Annu. Rev. Psychol. 1993. 44:265–315

THINKING

Keith J. Holyoak and Barbara A. Spellman

Department of Psychology, University of California, Los Angeles, California 90024

CONTENTS

INTRODUCTION

Reviewing an active field of research is a bit like writing an unauthorized mid-career biography. Your subject is not about to reveal its secrets to you, or even to stand still long enough to allow a coherent story to be constructed. The task is made especially difficult when the topic is as amorphous as thinking. As Oden put it in his prior review for this series, "Thinking, broadly defined, is nearly all of psychology; narrowly defined, it seems to be none of it"

0066-4308/93/0201-0265$02.00

(1987:203). Sometimes thinking is construed as a synonym for all "intelligent information processing," and sometimes it is construed as the umbrella term for a range of processes associated with "high-level" cognition, such as reasoning, categorization, and judgment and decision making. We emphasize the latter conception of thinking, but our chapter is not organized around those traditional subtopics, each of which has been reviewed in its own right (see Medin & Smith 1984; Payne et al 1992; Rips 1990). Rather, we are guided by a piece of folk psychology. Rips & Conrad (1989) found that lay people believe that virtually all "everyday" mental activities (e.g. reasoning and remembering) are kinds of thinking, and that thinking is a part of each kind of mental activity. In keeping with the perceived ubiquity of thinking in cognition, we review a number of general themes that have emerged in recent research on the topic, drawing a sprinkling of examples from work in a variety of subareas.

In surveying the field of thinking, three recent trends seem particularly noteworthy. 1. The rise of the connectionist paradigm has led to a critical reexamination of assumptions concerning the symbolic nature of human thinking. 2. Cognitive psychologists are taking seriously the notion that human thinking may be based on two very different systems; and there have been increased efforts to use evolutionary arguments, as well as biological evidence, to constrain cognitive theories. 3. Theoretical efforts have been directed at explaining how thinking is constrained by the content of what is thought about, and by the context in which thinking takes place. Our review is organized around these three themes, which are interconnected in various ways.

CONFLUENCE OF SYMBOLIC AND CONNECTIONIST PARADIGMS

Two Contrasting Paradigms

Our first theme is not simply the rise of connectionism, but rather the meeting and merging of two theoretical streams that have been channeled into the analysis of thinking. Human thinking (along with language) has generally been viewed as the *sine qua non* of symbolic mental activity. Since the cognitive revolution in the mid-20th century, thinking has been characterized as the product of a "physical symbol system." In 1990, Simon concluded that "The physical symbol system hypothesis has been tested so extensively over the past 30 years that it can now be regarded as fully established, although over less than the full gamut of activities that are called 'thinking'" (p. 3).

This sanguine assessment has been challenged, however, by those who have developed alternative "subsymbolic" paradigms, such as Hofstadter (1984), Rumelhart, McClelland, and the PDP Research Group (1986), and Smolensky (1988). Connectionist models, the most common instantiations of the subsymbolic approach,[1] consist of networks of relatively simple processing units connected by links. Processing involves a series of cycles of activity; on each cycle, units take on new states of activation as a function of their own prior activations, the activations of units to which they are connected, and the weights (excitatory or inhibitory) on the interconnecting links. Typical connectionist models embody some or all of four central ideas. First, control is distributed over the network of units, rather than localized in a central "executive." Second, knowledge is to varying extents distributed over sets of units, rather than identified with single units. Third, decision making is based on parallel constraint satisfaction, by which successive cycles of processing tend to converge on an activation pattern that best satisfies the constraints embodied in the weights on links. At convergence, the units with highest activations tend to support each other and inhibit their competitors. Fourth, learning consists of incremental revision of weights on the basis of either feedback concerning the performance of the network or internal constraints on weight patterns.

These characteristics of connectionist models contrast with the prototypical features of serial production systems, the style of model most closely associated with the symbolic approach to modeling cognition (Newell 1973). In a "classical" production system, knowledge is encoded locally in "condition-action" rules, perhaps coupled with a declarative semantic network (e.g. Anderson 1983). A central executive selects a single rule to fire on each processing cycle. When the condition of a rule is matched and that rule is selected to fire, then the action specified by that rule will be taken. The global behavior of a production system is more naturally characterized as serial generation of an action sequence, rather than parallel satisfaction of multiple constraints. Learning primarily consists of the addition of new productions, a process that requires the intervention of an executive controller that decides what new rules to build and when to build them (e.g. Anderson 1987; Rosenbloom et al 1991). It should be emphasized that some production systems developed in recent years, such as CAPS (Just & Carpenter 1992) and SOAR (Rosenbloom et al 1991), depart from the "classical" architecture in important ways; nonetheless, the above contrast captures in broad strokes the differences between the models associated with each of the two paradigms.

