that represent the properties of entities and preserve the relations between entities. These models are essentially envisioned situations representing the truth conditions of propositions, but they are not assumed to take on any particular subjective form (perceptual, propositional, etc.); rather, it is the structure of the models that is important.

Mental model theory was originally formulated to explain syllogistic reasoning, but it has been extended fairly broadly. The reasoning process begins with the formation of models representing situations based on the premises. Consider the following deduction (from Johnson-Laird et al., 1994):

Either there is a student or a professor in the room.

Actually, there is no professor in the room.

Therefore, there is a student in the room.

The initial model structure of the premises might be represented as:

```
| s |
```

where s refers to a situation involving a student in the room, and p refers to a situation involving a professor in the room. Each separate row represents a separate model to indicate that either one or the other condition is true in the world. When one model is ruled out in the second statement, then the other model can be used as the basis for a valid conclusion.

For more complex situations with heavy demands on working memory and for reasoning with conditionals, possible models may be represented implicitly by a  

"mental footnote." These footnotes constrain the content of alternate models if they need to be made explicit. As an example of explicit and implicit models, consider the conditional, "If there is an A on the board, then there is a 2 on the board" (from Johnson-Laird et al., 1994). One may need to represent only the premises in an explicit model:

```
[A] 2
```

The ellipses represent other implicit possibilities, such as the possible existence of a model of the world that contains only a 2. The brackets represent noting that any further models must not contain an A because A is exhaustively represented in the first model.

Once the explicit mental models are formed, they are then combined with one another to form the most parsimonious description of possible situations in the world. This revision process to reduce redundancy and eliminate inconsistencies is carried out according to a program of procedural rules. Mental model theorists emphasize that these rules are not the same as abstract, deductive rules. Instead, these rules serve to translate verbal propositions to a spatial or symbolic array representation. After the explicit models have been formed and combined, conclusions can be formulated based on what holds true in the models. Useful conclusions may arise featuring information not explicitly stated in the premises. However, a conclusion is not accepted until a search for alternative models has been carried out. The specifics of this search mechanism remain sketchy (see Johnson-Laird, 1989, for a possibility). If no alternative model is found, then the conclusion is accepted as valid. Conversely, discovering an alternative means returning to the process of formulating a conclusion based on the updated models.

The criteria for evaluating the mental model theory are similar to those put forth for rule-based theories. In particular, mental model theory needs to be able to account for situations in which people reason correctly or make mistakes. It should also explain the pattern of those mistakes, as well as the existence of content effects and biases, and the determinants of difficulty and reaction time in solving problems. The mental model account posits that people are generally logical, but may fail to validate conclusions using alternate models or may construct models poorly (that is, they fail to flesh them out when required). Errors that occur should not be random, but rather should match incomplete models. People should make more errors, and take longer to solve problems that require the construction of more explicit models (since each must be checked for inconsistencies with the others). Such measures should also correlate with the complexity of the models required, which requires more working memory load and a longer processing time as well. In the Wason task, mental model theory predicts that reasoners will only consider those cards that are represented in their explicit models, and will choose only those cards
whose hidden sides appear relevant to establishing the truth or falsity of the rule (Evans et al., 1993). Hence, the logically correct falsifying card will only be chosen when models have been fleshed out to include that card; some content domains will facilitate this process more than others.

2. Support for the Mental Model Theory of Reasoning

Evidence that people can reason across levels of abstractness and familiarity (Byrne, 1991; Evans, 1977; Marcus & Rips, 1979) can be taken as support for both rule-based and mental model theories. The use of mental models is quite general, in that it only requires a knowledge of the connective or quantificational terms present in the premises and conclusions. Yet it is purported to account for specific knowledge effects as well. The mental model account is consistent with reports of matching bias, if as claimed, negative premises tend to elicit representations of corresponding positive situations (Evans et al., 1993). Likewise, the facilitative versions of the Wason task may be describable as facilitating explicit representations of negative instances. This explanation is most convincing when applied to particular manipulations, such as the reduced-array task and facilitative versions with explicit "look for violations" instructions (Griggs & Cox, 1982; Johnson-Laird & Wason, 1970; Pol- lard & Evans, 1987). Evans (1989) concludes that the belief bias is a result of "selective scrutiny"; hence, one could imagine that certain versions of both rule-based and mental model theories could accommodate it.

There is some evidence that problems are rated as more difficult when a greater number of models is required (Johnson-Laird et al., 1992). Difficulty is also linked to the complexity of the initial model (Johnson-Laird & Byrne, 1989). Baur and Johnson-Laird (1993) report that a diagrammatic format of premises (that makes alternate possibilities more explicit) leads to faster and more valid conclusions than a verbal format of premises. In the domain of discourse, participants remembered passages of text that called for a single model of a spatial layout more easily than those that were consistent with more than one model (Johnson-Laird, 1989). Additionally, participants' errors on discourse memory have been found to be consistent with the formation of mental models of described situations (Bransford, Bar- clay, & Franks, 1972).

The predictions of the mental model theory at times seem difficult to distinguish from those of the rule-based theories. However, the mental models theory, but not the rule-based theory, can explain certain findings—such as the fact that modus tollens problems are easier to solve in biconditional than in conditional form, and the greater difficulty of inferences based on exclusive, rather than inclusive, disjunction (Baur & Johnson-Laird, 1993; Evans et al., 1993; Johnson-Laird & Byrne, 1991; Johnson-Laird et al., 1994). Likewise, the observation that people will suppress both invalid and valid inferences when encouraged to think about alternate causes of a consequent can be seen to support mental model theory, but not the rule-based theory (Byrne 1989; Johnson-Laird & Byrne, 1991).

3. Learning and Transfer with Mental Models

Mental models can indeed be taught, according to Johnson-Laird (1989). He points to studies of representation of beliefs about natural phenomena (Forbus, 1983; Gentner & Gentner, 1983; Keneman, 1986; McCloskey, Caramazza, & Green, 1980) to imply that a novice-to-expert shift involves mental models. However, as noted earlier, Johnson-Laird's theory of reasoning by mental models is not interchangeable with the mental model theories under consideration in the research cited above. The different mental model theories—while not completely unrelated—make different assumptions about representation, use distinct notational systems, and make use of different expectations about the role of world knowledge in models. Nonetheless, the mental models theory of reasoning leaves room for learning and improvement. For example, Bauer and Johnson-Laird (1993) note that diagrams may help people reason because they make alternate possibilities explicit.

On the issue of development, Johnson-Laird and Byrne (1991) posit that what is innate is a capacity to build mental models of the world and search for alternative models. Linguistic abilities, the capacity of working memory, and metacognition develop through learning and maturation. The authors cite evidence that syllogistic reasoning is hampered in 7-year-olds by a lack of understanding of quantificational terms, and that somewhat older children can reason with one, but not with multiple models (Inhelder & Piaget, 1964, described in Johnson-Laird & Byrne, 1991; Johnson-Laird, Oakhill, & Bull, 1986). Rips (1994a), however, counters that basic logical operations (such as "and") could hardly be learned if one did not already possess some kind of understanding of what "and" entails.

4. Criticisms of the Mental Model Account

Studies have not always upheld the mental model prediction that difficulty is related to the number of explicit models needed (Byrne & Handley, 1992; Rips, 1990). Worse, it is not always clear how to count models in a consistent way (e.g., Bonatti, 1994). Much of the time, predictions about difficulty really do not distinguish between the mental model and rule-based theories. There have also been reported cases where the number of models needed would be intractable, but people avoid fallacies and draw inferences (O'Brien, Braine, & Yang, 1994).

Many of the criticisms of mental model theory relate either to its lack of specificity or to the possibility that procedures for manipulating models are essentially rule-based and syntactic at heart. On the first point, critics have argued that the mechanisms for comparing and falsifying models are left unspecified (Pólk, 1993; Rips, 1994b). The theory is similarly vague about how people retrieve counterexamples and at what point they flesh out their models. Another area of difficulty is the nature of the representations used. In contrast to causal mental models (e.g., Gentner & Stevens, 1983; Haldor, 1993; Keneman, Boster, & Haldor, 1994; Kiers & Bovair, 1984; McCloskey, 1983; Miyake, 1986; Schwartz & Black, 1996; Tversky, 1991), which use explicit representational conventions derived from predicate
calculus or from qualitative reasoning to represent beliefs, mental models in the
deductive reasoning tradition rely largely on an intuitive set of spatial conventions.
For example, the statement “there is a tiger and there is an ox” would be notated in
the same way as the statement “the tiger is on the left of the ox” or even “the
tiger is hungrier than the ox.”

This representational indeterminacy means that much of the interpretation is
external to the actual model; there must be distinct processes not reflected in the
model representation to deal with different logical interpretations of the same
surface form. In essence, the mental models theory has been accused of making
assumptions that are not represented in the models themselves. Johnson-Laird and
his colleagues have responded to these charges (Johnson-Laird & Byrne, 1990, 1991,
1993), noting that all representation models require interpretation. Nonetheless, the
representational indeterminacy makes it difficult to connect these analog mental
models with long-term belief systems (in which “left of” and “hungrier than” are
distinct relations). In response researchers in the analog mental models tradition can
maintain that their interest is in on-line processing, not in long-term storage. Finally,
the question arises how implicit models can be unconscious yet in working memory
that reasoning processes should not be expected to be fully introspectible. Holyoak
and Spellman (1993) argue that the very notion of an implicit model is at odds with the
“vivid” quality of mental models.

Critics have also asserted that mental models are essentially rule-based. Johnson-
Laird and his colleagues claim that procedural rules for manipulating models are not
abstract inference rules. But routines for manipulating models can be seen as abstract
rules (Braune, 1993; Rips, 1988, 1990), as, for example, they must be general enough
to recognize identity between any two tokens connected by “=.” Rips (1994b)
points out that mental model representations of some entailments are isomorphic to
natural logical propositions and can be manipulated to the same effect. The two
hypothesis have also clashed on the issue of whether mental models are really semantically
privileged (see Johnson-Laird, 1989; Rips, 1994b). Lastly, Rips (1994a) has
issued a challenge for mental model theorists to demonstrate their usefulness and
predictive power over and above that of rule-based theories, as (he argues) mental
model accounts fare no better than rule-based accounts on the issues of learnability
and the use of world knowledge.

D. Pragmatic Reasoning Schemas

The theory of pragmatic reasoning schemas (PRS) (Cheng & Holyoak, 1985, 1989;
Holyoak & Cheng, 1995) draws on many of the same assumptions as analog mental
models—including structured representations and complex inferential processes
that operate over them. The theory arose in response to growing interest in content
effects in the Wason selection task and has been tested almost exclusively within this
paradigm. According to the PRS approach, content effects in deductive reasoning
reflect the centrality of goals and knowledge structures. People do not reason with
content-free syntactic rules, nor with analog mental models, when faced with realistic
situations; rather, they access and apply certain generalized sets of procedures
that are organized by classes of goals. PRS are abstract rule sets induced by prior
experience. Unlike purely syntactic rules, PRS are context-sensitive and may be
extended to interpret both logical (e.g., “if–then”) and nonlogical terms (e.g., “predict”).

1. Reasoning by Pragmatic Reasoning Schemas

Although PRS include knowledge structures evoked by “obligations,” “causations,”
“precautions,” and others, the prototypical and most frequently studied PRS is that of
“permissions.” The Drinking Age problem described above (“If a person is drinking
beer, then the person must be over 19 years of age”) is an example of a permission
schema. Permission schemas contain a core of four abstract rules (Cheng &
Holyoak, 1985):

Rule 1: If an action is to be taken, then the precondition must be satisfied.
Rule 2: If the action is not to be taken, then the precondition need not be satisfied.
Rule 3: If the precondition is satisfied, then the action may be taken.
Rule 4: If the precondition is not satisfied, then the action must not be taken.

These rules become available when a permission schema is evoked by the context
of a problem or situation. Once evoked, such rules lead to inferences that facilitate
performance on schema-relevant forms of the Wason task (such as the Drinking
Age problem). Facilitation is likely because Rules 2 and 3 above prevent certain fallacies
(denial of the antecedent, affirmation of the consequent) that tend to occur with
abstract and arbitrary versions of the Wason task. Note, however, that facilitation
is not a foregone conclusion: logically correct performance depends on which
schema is evoked (if one is evoked at all). Schemas are used when the purpose of a
rule statement becomes recognized. Specifically, permission–obligation schemas
apply to rules with a social purpose imposed by authority.

The PRS view claims to account for the pattern of content effects observed in
the Wason task. Familiar contexts (e.g., the Postal rule for people in countries with
such a rule) are the most strongly facilitative because of past specific experience
with such rules. However, performance with unfamiliar, but realistic, rules may still
be facilitated—the appropriate schema is likely to be recognizable as a general type
even though the specific rule has never been encountered. For example, the Sears
problem (“If a purchase exceeds $30 then the receipt must be approved by the
departmental manager”) is not likely to be specifically present in memory, nor is it
particularly likely to cue counterexamples as some researchers have suggested as a possible explanation (Griggs & Cox, 1982; Mankelow & Evans, 1979; Reich & Ruth, 1982). Rather, the Sears problem evokes a general permission schema that may be familiar to many people. Arbitrary or abstract problems (e.g., the haddock/gun rule, the original Wason task) do not evoke a schema and thus produce few logically correct responses. There is evidence that in the face of many abstract problems, participants may resort to nonlogcal strategies, such as matching the form of the conclusion to that of the premises (Mankelow & Evans, 1979; Reich & Ruth, 1982). Indeed, we have noted above that the matching bias occurs with abstract problems, but is quite unlikely to occur in problems with realistic content.

The PRS approach does not account for those occasions when participants reason logically on abstract problems, PRS theorists acknowledge that syntactic knowledge structures exist, but assert that the pragmatic level takes priority in reasoning (Cheng & Holyoak, 1985). Presumably, PRS would also take precedence over any domain-independent analog mental model procedures—though little is said on this point.

2. Support for Pragmatic Reasoning Schemas

Cheng and Holyoak (1985) attempted to facilitate performance on a conditional reasoning task by providing a rationale for problems that might otherwise appear arbitrary. They used a variant of the Postal rule ("If an envelope is sealed, then it must have a 20-cent stamp") and what they termed the Cholera rule ("If a passenger's form says 'Entering' on one side, then the other side must include 'cholera'.") They also varied the familiarity of the problem content by using two populations of participants—one of which was familiar with such postal rules. Both rules were presented with context (e.g., "You are a postal clerk working in a foreign country.") and either with or without a rationale (e.g., "The rationale for this regulation is to increase profit from personal mail, which is nearly always sealed").

The pattern of results confirmed predictions: the rationale versions were highly facilitatory for both populations of participants, while the no-rationale version of the Postal rule was facilitatory only for participants familiar with the rule. Thus, in the absence of specific experience with a domain, the presence of schema evoking context resulted in a higher percentage of correct responses than did the exact problem without such context (about 60%). More striking were the results of a permission problem given an abstract form ("If one is to take action A, then one must first satisfy precondition P."). While 61% of participants correctly solved the permission problem, only 19% solved a modified version of the letters and numbers Wason task that presumably did not evoke a permission schema (Cheng & Holyoak, 1985; see also Jackson & Griggs, 1990).

Instructions to check for "violations" of a rule were found to be more facilitatory than instructions to determine truth or falsity of a rule—but only for those problems with some meaningful content (Griggs, 1983; Yachnin, 1986). Thus, highlighting potential violations may make permission-obligation schemas easier to evoke in realistic problems. However, the impact of PRS goes beyond mere violation checking. Cheng and Holyoak (1985) found that when participants were asked to rephrase rules (from "only-if" to "it-then", or vice versa), they were more likely to insert a model (e.g., "can", "must") for permission statements than arbitrary statements. This task was linguistic in nature and involved accessing declarative knowledge. Lastly, due to the semantic nature of PRS rules, this approach can account for differences in acceptability between contrapositives (switching p and q: "If not everyone will die, then the bomb does not explode"). The rule-based approach must appeal to interference from pragmatic knowledge to account for this result.

3. Learning and Transfer of Pragmatic Reasoning Schemas

As for causal mental models in general, PRS theorists claim that these knowledge structures are acquired through ordinary life experiences. Indeed, performance on selection tasks with realistic content increases systematically between 10 and 18 years of age (Giroto, Light, & Collbourn, 1988; Overton, Ward, Noveck, Black, & O'Brien, 1987; Ward & Overton, 1990). Six- and seven-year-olds can perform respectably on a reduced array (binary) permission version, and 10–11-year-olds and 14–15-year-olds show high facilitation for a full (4-card) permission version (Giroto, Gills, Bayle, & Light, 1989). Cheng and Holyoak (1985) found that training adults on formal rules was ineffectual, but training on the nature of obligations facilitated performance on conditional rules that could be seen as evoking an obligation schema (but also see Tong et al., 1986).

The transfer performance seen on the Wason task (from content problem to content problem and abstract problem to abstract problem, but usually not from one problem type to the other) has implications for PRS and rule-based theories. As noted earlier, Cox and Griggs (1982) found that an unfamiliar rule, "If a person is wearing blue, then the person must be over 19," is facilitatory only when presented after the Drinking Age problem. The Clothing Age rule is not in itself a clear permission rule—it may be interpreted as simple co-occurrence, which would not facilitate reasoning performance. However, when the Clothing Age problem is presented after the Drinking Age problem, participants may reason by analogy to the earlier problem and thus interpret the second problem in terms of permission.

4. Criticisms of the Pragmatic Reasoning Schemas Account

As mentioned earlier, the PRS view cannot account for above-chance performance on abstract tasks (Evans, 1991). Furthermore, it has been argued that permission rules are reducible to logical rules (Rips, 1988). However, such a stance becomes difficult to defend in light of findings that (a) realistic content improves reasoning performance—so much so that children perform better with permission versions than adults do with abstract versions; (b) participants rephrase PRS and abstract veri-
sions of rules differently; (c) differences in acceptability occur for different phrasings with the same logical syntactic form; and (d) and PRS training and formal logical training lead to different levels of improvement on the selection task. PRS adherents emphasize that although PRS are indeed made up of abstract procedural rules of propositional form, these rules are context-sensitive in nature. Critics have also argued that the PRS account is insufficiently specified (Oakford & Chater, 1993) and that PRS are not the only ways that knowledge can be organized and applied (Johnson-Laird & Byrne, 1991). Manktelow and Over (1991) suggest that a more likely division between facilitative and nonfacilitative content might be between deontic (if-may/must) and indicative (if-then) conditionals. Almori and Sloman (1996) assert that even nondeontic contexts that evoke “clear expectations” can be facilitatory. Rips (1990) has noted that modal logics already have terms with which to handle these types of problems (the operators PERMISSABLE and OBLIGATORY).

Another line of criticism centers around a possible confound in Cheng and Holyoak’s design (Cosmides, 1989; Griggs & Cox, 1982; Jackson & Griggs, 1990). Although the PRS rule contained explicit negatives for not-p and not-q (e.g., “has not taken action A”), the original Wason task contained only implicit negatives. This is potentially a serious confound since Evans et al. (1983) have shown that explicit negatives facilitate logical performance and reduce matching bias. Jackson and Griggs (1990) found that facilitation on abstract permission and obligation rules disappeared when implicit negatives were used (though the use of explicit negatives did not facilitate performance on the arbitrary rules). However, this criticism overlooks the fact that Cheng and Holyoak (1985) adapted the arbitrary Wason rule to fit the explicit logical form of their permission rule.

In sum, PRS occupy a middle ground between the storage of specific content knowledge and content-free abstractions. Although they are triggered by context, the reasoning rules themselves are abstract and propositional in nature. They are brought about by general learning processes that may be the same as those underlying the acquisition of categories and general memory schemas, but PRS rules are restricted in application to certain deontic contexts. PRS rules have been presented as an endpoint—no account has been offered of how learning could proceed beyond the schemas toward content-free abstractions.

E. Social Contract Theory

The social contract theory (SCT) (Cosmides, 1989) was developed as an alternative to the PRS account of content effects in reasoning. According to Cosmides (1989), other theories fall short because they appeal to content-independent processes to explain content-dependent behavior (even the PRS theory posits that content-independent processes of induction are used to include content-specific schemas). In contrast, the SCT proposes a content-specific module that is specially adapted for reasoning about social contract problems (that is, those dealing with cooperation between two parties for mutual benefit). The idea is that SC schemas came about evolutionarily due to adaptation pressures resulting from living in social groups. Human beings innately possess mechanisms and procedures that exist specifically to reason about SCs and to detect cheaters—to ensure that no benefits are taken without the appropriate costs. Because SC reasoning is modular and innate, improvements due to learning and development are difficult to reconcile.

When triggered by appropriate content, these specialized mechanisms function by calling up procedural knowledge that leads to content-appropriate inferences. As with PRS, SC schemas generate inferences that may or may not be consistent with logic. According to Cosmides (1989), the problem contents that are consistently facilitatory are those that invoke the cost-benefit structure of social contracts. Such content would induce the reasoner to look for the “cost not paid” and the “benefit accepted” cards in order to investigate potential cheaters. When these cards are the same as the logically correct cards ("p" and "not-q"), performance will be facilitated. Proponents of SCT argue that SC rules are not reducible to logical rules. They further argue that PRS owe their facilitatory effects to the presence, stated or implied, of cost-benefit structure. SC reasoning is not simple memory cuing because it occurs with unfamiliar problems that have cost-benefit structure.