The symbolic and connectionist paradigms bear a rough but interesting correspondence to two different perspectives on thinking that have coexisted (with some degree of tension) over this century. The symbolic paradigm was shaped in large part by Newell & Simon's (1972) treatment of problem solving. Their approach primarily focused on "well-defined" problems, for which the problem solver knows at the outset what goal is to be achieved, what the

[1] Other types of subsymbolic models include classifier systems (Holland et al 1986; see Druhan & Mathews 1989, for a psychological application), and models based on flexible semantic networks such as Hofstadter's (1984; Hofstadter & Mitchell, 1993) Copycat model of analogy.

starting state is, and what operators are potentially relevant to achieving a solution. Problems that meet this description (such as proving geometry theorems or solving logical puzzles) approximately satisfy a "closed world" assumption: The pool of knowledge relevant to their solution, although possibly large, is nonetheless circumscribed. Newell & Simon characterized explicit thinking as conscious serial search through a specifiable space of possibilities, based in large part on heuristics that evaluate incremental progress toward goal attainment. In addition, they stressed the role of rapid recognition processes that match external inputs against knowledge stored in long-term memory (Chase & Simon 1973). Production systems emerged as the model that most directly embodied Newell & Simon's characterization of thinking.

In contrast, the earlier Gestalt psychologists, such as Duncker (1945) and Wertheimer (1945), stressed the solution of problems that are less well defined, and hence do not satisfy the closed world assumption. In working on a particular new "target" problem, for example, a reasoner may be reminded of a better-understood analogous "source" problem, perhaps drawn from a substantially different knowledge domain. The source analog may then suggest new goals or operators that might be used to solve the target problem. Because the bounds within which a potentially useful source analog may be found are not clearly circumscribed, analogical thinking can violate the closed-world assumption. More generally, Gestaltists emphasized that thinking may involve parallel integration of knowledge based on mechanisms that are largely unconscious, sometimes producing a "restructuring" of the problem representation.

The theoretical ideas of the Gestaltists were notoriously vague, and Simon (1986; Kaplan & Simon 1990) has shown that the symbolic paradigm can accommodate many of the empirical phenomena associated with such Gestalt concepts as "intuition" and "insight." Nonetheless, some alternative approach might provide a computational realization of the Gestalt perspective on thinking. As Rock & Palmer (1990) have pointed out, there is some affinity between Gestalt theory and current connectionist models. In particular, connectionist networks perform "soft" constraint satisfaction (i.e. each constraint has some influence on the overall behavior of the network, but is not as inviolate as a hard-and-fast rule). A constraint network based on partially convergent and partially discrepant knowledge may yield a coherent interpretation of a situation, so that "the whole is different from the sum of its parts."[2]

[2] The broad current interest in connectionism within psychology is in part attributable to the fact that, in different ways, it captures some of the flavor of both Gestalt psychology and behaviorism, the main intellectual rivals that dominated early 20th-century psychology. Roughly, parallel constraint satisfaction is reminiscent of Gestalt ideas, while learning by incremental weight adjustment over distributed representations is reminiscent of associationist conceptions of learning (see Estes 1991, for the latter perspective). Current cognitive psychologists generally hold basic conceptions of cognition that have been shaped to a large extent by reactions to the Gestaltist and associationist legacies. As a consequence, connectionism offers something for almost everyone to love and/or hate, in a mixture that is a function of selective focus and intellectual predispositions.

These two perspectives on thinking tend to bring with them different views of the relationship between "high-level" thinking and the broader spectrum of information processing, which includes perceptual and motor components. The symbolic approach has dealt most directly with "central" cognition, either leaving aside the problem of modeling input and output processing, or attempting to press models of high-level cognition downwards to serve also as models of perception and action (e.g. Anderson 1983; Rosenbloom et al 1991). Even theorists who acknowledge that "one thing wrong with much theorizing about cognition is that it does not pay much attention to perception on the one side or motor behavior on the other" (Newell 1990:159) are wont to find themselves, for the pragmatic reason that peripheral processes are highly complex, "committing this same sin" (p. 160).

In contrast, Gestalt psychologists emphasized that high-level thinking is in many ways akin to perception. Similar views have been expressed by recent proponents of the subsymbolic approach, such as Hofstadter (1984; Hofstadter & Mitchell, 1993). Lakoff (1993) reviews a wide range of linguistic evidence suggesting that human understanding of such abstract concepts as time, categories, and causality is based on metaphors derived from perceptuomotor experience. (See Mandler 1992 for a discussion of the implications of this view for cognitive development.) Perceptual and motor processes, as well as basic memory processes, clearly evolved long before high-level human cognition. A general principle of evolutionary biology is that mechanisms that initially evolved to serve one function may later be coopted to serve other functions (a type of change termed "exaptation"). Thus from an evolutionary perspective, it is reasonable to conjecture that the mechanisms of high-level cognition have important links to those that evolved earlier to support perception and action. As we note below, recent analyses of "implicit" cognition have drawn attention to the evolutionary development of human thinking (e.g. Reber, 1992). Connectionist models, which have been developed primarily in the context of work on perception and motor control (e.g. Jordan & Rosenbaum 1989), tend to encourage "outer to inner" theorizing, in which models of peripheral processes are extended in attempting to account for more central processes, rather than the reverse.