Cosmides (1989) manipulated familiarity and context in four experiments to pit the SCT against rule-based and availability (memory cuing) theories. SC rules (in the form of "If you take the benefit (p) then you pay the cost (\(q\))") were contrasted with switched SC rules (in the form of "If you pay the cost (p) then you take the benefit (\(q\))"). For SC rules, the cards corresponding to p and not-q were both relevant to cheater detection and logically correct. However, for switched SC rules, the cards corresponding to not-p and q were relevant for cheater detection, but logically incorrect. Rules were presented in the context of rather lengthy stories that varied in terms of familiarity and cost-benefit structure (SC-switched SC/non-SC). Cosmides found high facilitation for unfamiliar SC problems, but little facilitation for unfamiliar non-SC problems—ruling out pure memory cuing hypotheses and contrary to predictions of rule-based views. Additional experiments (Cosmides, 1989) pitted the SCT against the PRS theory. An SC rule was tested along with a non-SC permission rule with a social purpose (that is, the rule had the action-precondition representation, but no cost-benefit structure). SC responses were about twice as frequent for the SC rules than for the non-SC permission rules—suggesting a privileged status for reasoning about social contracts. It was concluded that facilitation on permission rules in past experiments occurred because the permission rules’ context led them to be interpreted as social contracts.

In the selection task, participants performed better with instructions to look for “violators,” which encourages a search for cheaters (Gigerenzer & Hug, 1992; Griggs, 1983; Yachanin & Tveney, 1982). Further, the Clothing Age rule discussed above was not facilitatory unless it was presented after the Drinking Age rule. Proponents of SCT have taken this to mean that the cost-benefit structure of the second rule was made apparent by comparison to the first. Lastly, for rules such as "If
an employee gets a pension, then that employee must have worked for the firm for at least 10 years; participants turned over different cards depending on who they took the perspective of the employer or the employee — this corresponds to the predictions of SCT, but not formal logic or PRS (Gigerenzer & Hug, 1992; Manktelow & Over, 1991; Politzer & Nguyen-Xuan, 1992).

Despite this evidence, the SCT account faces a number of serious criticisms. Evans et al. (1993) point out that even though non-SC permission rules led to less facilitation than SC rules (Cosmides, 1989; Gigerenzer & Hug, 1992), non-SC rules still led to a much higher percentage of correct choices than in a standard abstract Wason task. Thus, SCT does not provide a full account. A major point of debate is Cosmides's classification of some PRS problems as SCs. As Cheng and Holyoak (1989) point out, Cosmides (1989) waves between a definition of a social contract as an exchange situation where a cost must be paid and one where a requirement must be met. By this latter criteria, permission rules are somewhat questionably recategorized as SCs. For example, Cosmides classifies the Drinking Age problem as an SC, though attaining a certain age cannot easily be seen as paying a cost to an individual or group. In the Choker problem (mucolysis for protection against disease) and other examples, the rule is more of a conditional obligation than a paid cost. These cases facilitate performance readily (Cheng & Holyoak, 1985; Evans, 1982, Griggs & Cox, 1982), but are not convincing as SCs. Manktelow and Over (1990) found facilitation for PRS rules that are clearly not SC rules such as, "If you clean up split blood, then you must wear rubber gloves." Such a rule does not, in any sense, imply a cost-benefit exchange or opportunity for cheater detection. Cheng and Holyoak (1989) claim that even Cosmides's own examples (such as, "If a man eats cassava root, then he must have a tattoo on his face") may be more comfortably classified as permission rules rather than cost-benefit exchanges. Cosmides and Tooby (1992) counter that the "benefit" and "cost" terms do not presuppose the values that the parties in the exchange assign to the terms.

III. REASONING BY SIMILARITY: STRONG METHODS

A sizable share of the reasoning situations that people encounter do not seem to be resolved by formal abstract procedures, but instead by strong methods that rely heavily on content knowledge to support predictive inferences (Rumelhart, 1989). Such methods are characterized by Equation (2) in which the knowledge component K plays a critical role in the flow from evidence to inferences. The evidence is often assumed to be represented in terms of the presence or absence of features or values along dimensions (but see Section IV.A for an alternative assumption that the evidence is represented structurally). Categorization and statistical likelihood estimation are examples of mechanisms that operate over stored knowledge or experience.

Similarity is frequently posited as the key psychological construct underlying strong methods (e.g., Sloman, 1996). Hahn and Chater (1998) suggest that similarity-based reasoning can be distinguished from rule-based reasoning by its use of partial matches. Although situations vary in the extent to which they match a store example or category, there is no way to partially trigger a rule. Strong methods gain power and flexibility by relying on best matches rather than perfect matches. But at the quality of the match varies, so might the quality of the resulting inference. Weak forms of reasoning have a high threshold for activation (e.g., matching a rule to the evidence), but the logical validity of the resulting inference is guaranteed.

We focus our discussion of strong methods on a subset of knowledge-driven approaches, including (a) reasoning driven by the use of category representations (b) reasoning driven by computations over data (i.e., statistical and connectionist approaches), and (c) reasoning as a set of heuristics and biases. In section IV, we focus on the strong method of reasoning by analogy and consider wider implications of comparison processes.

A. Reasoning by Categorization

Categorization itself can be thought of as a reasoning process in which the category assignment is the conclusion and the factual evidence serves as premises (Holland, Holyoak, Nisbett, & Thagard, 1986) or in which the properties of an instance count as evidence to be explained by a theory-based category (Murphy & Medin 1985). However, we focus in this section on the property-level inferences that can be made by assigning instances to categories. Once category membership has been established, the knowledge stored in the category representation becomes a resource for generating inferences about the new member.

This inductive potential varies with the nature of the category representation. The number, kind, and validity of candidate inferences depend on how categories are structured (see Medin & Heit, Chapter 3, this volume, for detailed discussion of the various accounts). Some theorists posit that categories consist of sets of defining criteria. An alternative representational assumption is that categories are structured by prototypes that store information about the central tendency across members. In this case, the candidate inferences are based on which features are likely to be present or what values the dimensions toward, but these are reasonable guesses with no guarantee of validity. A third view is that category representation is exemplar-based; membership is determined by similarity to one or more specific exemplars. In this case, properties of particular exemplars are candidate inferences. A more recent viewpoint considers categories to be intuitive theories made up of properties organized by explanatory principles. In this case, candidate inference may be warranted by convincing causal links rather than correlational history.

The theory view of categories has provided a framework for investigations of what kinds of inferences are drawn about what kinds of categories. It has been argued (Carey, 1985; Keil, 1989) that people's understanding of natural kind categories, such as biological categories, is grounded in theories that specify the causal basis or nature of the category. Gelman and Coley (1991) suggest six properties that
characterize natural kinds categories: (a) rich inductive potential—the capacity to generate inferences; (b) nonobvious basis—they capture deep similarities that might not otherwise be noted; (c) essence—a unique core property responsible for the surface properties (see also Medin & Ortony, 1989); (d) existence of anomalies—reflecting the idea that core properties, not surface properties, constitute the essence of natural kinds; (e) division of linguistic labor—the recognition that expert knowledge of a category's basis may exceed one's own; and (f) corrigibility—the belief that as theories are revised, theory-laden categories can change. In contrast, artifact concepts—such as bicycle and grater—are typically structured around functional characteristics and, along with nominal kinds—such as triangle, bachelor, and pet—offer a narrower range of inductive potential than natural kinds (Kell, 1989, 1991b).

Within a domain, instances can be categorized at multiple levels of abstraction. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) interpreted the superordinate, basic, and subordinate levels of categorization in terms of varying degrees of informativeness (these levels may be specific to American undergraduates; see Medin & Heit, Chapter 3, this volume, for more about differing taxonomies). According to Rosch, the more specific the categorization, the more inferences are made, though the subordinate level does not add much beyond the privileged basic level. The few inferences made at more subordinate levels (e.g., if X is an animal, then it has a body) may be the most likely to be valid, but the least likely to be useful.

Although it is commonly assumed that taxonomic inheritance from superordinate to subordinate categories is a basic and relatively general reasoning process, recent evidence suggests that the story may be more complex. Sloman (1998) has shown that participants assigned to inductive projections with arbitrary properties did not, in fact, consistently follow taxonomic inference rules. For example, they might fail to infer that a property true of a superordinate category (e.g., electronic equipment) was true of its subordinate category (e.g., stereos). Instead, the similarity of the two categories predicted judgments. Sloman argues that this calls into question the inferential strategy of the rule “It belongs to Category X, so it must have the properties of Category X” and instead prompts consideration of a similarity-based strategy such as “It is like Example X, so it is likely to have the properties of Example X.”

**B. Reasoning Across Categories: Category-Based Induction**

Categories can also be useful for reasoning about other categories. Say a reasoner needs to know whether a particular property is true of dogs, but no relevant information is included within their concept of dogs. If the status of the property is known with regard to another category (e.g., cats), this other category can serve as a basis for inductive reference. The reasoning process in this case is not a matter of categorizing an instance as a cat (as described above), but instead amounts to using what is known about cats to evaluate a possible property of dogs.

Gelman and her colleagues have carried out extensive research on the development of category-based induction (CBI) (e.g., Davidson & Gelman, 1990; Gelman, 1988, 1989; Gelman & Coley, 1991; Gelman & Markman, 1986, 1987). For example, Gelman and Markman (1986) studied children's inductions using a paradigm in which a target instance was given the same label as one standard, but was highly similar in appearance to a second standard. The task pits perceptual similarity against category membership (as indicated by the common label) in order to ask whether children would use inductive projection based on the category given the two standards. Young children (like adults) tended to extend the property from the case with the shared label—demonstrating reliance on category membership as an inferential basis. Perceptual similarity was a factor, but shared category dominated.

Davidson and Gelman (1990) obtained a different pattern of results using novel objects (e.g., a gnu-like animal), novel labels, and unfamiliar properties (e.g., “has four stomachs”). Children were taught a property of one animal, and then asked whether or not the property would be present in another animal. The researchers found that 4- and 5-year-old children made more inferences for animals that were perceptually similar to the standard (about 75%) than for those that were perceptually dissimilar to the standard (about 45%). There was no effect of shared labels whether novel or familiar. In a version in which the correlation between similarity and common label in the stimulus set was increased, common label did have an effect when supported by appearances. When there was no conflict between labels and appearances, children based their inferences on appearances. Children's inductions appear to be influenced both by perceptual similarity and by common category labels.

Osherson, Smith, Wilkie, Lopez, and Shafir (1990) conducted an extensive study of adult CBI using arguments of the form “Cats have property P. Therefore mammals have property P!” Inductive arguments of this kind can be classified as general or specific. In general arguments, like the one above, the conclusion category is more general than (i.e., includes) the premise category/ies. In specific arguments, the conclusion category lies at the same level of specificity as the premise and belongs to the same superordinate category/ies (Osherson et al., 1990). An example is: “Cats have a left aortic arch. Therefore, badgers have a left aortic arch.”

The dependent measure in this paradigm is the strength of the argument that the property is true of the conclusion given that it is true of the premise(s). Note that the property in question is unlikely to be part of the representation of the conclusion category. The use of blank predicates (Osherson et al., 1990) sets up the task as a comparison between the premise and conclusion categories rather than evaluation of the validity of the property relative to the conclusion category. Although this clearly qualifies as a strong method of reasoning because the results are heavily dependent on category representations, it is (by design) less knowledge-intensive and more reliant on general processes than reasoning by categorization—the semantic content of the inference (the blank predicate) is disconnected from the semantic content of the conclusion category.

Osherson et al. (1990) have identified a constellation of phenomena in which the strength of category-based inductive arguments varies systematically with the
nature of the premise and conclusion categories and the relationship between them. In the similarity phenomenon, greater similarity between the premise and conclusion categories leads to greater argument strength. For example, given that mice have a property, one is more likely to attribute the property to rats than to elk. In addition, general arguments are stronger when they are based on more typical premises. This typicality phenomenon extends Rapoport’s (1975) finding of increased likelihood of generalization with greater typicality of the premise category. General arguments are also stronger with more homogeneous conclusion categories. Thus, in the homogeneity phenomenon, more specific conclusion categories lead to greater argument strength (robin → bird is stronger than robin → animal). However, Shafir, Smith, and Osherson (1990) describe an effect called the inclusion fallacy, according to which more general conclusions seem stronger than specific ones. In normative terms, people should be less willing to make a more wide-ranging conclusion, but given a premise about robins, the inductive argument for ostriches is sometimes considered weaker than the argument for birds (all birds including ostriches).

Using multiple premises, Osherson et al. (1990) collected evidence for a further set of phenomena that apply to both general and specific arguments. They found that the diversity of the premise categories predicts argument strength. If a property is known to be true for a set of far-ranging creatures (e.g., wasp and deer), it is more likely to apply to a new case than if it is known to be true of a cluster of more similar creatures (e.g., antelope and deer). Also, adding an additional premise at the lowest level category that includes the other premises and conclusions increases argument strength. This is called the monotonicity phenomenon since more evidence leads to greater argument strength. For example, collies, poodles, dalmations → dogs is stronger than an argument based on only two of the three premises. On the other hand, given a third premise from an outside category (collies, poodles, dragonflies → dogs), then the argument is often judged to be weaker than (collies, poodles → dogs). This phenomenon is known as nonmonotonicity because the argument strength decreases with the additional premise.

Osherson et al. (1990) propose a similarity-plus-average model to account for this set of findings. The strength of an argument depends, first, on the similarity of the premise and conclusion categories. In addition to accounting for the similarity phenomena described above, the similarity component of the model can account for the inclusion fallacy if the similarity between the general conclusion category (birds) and the premise category (robins) is greater than the similarity between the specific conclusion category (ostriches) and the premise category (robins). The second component in the model is average, the average similarity of the premise categories to the lowest-level category that spans both the premise and conclusion categories. Intuitively, argument strength increases with the extent to which the categories in the premises cover different areas within the space of the inclusive category (Lopez-Gelman, Guthed, & Smith, 1992). In an argument such as (robins → bluejays), the degree of coverage is the average similarity of robins to exemplars of the lowest inclusive category, bird. An argument like (trout → bluejays) should be less persuasive than (robins → bluejays) because the coverage between trout and animal—the lowest inclusive category in this case—is low. Coverage is critical to explaining the multiple-premise effects. Monotonicity is explained by assuming that the additional within-category premise increases the coverage of the inclusive category. Nonmonotonicity (as in the example used above) occurs because the additional premise from outside the category (dragonflies) forces the lowest-level inclusive category up to the level of animals rather than dogs. The category of animals is less well covered (by collies, poodles, and dragonflies) than the category of dogs is by collies and poodles alone. Less coverage means less argument strength, despite the greater number of premises.

Sloman (1993) offers a different interpretation of this set of reasoning phenomena, based on comparison of features, not categories. His feature-based induction model is instantiated as a connectionist model. The premises of an argument are encoded in terms of the strength of associations from input units representing the features of the premise categories to an output unit representing belief in the predicate. The strength of an argument is tested by applying the features of the conclusion category to the input vector to determine the degree of activation of the output unit. The prediction is that the degree of feature overlap between the premise and conclusion categories determines the argument strength. Similarity implicitly guides the process since like inputs lead to like outputs, but there is no explicit computation of similarity.

Sloman (1993) successfully modeled most of the core phenomena demonstrated by Osherson et al. (1990). For example, he explained diversity effects in terms of feature overlap, rather than by category coverage. Diverse premises lead to stronger arguments because they provide greater coverage of the feature space (more of the excitatory feature-to-output connections are activated). Sloman’s model predicts an additional phenomenon related to diversity called feature exclusion. Using a same-level conclusion category (weasel) rather than superordinate (mammal), diverse premises (foxes and rhinos) do not lead to greater argument strength than clustered premises (foxes and deer) because the premises do not provide broader coverage of the feature space of the conclusion category.

Sloman argues that the feature-overlap model is superior because (a) it is not dependent on the existence of stable category structures or explicit similarity computations, and (b) the core phenomena are explained by a unified account rather than requiring the two components of similarity and category-coverage. However, the feature-based approach does not account for the nonmonotonicity phenomenon (without an additional assumption of feature competition). Additionally, critics point out that Sloman’s emphasis on features is a somewhat dangerous move because there is no satisfying basis for determining the set of representing features. Furthermore, recent arguments in the study of conceptual structure (Markman & Wszolek, 1997; Murphy & Medin, 1985) and comparison processes (Markman & Gentner, 1993a; Medin, Goldstone, & Gentner, 1993) have emphasized the
importance of structured representations of causal and relational information. We will return to the role of structured representations and comparison processes in section IV.

C. Statistical Inference

Associative or statistical accounts of reasoning are based on the updating of statistical parameters via observational learning. In this manner, prior experience is used as a basis for prediction. Statistical regularities such as feature occurrences (frequencies) and co-occurrences (correlations) can be applied to predict unknowns as a function of the presence or absence of other features. Stored associations based on keeping track of occurrences or observable properties are an alternative to explicit knowledge structures such as rules, schemas, or concepts. Apparent subjectivity in people's interpretations of the statistical structure of a domain challenges the assumption that people construct accurate statistical models of the environment. For example, Chapman and Chapman (1969) found effects such as illusory correlations in which participants consider a correlation to be stronger or weaker than its objective basis due to the mediating influence of a prior belief or goal.

Bayes's theorem provides a formal framework for generating optimal inferences based on probability information. Any proposition (or feature, hypothesis, event, etc.) can be assigned a prior probability of being true. This likelihood of the proposition in a neutral context provides a base rate for prediction. Bayesian inference is the process of computing likelihoods as a function of the priors along with conditional probabilities representing the likelihood of a proposition being true given that one or more other propositions are true. Because there is a complex network of dependencies and interactions among propositions, effective prediction often involves a process of maximum likelihood estimation over a set of parameters.

Anderson (1991) explored the use of such a Bayesian framework as a psychological account grounded in evolutionary adaptation. In his "rational analysis," category membership is treated like any other feature that could be predicted via conditional probabilities. Category structure is constrained by the overall goal of optimizing inference performance. Predictions are made by combining the likelihoods of all possible classifications and their subsequent inferences. Anderson's model predicts a number of behavioral results in human categorization (e.g., prototypic effects, linear separability, base-rate effects, and more). But the Bayesian framework requires the simplifying assumptions that features are independent of one another and that categories are nonoverlapping. The psychological plausibility of this approach has been questioned by Ross and Murphy (1990), who show that participants generally make inferences on the basis of one best-fitting category, instead of using a probabilistic summation across multiple possible categories.

Statistical accounts of reasoning are challenged by evidence that people often fail to conform to the prescriptions of Bayesian inference. A classic example is base-rate neglect or the failure to take appropriate account of the prior probability in assessing a likelihood (Kahneman, Slovic, & Tversky, 1982; A. Tversky & Kahneman, 1981). For example, even if someone is 90% certain they just saw a duck-billed platypus, before believing them, it would be worth considering how likely it is in general for such a creature to make an appearance. In probability estimation tasks, it has been shown that participants tend to respond as though the base rate information were not present or did not matter. However, Holyoak and Spellman (1993) suggest this phenomenon may not be robust. There are a variety of circumstances involving increased salience or causal relevance of the base-rate information in which proper use of base rates is made (e.g., Bar-Hillel & Fischhoff, 1981). Additionally, much of the empirical evidence for base-rate neglect uses the methodological approach of establishing the base rates by providing summaries rather than a range of experience to allow statistical learning. Manis, Dowalina, Avis, and Car- doze (1980) show that base rates acquired through presentation of exemplars are used effectively.

D. Reasoning by Heuristics

Cognition may be effectively characterized as informal and subject to a variety of biases. A. Tversky and Kahneman (1973, 1983) have carried out an important research program exploring natural reasoning heuristics. They have shown that people do not always follow normative processes in their judgments and decisions. Instead of optimal performance (as established by either statistical or logical frameworks), reasoning often reflects the use of heuristic approaches that provide reasonable resolutions at minimal cost of time and processing resources. These shortcuts are not guaranteed to provide a correct answer (or any answer), and they are often linked to systematic biases in performance. Barsalou (1992) suggests that reasoning mechanisms may best be explained in terms of the interaction between a system capable of deductive logic and three pervasive cognitive biases that reflect the way information is compared (representativeness), brought to bear (availability), and stored (in organizing knowledge frames).