Systematicity and Symbols

It is unlikely, however, that connectionism will undermine the traditional view that human thinking requires a symbol system. The most fundamental argument for the necessity of symbolic representations was presented by Fodor & Pylyshyn (1988; Fodor & McLaughlin 1990). They pointed out that knowledge is systematic in the sense that the ability to think particular thoughts seems to imply the ability to think certain related thoughts. For example, if a person understands the meaning of the concepts "love," "boy," and "girl," and can understand the proposition "The boy loves the girl," then it would seem extremely bizarre if the person were nonetheless unable to understand the proposition "The girl loves the boy." More generally, it seems characteristic of thinking that if each concept in a set of potential constituent concepts is understood, and a relation structure (such as a frame for a predicate and its arguments) can be instantiated by one assignment of the constituent concepts, then the thinker can also instantiate the relation structure with other permissible assignments of the concepts. The need to represent this kind of systematic relational information was part of the motivation for Minsky's (1975) concept of frames, a type of symbolic relation structure that continues to be influential in modeling human cognition (Barsalou & Hale, 1992).

Systematic reasoning with composable constituents requires symbols. Newell (1990) describes the workings of a representational system: It can encode an external situation and external transformations; it can internally apply the encoded transformation to the encoded situation; and it can decode the result back to the environment—thereby predicting the external result of applying the transformation. (See Palmer 1978 for an earlier discussion of the nature of representations.) A representational system must be sufficiently flexible to predict the effects of all the distinct external situations and transformations that are important to the organism. Newell argues that as the diversity of the knowledge that an organism must represent and manipulate increases, it becomes increasingly difficult to find specialized representational systems to provide appropriate encodings. In what Newell terms "the Great Move," evolution developed a representational system that enables more complex representations to be composed from simpler ones.

This representational system must be able to share knowledge across many different contexts, because it will be impossible (owing to physical limits of the storage system) to store all the information potentially required for every task in a form in which it is immediately available. A "symbol" is fundamentally a locally available code that can provide access to distal information relevant to a task. In a symbol system, information acquired in one task context has the potential to be made available in a different task context. This is exactly what is required for systematic reasoning with composable constituents. In our example above, we can understand both "The boy loves the girl" and "The girl loves the boy" because the concepts "girl" and "boy" are represented in a manner that keeps each distinct from both the "lover" and "beloved" contexts; both are therefore available for use in either context. When multiple task contexts permit access to a shared pool of knowledge, by virtue of constituency relations, performance in one context will be systematically related to performance in others. The ability to use systematic relational knowledge across contexts enables analogical reasoning about novel situations (Falkenhainer et al 1989; Holyoak & Thagard 1989).

Systematicity is also a key feature of rules of inference, such as "If X sells Y to Z, then Z owns Y." Smith et al (1992) propose several empirical criteria that may reveal when some knowledge used in human reasoning is coded as abstract rules. One criterion, which applies for at least some well-established rules, is that it seems just as easy to draw inferences about unfamilar instantiations—including nonsense ones—as about familiar ones. Thus if we are told that "Henry sold the floogle to Sam," we immediately conclude that Sam now owns the floogle, whatever a floogle might be. The inference follows directly from the role that "floogle" plays in the argument structure of the rule, without any requirement that floogles resemble familiar objects that have been transferred from one owner to another.

Systematicity of relational correspondences (i.e. of correspondences between the arguments of multi-place predicates) also plays a role in judgments of perceptual similarity (Goldstone et al 1991). For example, suppose people are shown three pairs of geometric shapes, with each pair arranged vertically. One pair consists of two identical triangles, one of identical squares, and one of identical circles. People tend to evaluate the pair of triangles as more similar to the pair of squares than to the pair of circles. But if a square is now added as a third form below the two items in each of the pairs, the evaluation of similarity reverses: Two triangles and a square are viewed as less similar to three squares then to two circles and a square. This similarity reversal reflects differences in relational correspondences: Both the first and the third triad can readily be represented as "two same forms plus a square," whereas the middle triad is most naturally represented as "three squares." Thus the first and third triads match better in terms of relational correspondences. Goldstone et al demonstrated not only that systematic relations matter to similarity, but also that relational matches matter more (relative to matches of one-place predicates, such as "square") as the overall relational overlap between two complex figures increases. Their findings are difficult to interpret in terms of feature models of similarity (most notably Tversky's 1977 contrast model) that do not specify a role for systematic relational correspondences in similarity judgments. (Connectionist models that are implementations of feature models also fail to capture relational aspects of similarity.)