Much of the evidence for these effects comes from studies of probability estimation in which the outcome of the reasoning process is not a new belief, but a degree of confidence in a particular belief. We briefly discuss two such phenomena that reflect the role of stored knowledge. One example is the conjunction fallacy (A. Tversky & Kahneman, 1983), in which people seem to rely on a representativeness heuristic rather than probability theory to make a judgment. The likelihood of a conjunctive event that is typical of its kind is judged as greater than the likelihood of one of the single constituent events by itself. Despite the fact that logically the co-occurrence of two events cannot be more likely than the occurrence of one of the events alone, people judge (for example) that it's more likely that Bjorn Borg lost the first game but won the set than that he lost the first game. In this case, knowledge about representative cases takes precedence over an analysis grounded in formal likelihood estimation or logic. Another example is the availability heuristic—
the use of easily accessible knowledge to make a judgment (A. Tversky & Kahneman, 1973). For example, in judging the probability of a car being blue, experience with blue cars that is recent or salient often increases the probability estimate. Again, people's ability to collect and apply probability information objectively is called into question. In these examples, the reasoning process seems to reflect reliance on knowledge content rather than strict use of domain-general principles.

E. Connectionist Inference

Connectionist models provide several interesting interpretations of reasoning within the framework of parallel, distributed processing. In such systems, a large number of simple processing units loosely inspired by the networks of neurons in the brain take on activation values as a function of input from the environment along with weighted signals from other connected units. Recently, the statistical nature of connectionist models has been emphasized (see Smolensky, Mozer, & Rumelhart, 1996) to demonstrate the formal groundwork on which the models operate. This emphasis reflects the challenges connectionism has faced, such as (a) it is difficult or impossible to tell exactly how or why the models do what they do, and (b) they can be made to do nearly anything so they are uninteresting. Jacobs, Jordan, Nowlan, and Hinton (1991) describe a connectionist design principle for creating systems known as mixture models that instantiate principles of Bayesian inference.

Connectionist models are also intriguing to many researchers for their psychological and neuropsychological plausibility rather than as a way of doing statistics. Such models are often trained using a learning algorithm to perform pattern association or a function mapping between a set of inputs to a set of outputs. The systems produce inferences by generalizing from training experience. A novel input is treated like (or leads to the same output as) the training instance(s) with which it shares the most in common or a configuration of particularly critical features—similar inputs lead to similar outputs (Rumelhart, Durbin, Golden, & Chauvin, 1996).

Connectionist models based on recurrent architectures perform a type of processing known as constraint satisfaction involving a series of small, local adjustments toward eventually settling at a stable global state. Examples of such systems as psychological models include Holyoak and Thagard's (1989) ACME model of analogical mapping and Rumelhart, Smolensky, McClelland, and Hinton's (1986) account of schema-based processing (see Holyoak & Thagard, 1995, for further discussion of connectionist vs. symbolic paradigms in thinking and hybrid connectionist-symbolic accounts such as ACME). In these setting systems, the activation level of processing units represent degree of belief in hypotheses about properties of the environment. Connections between units represent associations or the degree to which one property is consistent with or predicts the other. The activations of the units of the system are continually updated to effect small increases in the overall goodness (consistency of the configuration of global activation with prior knowledge as stored in the connections between units) of the interpretation of an input. Such systems, along with pattern association devices, have the interesting properties of completing partial patterns of input activation with inferred values in a manner reflecting the statistical structure of the training data and even overcoming incongruous hypotheses. Rumelhart (1989) discusses how a setting system of this sort featuring a soft clamping mechanism on the inputs will naturally exhibit a range of performance along a continuum from memory to reasoning. Depending on the match between the input and the knowledge stored in the weights, the system will amplify a recognized pattern, complete a partial pattern, or generalize from appropriately similar examples to a novel variation.

IV. STRUCTURAL ALIGNMENT IN REASONING

Analogy can be defined as the perception of relational commonalities between domains that may be dissimilar on the surface, or as a kind of reasoning based on the assumption that two things that are similar in some ways will be similar in others. In fact, the two definitions are related. Relational correspondences between a base and target often lead to further candidate inferences. Rips (1990) noted that analogy occupies an intermediate position with respect to his strict-true criteria, as discussed above. Analogy also holds an intermediate position in the weak-strong framework for reasoning we have set forth. Because it is clearly knowledge-intensive, it qualifies as a form of strong reasoning, yet the processing mechanisms are fairly well specified and are relatively independent of other cognitive processes. (Of course, as with other forms of reasoning, the representations over which it operates are influenced by other processes.)

In this section, we begin by outlining the characteristics of analogy and its relative, similarity, with a particular focus on the underlying mechanism of structural alignment and mapping. Then we consider how these processes can inform our understanding of the strong and weak methods discussed above. The perspective raised by this consideration of comparison processes suggests a general view of reasoning as a continuum from strong to weak methods, rather than a dichotomy between systems. We speculate on the unifying claim that structural alignment processes may not only guide strong forms of reasoning, but may contribute to the development of weak methods as well.

A. Reasoning by Analogy

Analogy research has focused mainly on the mapping process used to establish correspondences between two situations. A familiar situation, referred to as the base or source analog, is used as a model from which to map inferences to the unfamiliar situation or target. Such mappings can be decomposed into two subprocesses: (a) alignment of the two representational structures, and (b) projection of inference from one to the other. The subprocesses are intimately linked because the nature of
the alignment process constrains the candidate inferences that result. We present a basic overview of analogical mapping and inference. In addition, we explore how the structural alignment account of analogy can be extended to other comparison processes such as similarity. Lastly, because the process of reasoning by analogy about a target domain often depends on accessing potential base analogs stored in memory, we briefly discuss the retrieval component of analogical thinking.

1. Analogy: Theoretical Claims and Models

According to Gentner’s (1983, 1989; Gentner & Markman, 1997) structure-mapping theory, analogical mapping is a matter of establishing a structural alignment between two represented situations and then projecting inferences. The theory assumes the existence of structured representations made up of properties (such as objects and attributes) and relations connecting the properties. An alignment consists of an explicit set of correspondences between the sets of representational elements of two situations with a focus on matching relational predicates. The alignment is determined according to a set of constraints that guarantee structural consistency: (a) there must be one-to-one correspondence between the mapped elements in the base and target, and (b) there must be parallel connectivity such that the arguments of corresponding predicates also correspond. In addition, the selection of an alignment is guided by the systematity principle: a system of relations connected by higher-order constraining relations such as causal relations is more preferred in mapping than an equal number of independent matches. Thus, if analogical similarity is a matter of common relational structure, then a base domain with a richly linked system of connected relations can yield candidate inferences by guiding completion of the corresponding structure in the target (Bowdle & Gentner, 1997). The systematity principle underscores a preference for coherence and causal predictive power in analogical processing. Table 1 (adapted from Gentner & Markman, 1997) lists seven key phenomena of analogical mapping.

Table 1. Seven Phenomena of Analogy

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Structural consistency</td>
<td>Analogical mapping involves one-to-one correspondence and parallel connectivity.</td>
</tr>
<tr>
<td>2. Candidate inferences</td>
<td>Analogical inferences are generated via structured completion.</td>
</tr>
<tr>
<td>3. Systematity</td>
<td>People prefer connected relations rather than collections of isolated relations.</td>
</tr>
<tr>
<td>4. Relational focus</td>
<td>Relational matches are made whether or not the objects making up the relations also match.</td>
</tr>
<tr>
<td>5. Interactive interpretation</td>
<td>The interpretation of an analogy depends on both terms: the same term yields different interpretations in different comparisons.</td>
</tr>
<tr>
<td>6. Multiple interpretations</td>
<td>Analogy allows multiple interpretations of a single comparison.</td>
</tr>
<tr>
<td>7. Cross mapping</td>
<td>Adults generally perceive both interpretations of a cross-mapping and typically prefer the relational interpretation.</td>
</tr>
</tbody>
</table>

Adapted from Gentner and Markman, 1997.

There are three main factors that contribute to the role of analogical mapping in analogical reasoning. These three factors are: (1) the representation of analogical structures, (2) the mapping process, and (3) the inference of analogical relations. The representation of analogical structures is the process of encoding the analogical structures in memory. The mapping process is the process of aligning the analogical structures with the target structures. The inference of analogical relations is the process of drawing the analogical relations from the aligned analogical structures.

As a part of the mapping process, analogy can act as a mechanism for learning or knowledge change through re-presentation of the constituent predicates of the analogs or highlighting the common structure between analogs (C. A. Clement & Gentner, 1991; J. Clement, 1988; Gentner, 1989; Holyoak & Thagard, 1989). Re-presentation (or adaptation) involves altering the representation of one or both
analogies to improve the quality of the match (see Holyoak, Novick, & Melz, 1994; Kass, 1994; Keane, 1996; Kowalski & Gentner, 1996). Schema abstraction occurs when a common system arrives at in understanding an analogy is rendered salient—thereby increasing the possibility that it will be used again later (Gick & Holyoak, 1983; Hayes-Roth & McDermott, 1978). Through these mechanisms, analogy can promote the formation of new relational categories and abstract rules.

2. Systematicity as a Constraint on Inference

The role of relational structure in analogical processing is more specific than a global preference for relational commonalities over attribute or object matches. As noted above, the analogical interpretation process seeks matches consisting of interconnected systems of relations. This preference for interpretations of analogies that align systems of predicates connected by higher-order constraining relations is known as the *systematicity* principle (Gentner, 1983, 1989). This claim that comparison acts to promote systems of interconnected knowledge is crucial to analogy's viability as a reasoning process. If the comparison process generated a set of isolated features, there would be no natural basis by which to constrain the inferences derived from the match.

A useful experimental methodology for observing the effects of structure in comparison takes advantage of *cross-mapping* in which structural commonalities conflict with object matches (Gentner & Toupin, 1986; Goldstone & Medin, 1994a; Markman & Gentner, 1993b; Ross, 1987). For example, in a comparison between “Spot bit Fido” and “Fido bit Rover,” Fido is cross-mapped. When presented with cross-mapped comparisons, participants can compute both alignments (see Table 1). Adults (though not young children) typically prefer the relational alignment. The preference becomes more salient when the cross-mapping includes higher order relational structure (Gentner & Rattnermann, 1991; Gentner & Toupin, 1986; Markman & Gentner, 1993b).

In order to test the role of systematicity as a constraint on inference, C. A. Clement and Gentner (1991) showed participants analogous scenarios and asked them to judge which of two lower-order assertions shared by the base and target was most important to the match. Participants chose the assertion that was connected to matching causal antecedents—their choice was based not only on the goodness of the local match, but also on whether it was connected to a larger matching system. In addition, inferences projected from one scenario to the other were governed by systematicity—inferences were made in order to complete a causal system.