If thinking depends on symbol systems, as the arguments of Fodor & Pylyshyn (1988) and Newell (1990) imply, connectionist models of thinking face the formidable challenge of implementing symbolic processing within the constraints imposed by the simplicity of units and links (e.g. Dyer 1991). As McCarthy (1988) has observed, the representational power of connectionist models is generally restricted to unary (i.e. one-place) predicates applied to a single fixed object. An adequate model of human thinking, however, requires representations with at least the logical power of the first-order polyadic predicate calculus (Stenning & Oaksford, 1993): That is, it must be able to express relations among multiple objects that fill particular roles associated with the arguments of predicates (e.g. the "lover" and the "beloved" roles associated with the predicate "love"). Accordingly, a crucial requirement for systematic reasoning is a solution to the "binding problem": the need to keep track of what roles are being played by each constituent. (For example, distinguishing "John loves Mary" from "Mary loves John" requires a way to encode which object fills which argument slot.) Humans can obviously make such distinctions and can code binding information in long-term memory (although it takes longer to recognize previously encountered role bindings than simply to recognize recurrences of objects; Ratcliff & McKoon 1989).

Simple connectionist representations, however, lack constituency relations. Unlike symbolic representations in which links between elements typically define meaningful relationships, the links in connectionist models merely serve to transmit activation between units. As a result, connectionist models do not guarantee systematicity of thinking in principle; in practice, most current connectionist models fail to deal with anything like the systematic knowledge involved in everyday human reasoning. For example, binding information is conspicuously absent in connectionist models such as that used by Rumelhart et al (1986) to represent a "room schema." In their model, a "kitchen," for example, would be represented by a vector of features (i.e. unary predicates) such as "has refrigerator," "has stove," "has sink," and so on. Lacking any capability of expressing multi-place predicates and their role bindings, the model is unable to distinguish a "normal" kitchen from a room with a refrigerator in the sink with a stove piled on top of it.[3] Thus, a key theoretical challenge facing the connectionist approach to meaning is to show how distributed representations of individual concepts could function symbolically as constituents of more complex relation structures. (See Farah & McClelland 1991 for an analysis of neurological data consistent with the possibility that individual concepts have distributed representations in semantic memory.)

The fact that human cognition has both symbolic and subsymbolic aspects encourages various attempts to integrate the approaches. A number of suggestions for hybrid "symbolic-connectionist" models have been offered (e.g. Dyer 1991; Holyoak 1991; Minsky 1991). These models can be divided roughly into two classes. One class of models maintains a core of "traditional" symbolic machinery (e.g. discrete propositions and rules) to represent relation structures, while adding connectionist-style mechanisms for "soft" constraint satisfaction. The second class of models seeks to develop connectionist representations of relation structures by introducing techniques for handling the binding of objects to roles. We review examples of each of these approaches to integrating the two theoretical perspectives.

[3] Connectionist models typically introduce units that respond selectively to combinations (e.g. conjunctions) of inputs, allowing the expression of Boolean operations. Such capability (equivalent to introducing the operators "and," "or," etc) is sometimes characterized as capturing "relational" information (e.g. Estes 1991). However, Boolean operations on a finite set of elements do not suffice to represent relation structures based on multi-place predicates. A typical connectionist model might be able to roughly represent the propositional conjunction "room has sink and room has stove" by including a unit that becomes active just in case both the "room has sink" and "room has stove" units are on. But Boolean operations on propositions do not distinguish "stove is beside sink in room" from "stove is on top of sink in room," "sink is on top of stove in room," and so on. This broader ability to represent argument bindings for multi-place predicates is lacking in typical connectionist representation schemes.

Soft Constraint Satisfaction in Reasoning

The generation and evaluation of beliefs—the central task of induction—has a holistic quality that has posed grave difficulty for theoretical treatments. Tweney (1990) identified the complex interrelatedness of hypotheses as a major challenge for computational theories of scientific reasoning. Fodor (1983) has taken the pessimistic position that little progress is to be expected in understanding central cognition because the facts relevant to any belief cannot be circumscribed (i.e. we do not operate within a closed world) and the degree of confirmation of any hypothesis is sensitive to properties of the whole system of beliefs. As Quine (1961:41) put it, "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body." A psychological theory of induction must identify mechanisms that can cope with the holistic quality of hypothesis evaluation (Holland et al 1986).

One mechanism with the requisite properties is parallel constraint satisfaction, a basic capability of connectionist models. In a connectionist network, local computations involving individual units interact to generate stable global patterns of activity over the entire network. Models that perform "soft" constraint satisfaction over units corresponding to relation structures can attempt to capitalize on the complementary strengths of symbolic representation and connectionist processing. Such symbolic-connectionist models can make inferences based on incomplete information, which standard symbolic systems are often unable to do, using knowledge that distributed connectionist systems cannot readily represent. Models of this sort have been used to account for psychological data concerning text comprehension, analogical reasoning, and evaluation of explanations.