3. Similarity Is Like Analogy

The framework developed for analogy extends naturally to literal similarity (Gentner & Markman, 1993, 1995, 1997; Goldstone, 1994b; Goldstone, Medin, & Gentner, 1991; Markman & Gentner, 1993a, c; Medin et al., 1993). Specifically, the alignment of relational structure is also crucial to similarity comparisons. The distinction between analogy and literal similarity can be thought of as within a similar space defined by the degree of object-attribute similarity and the degree of relational similarity (Gentner, 1989). Similarity occurs when comparisons exhibit a high degree of relational similarity with very little attribute similarity. As the amount of attribute similarity increases, the comparison shifts towards literal similarity. Literal similarity matches are easier to make than analogy because the alignment of relational structure is supported by object matches.

Markman and Gentner (1993b) found evidence that similarity comparisons induce a structural alignment. Participants viewed pairs of pictures—in one scene a woman was shown giving food to a squirrel; in the other, a similar-looking woman was shown receiving food from a man. One group of participants rated how similar the two scenes were to each other, while a control group rated the two scenes’ aesthetic value. All participants were then asked to map the woman in the first picture to an element of the second. Participants who had first rated the similarity of the scenes made significantly more relational mappings (i.e., woman to squirrel) than did participants in the control condition. The very act of carrying out a similar comparison can induce a structural alignment and increase the likelihood of making matches on the basis of shared relations rather than object similarity.

In addition to relational focus, the critical finding that systematically guided inference (C. A. Clement & Gentner, 1991) also carries over to similarity comparisons. Bowdle and Gentner (1997) gave participants pairs of similar scenarios (without distinguishing a base or target) and asked for inferences. It was found that participants preferred to make inferences from a systematic structure to a less systematic structure and also judged comparisons to be more informative in this direction of the reverse.

4. Analogical Retrieval

So far our focus has been on analogical mapping once the base and target have been established. Explaining the use of analogy and similarity in reasoning requires some account of how potential analogs are accessed in long-term memory. There is considerable evidence that similarity-based retrieval, unlike the mapping process, more influenced by surface similarity than structural similarity. Strong similarity content effects seem to dominate in mindings and to limit the transfer of leamings across domains (Gentner, Rattnermann, & Forbus, 1993; Holyoak & Koh, 1987; Keane, 1988; Novick, 1988a, 1988b; Reed, 1987; Ross, 1984, 1987, 1989).

In Gick and Holyoak's (1980, 1983) classic studies, participants often failed to access potentially useful analogs. The rate of successful solutions of a very difficult problem tripled (from a baseline of 10%) for participants given an analogous story prior to the problem; but even so, the majority of participants failed to benefit from the analogy. However, when these nonsolvers were given a hint to think about the story they had heard, the solution rate approximately tripled again to 80–90
Because no new information was given about the story, it can be concluded that the analog was available in memory, but was not spontaneously retrieved. The structural similarity between the story and the problem was sufficient to carry out the mapping with both analogs present in working memory, but not sufficient to produce spontaneous retrieval.

To test this functional distinction between kinds of similarity, Gentner et al. (1993) gave participants a large set of stories to remember and then later provided new stories that varied in their surface and relational similarity to the originals. Participants were asked to write out any original stories they were reminded of—"the remembrances that resulted were strongly governed by surface commonalities such as similar characters. However, when asked to rate the similarity and inferential soundness of pairs of stories, the same participants relied primarily on higher-order relational commonalities, such as matching causal structure. Participants even rated their own surface-similar reminiscences as poor matches. This dissociation is also found in problem-solving tasks: reminiscences of prior problems are strongly influenced by surface similarity, but structural similarity better predicts success in solving the problem (e.g., Ross, 1987).

B. Reasoning by Categories Revisited.

At the outset of this chapter, we suggested that structural alignment could serve as a unifying framework for a discussion of reasoning. We now begin our effort to make good on this proposal. We have seen how the comparison processes of analogy and similarity can function as a mechanism for extending knowledge. Similarity and categorization are quite intricately and somewhat controversially linked to one another—in ways ranging from accounts of one in terms of the other to dissociations between them (see Medin & Heit, this volume). We leave the debate aside except to note that if similarity and structural alignment is "explanatory" as a source of constraints on category coherence (Goldstone, 1994a) and hierarchical organization (Markman & Winn, 1997), then inferences that arise from categorization (as in section III. A) may in fact be derived from structured comparison between instances and category representations (see Kurtz & Gentner, 1998). The use of CBI (section III. B) to extrapolate knowledge on the basis of category structure is generally considered distinct from similarity-based inference (the projection of knowledge from one specific instance to another). Nonetheless, viewing CBI from the vantage of structural alignment yields some useful insights.

1. Comparison and Category-Based Induction

Most models of CBI utilize a notion of similarity as the degree of match between sets of features. Evidence suggests that adults' property induction is guided by structural similarity rather than flat similarity (as in Osherson et al., 1990) or featural overlap (as in Sloman, 1993). Lasaline (1996) found that argument strength increased when there was a causally connected inference that could be carried over as a candidate inference. For example, adding the relational premise "For Animal B, a weak immune system causes an acute sense of smell" increased the strength of the following argument:

Animal A has a weak immune system, skin that has no pigment, and dry flaky skin
Animal B has a weak immune system and an acute sense of smell

Therefore, Animal A also has an acute sense of smell.

Further evidence that people are sensitive to connected systems of relations in induction comes from Heit and Rubenstein (1994). They found that participants make stronger inferences when the kind of property to be inferred (anatomical or behavioral) matches the kind of similarity between the animals (anatomical or behavioral). For example, participants made stronger behavioral inferences from tuna to whales (because both swim) than from bears to whales. Stronger anatomical inferences were made from whales to bears (because both are mammals). If we assume that anatomy and behavior are represented by different systems of semantic relations, then these findings support the claim that adults are strongly influenced by relational focus and systematicity in drawing inductive inferences.

Wu and Gentner (1998) found further evidence for structural alignment effects in category-based inference. Participants were given descriptions of two standard animals and a target animal and asked to make an inference about the target based on one of the standards. The target shared one property with Animal A and two properties with Animal B; thus its featural similarity was higher with B than with A. However, the description of Animal A also included the information that the shared property causes an additional property. Participants strongly preferred the inference connected to causal structure (from Animal A) despite the greater feature overlap favoring the inference from Animal B. The pattern of findings suggests that people carry out structural alignment during CBI.

Given this evidence, the question arises whether structural alignment might also explain the patterns of responding found in studies of CBI discussed above. We focus our discussion on the nonmonotonicity effect (Osherson et al., 1990; Lopez et al., 1992) in which inference strength goes down with the addition of a distant premise. For example, Argument 1 below is stronger than argument 2:

Argument 1: (crow, peacock) → bird
Argument 2: (crow, peacock, rabbit) → bird

This runs contrary to the monotonicity prediction that increasing the number of premises should increase inductive strength. Osherson et al. (1990) argue that nonmonotonicity effects are due to category coverage. By enlarging the size of the covering category, the average similarity of the argument categories to the covering
category is diluted. However, consider Sloman's (1993) example in which Argument 1 is stronger than Argument 2:

- Argument 1: (crocodile) \rightarrow alligator
- Argument 2: (crocodile and king snake) \rightarrow alligator

This effect is not predicted by coverage. Note, however, that crocodile has a richer match with alligator than that arising from the alignment of crocodile and king snake. Some instances of nonmonotonicity may reflect structural alignment of the premise categories (Gentner & Medina, 1998). Consider again the "bird" example from above. In Argument 1, the premises (crow and peacock) are strongly alignable—yielding a rich "premise schema" that can project inferences to the conclusion category with which this schema is also strongly alignable. The addition of a difficult-to-align premise (rabbit) in Argument 2 invites a retreat from strong alignment-based reasoning. The reasoning process could rely on this less forceful alignment—one that must incorporate rabbit in with crow and peacock—or instead shift to taxonomic reasoning. This shift can be seen as going from a strong reasoning process (close alignment) to a somewhat more domain-general (or less strong) reasoning process (category-based inheritance). Category-based inheritance is less dependent on the specific knowledge structures associated with the entities than is literal similarity alignment. For example, tiger and paramecium are equally entitled to inherit the properties of animal, even though tiger yields a more satisfying alignment with animal than does paramecium.

2. Analogical Induction in Development

Even more than adults, children often need to reason in the absence of useful background knowledge. Indeed, it is sometimes stated that the inductive processes used by children are comparison based—as opposed to the CBI used by adults (Carey, 1985; Inagaki, 1989). Lopez et al. (1992) used inductive inference problems like those of Odherson et al. (1990) to demonstrate a developmental pattern of early reliance on similarity with category-based reasoning entering later. Both kindergarteners and second graders showed similarity-driven effects such as the influence of premise-conclusion similarity and typicality of the premise. However, only the second graders showed category coverage effects such as premise diversity and monotonicity. As in the Odherson et al. studies, it was found that adults appeared to use both comparison-based and category-based processes. Adults with different kinds of expertise showed different patterns of inductive reasoning in the domain of trees (Medin, Lynch, & Coley, 1997). Participants with expertise in ecological or causal knowledge (maintenance workers and landscapers) tended not to use diversity to guide their reasoning. This may reflect direct comparison of sophisticated knowledge structures rather than the use of the more domain-general (weaker) method.

Carey (1985) asked children to make inductive inferences about properties. For example, children were told that a dog had a spleen and then asked if bugs we likely to have a spleen. She found that before age 10, children tended to base the inductive attributions of biological properties on the similarity of the target object to humans. This pattern suggests that children might be using a well-understood species (i.e., their own species) to reason about less familiar species; that is, the were drawing an analogy. The research program of Inagaki, Hatano and colleagues (Inagaki, 1989, 1990; Inagaki & Hatano, 1987, 1991; Inagaki & Sugiya, 1998) suggests that (a) children use structure-mapping processes in inducing new features of animals, but (b) the flexibility and sophistication of their reasoning is limited by incomplete domain representations. Inagaki and Hatano (1987) examine children's spontaneous use of analogy in inductive problems from the biology domain. They asked 5–6-year-old children questions like, "What would happen if a rabbit were continually given more water?" As in Carey (1985), the children often made explicit analogies to humans such as: "We can't keep it [the rabbit] forever in the same size. Because, like me, if I were a rabbit, I would be 5 years old and become bigger and bigger." The personification responses were often reasonable and tended to be more correct than non-personification responses. When asked questions for which the analogy with humans would yield incorrect responses, children were far less likely to use the analogy. In addition, children were more likely to use the analogy to humans for target entities more similar humans. This is consistent with high similarity facilitating alignment and inference projection.