Kintsch (1988) has developed a symbolic-connectionist model to deal with the resolution of ambiguities during text comprehension. His "construction-integration" model has four main components: 1. initial parallel activation of memory concepts corresponding to words in the text, together with formation of propositions by parsing rules; 2. spreading of activation to a small number of close associates of the text concepts; 3. inferring additional propositions by inference rules; and 4. creating excitatory and inhibitory links, with associated weights, between units representing activated concepts and propositions, and allowing the network to settle. The entire process is iterative. A small portion of text is processed, the units active after the settling process are maintained, and then the cycle is repeated with the next portion of text. In addition to accounting for psycholinguistic data on text comprehension, the construction-integration model has been extended to simulate levels of expertise in planning routine computing tasks (Mannes & Kintsch 1991).

Symbolic-connectionist models have been developed to account for two of the basic processes in analogical reasoning—retrieving useful analogs from memory and mapping the elements of a known situation (the source analog) and a new situation (the target analog) to identify useful correspondences. Because analogical mapping requires finding correspondences on the basis of relation structure, most distributed connectionist models lack the requisite representational tools to do it. Purely symbolic models have difficulty avoiding combinatorial explosion when searching for possible analogs in a large memory store and when searching for optimal mappings between two analogs. The two symbolic-connectionist models—the ACME model of Holyoak & Thagard (1989), which does analogical mapping, and the ARCS model of Thagard et al (1990), which does analogical retrieval—operate by taking symbolic, predicate-calculus-style representations of situations as inputs, applying a small set of abstract constraints to build a network of units representing possible mappings between elements of two analogs, and then allowing parallel constraint satisfaction to settle the network into a stable state in which asymptotic activations of units reflect degree of confidence in possible mappings. The constraints on mapping lead to preferences for sets of mapping hypotheses that yield isomorphic correspondences, link similar elements, and map elements of special importance. These same constraints (with differing relative impacts) operate in both the mapping and retrieval models. The mapping model has been applied successfully to model human judgments about complex naturalistic analogies (Spellman & Holyoak 1992) and has been extended to account for data concerning analogical transfer in mathematical problem solving (Holyoak et al, 1993). A similar constraint-satisfaction model has been proposed by Goldstone & Medin (1993) to account for the role of relational correspondences in similarity judgments.

Thagard (1989, 1992) has shown that the problem of evaluating competing explanations can be addressed by a symbolic-connectionist model of explanatory coherence, ECHO. The model takes as inputs symbolic representations of basic explanatory relations between propositions corresponding to data and explanatory hypotheses. The system then builds a constraint network linking units representing the propositions, using a small number of very general constraints that support explanations with greater explanatory breadth (more links to data), greater simplicity (fewer constitutent assumptions), and greater correspondence to analogous explanations of other phenomena. Relations of mutual coherence (modeled by symmetrical excitatory links) hold between hypotheses and the data they explain; relations of competition (inhibitory links) hold between rival hypotheses. Parallel constraint satisfaction settles the network into an asymptotic state in which units representing the most mutually coherent hypotheses and data are active and units representing inconsistent rivals are deactivated. Thagard (1989) showed that ECHO can model a number of realistic cases of explanation evaluation in both scientific and legal contexts; Schank & Ranney (1991, 1992; Ranney, 1993) have used the model to account for students' belief revision in the context of physics problems; and Read & Marcus-Newhall (1993) have applied the model to the evaluation of explanations of everyday events.

The role of constraint satisfaction in human reasoning may explain a set of reasoning and memory phenomena that have sometimes been interpreted as evidence for "mental models" (Johnson-Laird 1983; Johnson-Laird & Byrne 1991). In syllogistic reasoning tasks, people tend to perform poorly when the premises admit of multiple consistent instantiations; and comprehension of described spatial relations is impaired when the description cannot be mapped onto a single determinate array. As Stenning & Oaksford (1993) and Stenning & Oberlander (1992) have noted, connectionist networks have the property of "self-completion": Given a fragmentary input, they naturally settle into a state representing a coherent, unified interpretation of the input. Although such networks may be massively parallel at the level of unit activity, they nonetheless are radically serial at the level of network states. Thus constraint satisfaction is well-suited to reasoning tasks in which a single unified interpretation of the input is both possible and desirable (i.e. the interpretation corresponds to a unique stable state of the network) but badly suited to reasoning tasks in which a single unified interpretation is not possible (as in coding indeterminate spatial descriptions) or not desirable (as in syllogistic tasks for which identifying an acceptable conclusion depends on all possible consistent instantiations of the premises, rather than a single instantiation). We return to the topic of mental models when we discuss "vivid representations" below.