Viewing this process as analogy suggests that children based their inferences of knowledge of humans because their knowledge of humans is deeper and more systematic than their knowledge of other animals. If this is the correct account, then the same pattern should hold for other animals about which children possess expertise. Inagaki (1990) found that children who raised goldfish were highly likely to use their knowledge about goldfish to make inductive inferences about unfamiliar aquatic animals such as frogs. Interestingly, the goldfish-raising children also used the person analogy for frogs almost twice as often as the control subjects. It is possible that the experience of raising goldfish had led the children to compare humans and goldfish and to abstract underlying commonalities. Goldfish raisers were more likely to draw analogies from people to goldfish than non-goldfish raisers, despite the fact that they demonstrably knew more about goldfish. Deeper knowledge about animals apparently led children better able to notice and use cross-species analogies (see also Kotsowsky & Gentner, 1996). Developmental gains in inductive accuracy may result through increasingly sophisticated analogical comparison rather than a shift from comparison to category use. The distinction lies between comparison based on overall similarity and more informed analogical comparison which similarity with respect to specific causal or functional systems determines common system and the projected inference.
C. Reasoning by Rules Revisited

Deductive inference rules are the quintessential case of a purely syntactic, content-independent (weak) form of reasoning. Their operation is in principle unaffected by the semantic content of the representations to which they apply. Does comparison have any role to play in such processes? One suggestion from norm theory (Kahneman & Miller, 1986) is that people use analogical reasoning to compare a target situation to ad hoc counterfactual alternatives (see also E. E. Smith & Osherson, 1989). In this section we seek to establish three links between structural alignment and deductive rules (Gentner & Medina, 1997, 1998). First, we suggest that rule application typically entails structural alignment. Second, people are often more accurate with concrete analogies than with the structurally equivalent rule. Third, alignment processes may provide a route to acquiring deductive rules.

1. Structural Alignment in Rule Application

The use of rules in reasoning may actually depend on comparison processes. Wittgenstein's (1953) discussion of rule following emphasizes the interdependence of similarity and rules. He argues that it is a mistake to think similarity can be established independently of rule-governed activities because similarity is unconstrained. For example, given the series "2, 4, 6, 8, ..." it is indeterminate what should come next. Our sense of similarity alone might suggest monotonic increase, even numbers, or single-digit numbers. A rule such as "+ 2" is what provides a constrained basis for continuing the number sequence. Goodman (1972) adopted this line of argument to attack similarity as a basis for concepts. However, Wittgenstein argued the converse point as well—similarity plays a role in the application of rules. A represented rule like "+ 2" lacks a basis for determining when and how to apply the rule to specific instances (see Gentner & Medina, 1998, for more on this and related points).

The application for a deductive rule requires a firm structural binding of the abstract components of the rule with the specific premise and conclusion assertions. This binding or unification of rule and evidence can be seen as an alignment process. Imagine a continuum from analogy to relational abstraction. Both involve overlap in relational structure, but they vary in the concreteness of the base domain. In analogical mapping the objects of the concrete base must be aligned with the objects and relations of the target. When an abstraction acts as the base, there are only variables to fill rather than concrete objects to overcome. Thus, an abstract rule that is clearly represented by the learner should be easy to align.

2. Analogies as Surrogates for Rules

E. E. Smith et al. (1992) note that deductive rule following should be at least as accurate with abstract or unfamiliar materials. This assumption is challenged by content effects in reasoning. Participants given Wason's (1968) selection task typically fail in the application of conditional inference rules to abstract and unfamiliar material, but perform far better if familiar concrete materials are used. Additionally, an important pattern of transfer effects has been established using the selection task. The canonical result is that subjects revert to poor performance when they receive the abstract (i.e., letters and numbers) version of the task, even following a facilitating context problem, such as the Drinking Age problem (Cox & Griggs, 1982; Johnson-Laird et al., 1972; Wason & Shapiro, 1971). However, Cox and Griggs (1982) also demonstrated that an arbitrary specific rule, such as a Clothing Age rule ("If a person is wearing blue then the person must be over 19") shows facilitation after presentation of the Drinking Age rule.

The facilitation effects found in the Wason task are usually ascribed to "familiarity" with the rule and material used. Our framework allows us to be more specific and distinguish two different sources of facilitation: (a) transparency—the overall similarity between the prior knowledge representation and the representation of the conditional statement to be tested; and (b) systematicity, or the availability of higher-order structure that supports the application of the conditional inference. The role of transparency, or overall similarity, is straightforward. Successful transfer hinges on recognizing that the two problems are similar. As discussed above, analogical transfer between two specific problems is rare without surface similarity (or explicit hints). Thus, the high similarity between the Drinking Age scenario and the Clothing Age scenario leads to analogical reminding of the former given the latter, and facilitates positive transfer. But when subjects are given a specific case followed by an abstract case, the surface similarity may be insufficient to lead to a reminding.

The second factor is systematicity—being reminded of a stored scenario is only useful if that scenario contains a system of higher-order relations that is correlated with the structure of the rule (Gentner et al., 1993; Holyoak & Koh, 1987; Ross, 1987, 1989). For example, Wason and Shapiro's thematic rule "Every time I go to Manchester I travel by car" produces facilitation because people have well-established planning schemas that link particular destinations with modes of transport. These schemas have the desired implicative structure (One is not tempted to infer $q \rightarrow p$: "If I am traveling by car I am going to Manchester"). These can be mapped onto the conditional scenario to provide the appropriate constraint. In contrast, a rule such as "Every time I go into the kitchen I wear my brown shoes"—even though it contains highly familiar elements—should produce little if any facilitation for the conditional rule, perhaps in part because $p \rightarrow q$ and $q \rightarrow p$ are equally plausible (or implausible) in this scenario. Maximal facilitation requires not only familiar elements but appropriate relations among the elements.

Although experimental studies that demonstrate strong facilitation effects typically combine transparency and systematicity (e.g., Griggs and Cox's drinking task and Johnson-Laird et al.'s postal rule task), these effects are conceptually separable (Gentner & Taurin, 1986). For example, Johnson-Laird and Shapiro (using a different set of reasoning problems) found that realistic relations facilitate perfor-
rance when they correlate with the logical structure of the problem, but hinder
the deduction process when they do not (Wason & Johnson-Laird, 1972).

As expected on this account, children are highly sensitive to context in reasoning
problems. Girotto et al. (1988) gave children a simplified version of the selection
problem following Cheng and Holyoak's (1985) paradigm; they used obligation and
permission rules accompanied with brief rationales, as well as arbitrary rules. All the
rules were unfamiliar to the children, and they were introduced in a game situation
with toy bees and an imaginary beehive. An obligation rule, for instance, was, "If
a bee buzzes, then it must stay outside," followed by the rationale that the queen
bee wanted to avoid spreading the disease to baby bees. An arbitrary rule was, for example,"If a bee buzzes, then it is outside." Children were then asked which of the bees
should be checked (i.e., those inside, those outside, those that buzz, those that don't).
As expected, children performed better with meaningful rules than with arbitrary
rules (70% vs. 11% correct in 9- and 10-year-olds).

3. Structural Alignment and the Acquisition of Rules

Analogies during learning can lead to highlighting and abstraction of common
structural systems; thus comparison can orient learners towards systems of inter-
connected knowledge (e.g., systems linked by higher-order causal, mathematical, or
perceptual relations). We suggest a progression from reasoning based on overall,
literal similarity to reasoning based on higher-level abstract similarity — that is, in the
limit on rule-application. With repeated alignment and abstraction, similarity com-
parisons evolve from being initially perceptual and context-bound to become
increasingly framed in terms of common higher-order structure. The structural
alignment and mapping process grades naturally from highly concrete and literally
similar comparisons to purely abstract comparisons. This predicted pattern parallels
a developmental trend from overall similarity to relational similarity and abstract
mappings, which Gentner and Rattermann (1991) called the order of similarity. As
Quine puts it (1969, p. 167), we "retain different similarity standards . . . for use in
different contexts." Of course, comparison is only one of many learning mechanisms
involved in the route to abstract cognition, but its potential role in bridging
strong and weak methods of reasoning is unique. In the current framework, rule-
governed processes are based on abstract structural similarity, but they may coexist
with processes governed by more concrete alignments.

Kotowsky and Gentner (1996; Gentner, Rattermann, Markman, & Kotowsky,
1995) investigated the possibility that comparison processes might promote
children's learning about higher-order perceptual relations such as symmetry or monotonicity. In particular, they focused particularly on the ability to perceive cross-
dimensional matches in abstract representation, which require an abstract appreciation of common relational structure. Prior research shows a relational shift: preschool children are generally unable to appreciate higher-order perceptual patterns such as symmetry and monotonicity (Chapman, 1977; Halford, 1987, 1992; Smith, 1984, 1989, 1993). The mate-

tials were perceptual patterns—groups of simple shapes that could be perceived readily — so that higher-order relational commonalities could be manipulated inde-
pendently of object similarity. Children were shown a standard embodying son
relational structure such as symmetry (e.g., XoX) and were asked to say which of to
other figures it was most similar to. The two alternatives were (a) the relational (co-
rect) choice, which shared a higher-order relation (e.g., symmetry) with the stan-
dard (HiH); and (b) a foil that was composed of the same elements as the relation
choice, but was rearranged to remove the higher-order pattern (HiH). Since both
choices were equally dissimilar to the standard in terms of object attributes, and t
relational choice exceeded in relational overlap, it was (by adult standards) the ec
winner.

The key variable was the degree of concrete lower-order similarity between t
standard and the relational choice. For same-dimension trials, the relational cho
had the same dimension of change as the standard: the match was the same
and easily perceivable (e.g., big-little-big/big-little-big). In cross-dimension trials, t
match was solely at the abstract higher-order level (e.g., big-little-big/dark-light
dark). When given mixed sets of these similarity traits (without feedback), 4-year
olds were correct on the within-dimension (close similarity) trials, but chose ra
domly on the cross-dimension trials. However, simply presenting children with
concrete "easy" matches before the abstract cross-dimensional matches led
significantly better performance on the cross-dimensional matches. Kotowsky a
Gentner (1996) suggest a mechanism of progressive alignment whereby initially con
crete, dimensionally specific representations are rendered more abstract by comp
gression and alignment. Even for novices, close matches are easy to perceive beca
they are, in a sense, automatically aligned. This alignment results in a slight hi
lighting of the common relational structure. After repeating such alignments, t
higher-order relational structure becomes strong enough so that a partial match c
be made even in a cross-dimensional pairing.