Reflexive Reasoning Using Dynamic Binding

Whereas the models discussed above involve various hybridizations of connectionist processing mechanisms and symbolic representations, a second class of models attempts to provide pure connectionist-style representations of complex relational knowledge. Achieving this goal requires a mechanism for coding bindings of properties and relations to sets of individuals. In contrast to purely symbolic models (e.g. Anderson 1983) in which bindings are represented by unanalyzed elements of notation (e.g. labeled arcs in a semantic network), the connectionist approach represents bindings by more global properties of network states. One general proposal has been to introduce distributed representations in which both the argument slots associated with a predicate, and the objects that fill the slots, are represented as patterns of activity over pools of shared units. For example, Smolensky (1990) developed a representation of argument bindings based on taking the tensor product of appropriate vectors representing the predicate and each of the fillers of its argument slots. Halford et al (1993) have proposed a model based on tensor-product representations to account for constraints on analogical reasoning, as well as for capacity limits on working memory.

A very different approach, based on mechanisms that neurophysiological evidence suggests may play a role in mammalian vision, involves using oscillations of unit firings to represent transient bindings between objects and the argument slots in propositions and rules. A number of researchers have suggested that temporal synchrony can be used to bind features to object representations in visual perception (Hummel & Biederman 1992; von der Malsburg 1981). Shastri & Ajjanagadde (1993) have developed a detailed computational model that uses temporal dynamics to code the relation structure of propositions and rules. Dynamic bindings in working memory are represented by units firing in phase. Consider a proposition such as "John gave the book to Mary." On a single phase, a unit representing the object John will fire in synchrony with a unit representing the "giver" role; in a different phase the unit for Mary will fire in synchrony with a unit for the "recipient" role. The system is object-based, in the sense that each time slice is occupied by the firing of a single active object unit together with units for all the argument roles that the object fills. Bindings are systematically propagated to make inferences by means of links between units for argument slots. For example, in a rule stating that "If someone receives something, then they own it," the "recipient" role in the antecedent of the rule will be connected to the "owner" role in the consequent. Accordingly, if Mary is dynamically bound to the "recipient" role (by phase locking firing of the "Mary" and "recipient" units), then Mary will become bound to the "owner" role as well (i.e. the unit for Mary will fire in phase with units for *both* relevant roles). Shastri & Ajjanagadde show that their model can answer questions based on inference rules in time that is linear with the length of the inference chain but independent of the number of rules in memory—the most efficient performance pattern theoretically possible.

Shastri & Ajjanagadde (1993) note a number of interesting psychological implications of their dynamic binding model. In particular, they distinguish between two forms of reasoning, which they term "reflexive" and "reflective." Reflexive reasoning is based on spontaneous and efficient inferences drawn in the course of everyday understanding, whereas reflective reasoning is the deliberate and effortful deliberation required in conscious planning and problem solving. It is intriguing that humans are far better at text comprehension than, for example, syllogistic reasoning, even though the formal logical complexity of the former task is much greater than that of the latter (Stenning & Oaksford, 1993). In terms of the Shastri & Ajjanagadde model, text comprehension mainly involves reflexive reasoning, whereas syllogistic inference requires reflective reasoning. Fluent comprehension draws upon a rich network of stored rules, which are used in conjunction with the input to establish a coherent, elaborated model of the situation. Reflexive reasoning of the sort involved in ordinary comprehension relies on dynamic binding of objects to argument slots in preexisting rules. These rules have been encoded into long-term memory, with appropriate interconnections between their arguments. In contrast, reflective reasoning requires manipulation of knowledge in the absence of relevant prestored rules. An arbitrary deductive syllogism (e.g. " If all artists are beekeepers, and some beekeepers are chemists, what follows?") is unrelated to any stored rules; rather, understanding the premises requires setting up de novo "rules" (e.g. "If someone is an artist, then that person is a beekeeper") for each problem.

Shastri & Ajjanagadde's model predicts that reflexive reasoning will be constrained by limits on the number of multiply instantiated predicates, as well as by patterns of variable repetition across the arguments of a rule. The model also makes predictions about the limits of the information that can be active simultaneously in working memory. Although the number of active argument units is potentially unlimited, the number of objects that can be reasoned about in a single session is limited to the number of distinct phases available (because only one object unit may fire in a single phase). Given plausible assumptions about the speed of neural activity, this limit on the number of active objects can be calculated as being five or fewer. This figure is strikingly similar to Miller's (1956) estimate of short-term memory capacity and is consistent with work by Halford & Wilson (1980) indicating that adults cannot simultaneously represent relations involving more than four elements. For example, recent empirical evidence (described by Halford et al, 1993) confirms a limit that will be recognized by anyone who has worked with statistical interactions: The most complex statistical relation that people can deal with in working memory is a 3-way interaction (which involves three independent variables and one dependent variable, for a total of four dimensions). Experimental studies of people's memory for bindings between individuals and properties have revealed similar capacity limits, as well as error patterns consistent with distributed representations of bindings (Stenning & Levy 1988; Stenning et al 1988). Recent work has extended the temporal-synchrony approach to other forms of reasoning. Hummel & Holyoak (1992) have shown that the principles embodied in Holyoak & Thagard's (1989) ACME model of analogical mapping can be captured by a model that encodes propositional structure by temporal synchrony.

An interesting feature of the synchrony approach is that the need to minimize "cross talk" between the constituents of relation structures encourages postulating specific types of serial processing at the "micro" level of temporal phases. For example, in the Shastri & Ajjanagadde model only one object is allowed to fire in each time slice. It is also noteworthy that their model combines localist representations of concepts with distributed control, and thus exemplifies a theoretical "middle ground" between traditional production systems and fully distributed connectionist networks. It is possible that attempts to develop connectionist models of symbol systems will cast new light on the limits of parallel information processing. In addition, connectionist models may provide more effective implementations of the flexible recognition processes based on long-term memory that appear crucial to expertise (Chase & Simon 1973). More generally, the confluence of the symbolic and connectionist paradigms seems likely to deepen our understanding of the kinds of computations that constitute human thinking.

IMPLICIT THINKING AND COGNITIVE EVOLUTION

One interpretation of the contrast between connectionist and symbolic approaches to thinking is that human knowledge depends on two distinct cognitive systems. In fact, many theorists of diverse persuasions have been led to propose cognitive dichotomies, which have been given a rather bewildering array of labels: unconscious vs conscious, procedural vs declarative, automatic vs controlled, reflexive vs reflective, and many others. These distinctions do not always divide cognition along the same lines, nor are particular cognitive functions necessarily associated uniquely with particular halves of the dichotomy (Kihlstrom 1987). Nonetheless, there are tantalizing similarities among the proposed dichotomies. In particular, the first member of each pair is generally viewed as involving unconscious mental processes, a topic that has seen a recent resurgence of interest among experimental psychologists (see *American Psychologist* 47(6) for reviews from various perspectives).

We discuss some of the evidence for a gross cognitive dichotomy, which we term (following Reber, 1992) implicit vs explicit cognition. Reber argues that this dichotomy can be understood in terms of the evolutionary constraints that have molded human cognition, a type of argument that has attracted considerable attention lately (Anderson 1990). We therefore also review the broader issues raised by the use of evolutionary and adaptationist arguments in analyses of cognition.

Acquisition of Implicit Knowledge

The fact that at least some of our knowledge is conscious, explicit, and verbalizable is incontrovertible. At the same time, motor skills provide clear examples of knowledge that can be acquired without awareness and maintained in some implicit form that is not readily verbalized (e.g. Pew 1974). The more controversial claim is that some of the knowledge that underlies higher-level thinking tasks is also implicit in much the same ways as are motor skills. Implicit knowledge, as we use the term here, has a number of important characteristics. It is (*a*) knowledge about covariations in the environment, (*b*) learned by exposures to stimuli exhibiting the covariations, (*c*) obtainable without intention or awareness (although in some cases similar knowledge might be obtained explicitly), and (*d*) demonstrated by improved performance on tasks that seem to require thinking (e.g. generalization and prediction); but it is knowledge that does not have a fully explicit representation in that (*e*) it is not fully verbalizable and (*f*) it is not manipulable in the sense that it cannot be re-represented explicitly to serve as input to other procedures.

We begin by considering evidence concerning the acquisition of implicit knowledge—that is, implicit *learning*. Implicit learning may well be related to implicit memory, a topic that has received much attention in recent years (see Richardson-Klavehn & Bjork 1988; Roediger 1990; and Schacter et al 1993 for reviews). Work on implicit memory, however, typically focuses on the effect of specific events on subsequent task performance, whereas work on implicit learning focuses on the cumulative impact of multiple events involving different (although related) stimuli, emphasizing the acquisition of knowledge about overall regularities in the stimuli rather than about the details of single learning events. Implicit learning has been demonstrated in the laboratory using many different techniques which, until recently, have been explored in relative isolation. Seger (1992) has provided an integrated review of three major methodologies: artificial-grammar learning, learning to control the behavior of dynamic systems, and sequence-learning tasks.

In the typical artificial-grammar learning procedure (see Reber 1989 for a review of the extensive work from his laboratory, which began with the seminal study of Reber 1967), the experimenter constructs a finite-state grammar that generates "grammatical" letter-strings. In the learning phase, subjects are exposed to some of those strings; they may be told to observe or memorize the strings or, in an intentional learning task, to observe and try to figure out the rules that govern the regularities in the strings. In the test phase subjects are typically asked to make "grammaticality" judgments for both old and novel strings. They are told that the strings they have seen were generated according to rules; they are then asked to judge whether or not various test strings follow those rules. In addition, they are often asked to verbalize their knowledge about the grammar and explain how they made their grammaticality judgments. Two consistent results have emerged from these studies: 1. subjects can usually distinguish grammatical from nongrammatical letter strings at an above-chance although far-from-perfect level, and 2. subjects cannot fully articulate the rules they are using to make those judgments.