Applying this notion of progressive alignment to deductive reasoning leads
to the prediction that content effects should occur at a range of different levels
abstraction from highly specific through progressively more abstract. The level
abstraction achieved will vary with learner and with topic or domain. This predi-
tion that facilitation can occur not only with highly specific scenarios, as discussed abo
but also with abstract schemas. This seems to be the case. As Cheng and Holyo
(1985, 1989) have argued, the higher-order relation of transitivity can abstrac
ly represented in a schema that would include rules such as "If A is taken, then Preconditi
must be satisfied." Cheng and Holyoak (1985) test performance on a selection problem that described a permission situation abstract using the rule "If one is to take action A, then one must first satisfy preconditi
". Subjects gave 61% correct answers in the abstract permission problem, in con
trast with a 19% success rate in the Wason card problem for the same subjects. T
facilitation effect of systematicity with abstract stimuli was almost as strong as t
obtained with concrete similarity (e.g., 81% in Johnson-Laird et al., 1972).
Cheng and Holyoak's proposal that pragmatic reasoning schemas are "abstract knowledge structures induced from ordinary life experiences" (p. 395) is consistent with our proposal of a continuum of abstraction. However, there are some differences between their pragmatic approach and the present framework. First, they contend that the schematic structures that guide everyday reasoning derive primarily from experience with classes of goal-related situations. We suggest more broadly that the essential element is a well-established higher-order relational schema that supports conditional reasoning. The pragmatic dimension perse is not the source of facilitation. Goal-oriented contexts are undoubtedly a rich source of meaningful schemas, but (a) many regularities are learned across a variety of different goal scenarios, making it unlikely that their structure is derived from one particular type of goal; and (b) humans attend to many kinds of regularities in the world, not just to those that influence goal achievement. We learn many higher-order relational schemas that are not fundamentally goal-oriented (e.g., causality, perceptual higher-order relations, mathematical relatedness, and so on). For example, consider this rule:

If a pattern is symmetric, it has some identical components.

Although this rule could hardly be described as goal-driven, it seems likely to result in correct performance in a selection task. It is reasonably intuitive to resist the "affirming the consequent" fallacy— that is, to see that "some identical components" does not imply "symmetric"; and it is also intuitive to accept the contrapositive ("denying the consequent, the correct inference")— that is, to see that "no identical components" implies "not symmetric."

A second difference is that we do not suggest that pragmatic reasoning schemas are the end point of learning. With experience and instruction, people can and do develop abstract schemas that are not embedded in pragmatic contexts. For example, Rips and Conrad (1983) found that a one-quarter course in elementary logic substantially improved subjects' ability to evaluate propositional arguments, many of which contained conditionals. For subjects sufficiently trained in conditional logic, the implications of complex rules may be readily available. Finally, whereas Cheng and Holyoak assume no special role for language in this evolution, Gentner and Medina (1998) suggest that the acquisition of relational language may play an important role in analogical processing (Gentner & Rattermann, 1991) and relational abstraction (Kotovsky & Gentner, 1996).

D. How Do the Different Views of Reasoning Fit Together?

We have organized our discussion of reasoning along a continuum with two extremes. Weak methods of deductive reasoning consist of the application of content-free syntactic inference rules that operate on the logical form of representations. Strong methods of reasoning rely heavily on specific experience and knowledge representations. Some researchers have proposed that similarity-based and rule-based processes are both important in human cognition, but that they function as different cognitive systems. Smolensky (1988) draws a distinction between two different mechanisms: a conscious rule interpreter that functions algorithmically, and an intuitive processor that operates at the subsymbolic level. Perhaps the most clearly articulated proposal is Sloman's (1996) argument for the existence of two separate systems of reasoning that operate independently and in parallel. The associative system encodes covariation of features in the environment and makes predictions based on statistical regularities. Sloman's account restricts similarity to these associative, subsymbolic processes, while the rule-based system operates on structured symbolic representations.

As evidence for the existence of two independent systems of reasoning, Sloman cites the inclusion fallacy and the conjunction fallacy, where similarity and rules lead to contradictory conclusions (Shafir et al., 1990; E. E. Smith & Osherson, 1989; A. Tversky & Kahneman, 1983). In the "Linda the bank teller" example of the conjunction fallacy (A. Tversky & Kahneman, 1983), participants judged that Linda (who, as a student, "was deeply concerned with issues of discrimination and social justice") was more likely to be "a bank teller and active in the feminist movement" than "a bank teller." One explanation of this phenomenon is that participants were swayed by the greater similarity of the description of Linda to a "feminist bank teller" than to the typical bank teller (E. E. Smith & Osherson, 1989). Sloman notes that people are often simultaneously attracted to both of two contradictory conclusions—the correct "bank teller" solution and also the "feminist bank teller" solution—and interprets this as evidence for two parallel systems of reasoning.

However, as Gentner and Medina (1998) note, contradictory responses can be generated within a single comparison-based reasoning system in at least three ways. First, the retrieval process may produce more than one possibility for a given contextual cue. Second, even after the pair to be aligned has been selected, the local-to-global alignment process (see section IV.A.1) can lead to contradictory responses over time. Early responses are dominated by local object matches, whereas later responses reflect an alignment of relational structure (Falkenhainer et al., 1989; Goldstone & Medin, 1994a; Ratchiff & McKoon, 1989). Goldstone (1994a) found evidence for this temporal shift from object matches to relational matches using a deadline-same-different task. A third way that contradictory responses may arise is that the same comparison (even with the same correspondences) can give rise to alternative interpretations. For example, the statement that a given battle "is the mother of battles" could mean that it is the biggest (as a parent is larger than her offspring) or that it will engender a host of others (which may be larger than the parent). People can experience simultaneous awareness of these possibilities. In all three of these cases, the contradictory responses arise within one system.

A more fundamental difficulty with Sloman's proposal is that classifying all similarity with association neglects the evidence that the comparison process is structure-sensitive (Gentner & Clement, 1988; Gentner & Markman, 1994, 1997; Goldstone & Medin, 1994a; Markman & Gentner, 1993c, 1996; Medin et al., 1993).
When similarity is considered as structural alignment, a range of reasoning performance can be understood as a continuum between strong and weak methods. Furthermore, structural comparison can act as a bridge by which similarity-based processes lead to the development of abstract rules. This view of gradual abstraction of initially conservative, context-specific representations is consistent with the proposal that abstractions can arise from comparison across highly specific instances (Cheng & Holyoak, 1985; Elio & Anderson, 1981; Forbus & Gentner, 1986; Gentner & Medina, 1998; Gick & Holyoak, 1983; Medin & Hixon, 1989). As Gentner and Medina argue, there is a graceful learning continuum from a fully concrete mapping, in which the objects transparently match their intended correspondents, to an analogous mapping in which a relational structure is imported to a new domain with no support (or even with conflict) from the object matches, to a fully abstract mapping in which the base domain contains variables, the target contains objects, and the mapping qualities as rule application.

V. SUMMARY AND DISCUSSION

Before closing it is worth briefly noting the rise of new approaches that may lead to rapid change in the field. First, research in cognitive neuroscience may eventually make it possible to link the higher-order representations and processes posited in psychological accounts of reasoning to their neurobiological instantiations. At this point the exact form of the connection and which techniques (e.g., neuroimaging) will provide constraints on theories of high-level cognition are hard to predict.

A second area from which accounts of reasoning stand to benefit is cross-cultural research. Strong forms of reasoning would be expected to show cultural and developmental differences due to variations in the knowledge brought to bear. For example, Lopez, Atran, Coley, Medin, and Smith (1997) found that inductive reasoning among the Hua-Mayans was much like that of American subjects in making heavy use of similarity and typicality, but different in that Mayans did not make use of premise diversity. When premise variability was high, Mayans often drew on their ecological knowledge concerning relations among the creatures. The pattern of performance differences may reflect Mayans' superior knowledge of their ecology. Choi, Nisbett, and Smith (1999) found that Koreans were less likely than Americans to use categories for inductive inference when presented with specific arguments in which the covering category must be generated. Nisbett and his colleagues suggest that this difference may reflect different reasoning styles; in particular, a general pattern of greater inferences on categories among Europeans and Americans than among East Asians. As another example, East Asians have been found to focus relatively more on situational factors than Europeans and Americans in making causal attributions (Choi, Nisbett, & Norenzayan, 1997).

Weak methods might be expected to be more stable and universal, because they are domain-independent and, according to some researchers, originate mutually.

However, cross-cultural research suggests caution in the belief that particular weak methods are fundamental or universal in cognition. For example, Carraher, Carraher, and Schliemann (1985) used naturalistic methods to study mathematical knowledge in real-world contexts. Brazilian schoolchildren who worked as street vendors were skilled at arithmetic performance in the marketplace when the problems were formulated in the context of fruit sales, but could not solve equivalent problems in the abstract version used in formal instruction. The mathematical ability of these children seemed to be situated in a particular context.

Within and across cultures, there is evidence for both strong and weak methods of reasoning. Neither form can be relegated to a peripheral role in reasoning. It is clear that people can and do apply domain-general rules such as modus ponens; it is equally clear that they are not the whole story. As Newell and Simon noted, when stored data from relevant experience is available, people typically rely on it rather than invoking a more abstract, universal algorithm. Strong methods driven by cases, categories, statistical summaries, and heuristics are a useful way to get on with the goals of understanding and prediction. We agree with Holyoak and Spellman (1993) that "the psychological difficulty of inferences seems to depend more on the relationship between the content of the premises and prior knowledge than on the logical form of the reasoning involved" (p. 292).

The structural alignment view points toward a way of linking strong and weak methods. Structural alignment and projection can guide the application of either concrete or abstract prior knowledge to new situations, yielding inferences beyond the available evidence. Progressive alignment processes can also lead to the development of new abstract forms. Comparison-based computation takes advantage of stored knowledge as a strong method, but with a specificity of processing usually associated only with weak methods. This potential for compatibility among various methods may be a promising sign for a unified account of human reasoning.

References


Polk, T. (1994). Mental models, more or less. Behavioral and Brain Sciences, 16.


To account for human knowledge, one must specify the innate initial state and mechanisms intervening between birth and the stable adult state. Thus, seven foundational issues in cognitive science are developmental questions at heart. These include: all facets of the nativist-empiricist debate. Philosophers who silenced the debate were not interested in development per se, but rather were interested in fundamental issues within epistemology. They saw that no account of human knowledge could stand if one cannot imagine, at least in principle, how knowledge is acquired. Nativists such as Kant and empiricists such as Hume appealed to learnability (among other things) as a criterion for accepting any account of the nature of the human mind.

The modern study of cognitive development engages these same debates, and studies of development provide one source of data relevant to characterizing cognitive architecture. Developmental studies provide evidence for specialized learning mechanisms in different content domains, and also provide a unique window on such classic issues as how to draw a principled distinction between perception and cognition. In addition to constraining our theories of adult cognition, development poses problems of its own. Those discussed in this chapter include continuity versus discontinuity within cognitive development, existence of critical periods in development, and the role of maturation in cognitive development.