In dynamic systems tasks, subjects are asked to learn to control the output of a rule-governed system by manipulating input into the system. For example, subjects may be asked to try to control the output of a simulated sugar production factory by typing in the number of workers the company should employ to reach a specified level of production. On each trial the subject types in a number and then is told how much sugar will be produced. The rules underlying the system are linear equations (sometimes with a random error factor added) and always depend on either the previous or current input or output. Berry & Broadbent (1984) used rules of the form: output = 2 × current input − previous output + error. In that study, subjects were able to learn to control the system but were not able to verbalize how they did so. Berry & Broadbent (1988) and Hayes & Broadbent (1988) compared performance on a task with one of two underlying rules—a "salient" rule (output = current input − 2 + error) and a "nonsalient" rule (output = previous input − 2 + error). Under standard learning conditions, subjects receiving the first rule were better able to control the system, were more likely to be able to state the relationship between the variables of the task in protocols, and were more accurate on a questionnaire asking what they could do to control the system given specific circumstances.

Dynamic systems tasks and artificial grammar tasks are similar in that (*a*) subjects seem to acquire knowledge through exposure to repeated exemplars, and (*b*) subjects' performance exceeds their ability to verbalize their knowledge. The tasks differ, however, in that subjects in the dynamic systems tasks are explicitly trying to achieve an objective goal (to maintain production at a specific level). Thus it seems likely that all subjects in the dynamic systems task, unlike subjects in the artificial-grammar task, are intentionally seeking the systems' underlying rules.

In sequence-learning tasks, subjects' implicit knowledge is usually demonstrated by a decrease in reaction time to events generated by underlying rules relative to events generated randomly—without the subject being able to articulate the rule or make explicit predictions about subsequent events (e.g. Cohen et al 1990; Stadler 1989; Willingham et al 1989; but Kushner et al 1991; see Seger 1992). In a study by Nissen & Bullemer (1987), for example, subjects were to press a corresponding button after each in a series of flashes of light. Subjects showed a decrease in reaction time to push buttons when the pattern of flashes was repeated relative to random light sequences. These results are similar to those obtained in the Hebb (1961) digits task, in which subjects echo strings of digits, some of which are repeated. Subjects make fewer errors on repeated strings than on other strings, a result that has been shown to hold regardless of subjects' level of awareness of the repetition (McKelvie 1987). In other sequence-learning tasks (e.g. Lewicki et al 1988), subjects show a decrease in reaction time when responding to targets whose position is predictable from the positions of items in earlier trials. When questioned, however, subjects report no knowledge of the sequences underlying the task.

Evidence of implicit learning also comes from studies of "intuitive physics." As they interact with the physical world, humans acquire knowledge about the complex rule-governed behavior of moving objects. (See Wellman & Gelman 1992 for a recent review of developmental aspects.) People often make systematic errors in predicting (by verbalizing or drawing) the future motion of physical objects as they exit from curvilinear tubes, are released from strings constraining their motion, or are dropped from moving carriers. Yet implicit knowledge of physical rules is demonstrated when subjects view contrived videotapes showing what objects would look like following the paths that subjects predict (Kaiser et al 1985; Shanon 1976; but see McCloskey & Kohl 1983). Even 6-year-old subjects recognize that these trajectories "look wrong" (Kaiser & Proffitt 1984). Such recognition only occurs, however, for the physics of objects that behave like point masses (such as the examples described above); for more complex motion involving multidimensional relationships (e.g. rotating objects), subjects have great difficulty in distinguishing possible from impossible motions while viewing simulations of ongoing events (Proffitt et al 1990). In addition, explicit knowledge (such as that produced by training in physics) often does not increase the accuracy of predictions (e.g. Proffitt et al 1990).

The fact that people apparently acquire implicit knowledge of only certain types of physical regularities is consistent with other evidence indicating that there are limits or constraints on which covariations can be learned implicitly (Seger 1992). A variety of factors apparently contributes to those limitations, of which two major classes are (*a*) the "simplicity" of the rules underlying the covariations, and (*b*) biases—both innate and learned—that favor learning certain kinds of covariation.

Some rules may be too difficult to learn implicitly. There may be limits on our ability to detect covariations over spatial and temporal distances (e.g. Broadbent et al 1986; Cleeremans & McClelland 1991), or between a large number of interacting variables (e.g. Proffitt et al 1990). Other rules may be, in a sense, "too easy" to learn implicitly. If a covariation is sought and discovered consciously, then it becomes explicit: Performance is more accurate, the rule is more likely to be verbalizable, and there is no dissociation between verbalization and performance. Explicit learning can be encouraged by instructions to search for rules underlying the covariations in the exemplars; however, such instructions only seem to improve performance when the rules are, in fact, discoverable. Reber et al (1980) found that instructions interacted with the way the learning exemplars were presented. When the exemplar presentation made the underlying rules more obvious, subjects instructed to search for rules performed better than those not so instructed (also Servan-Schreiber & Anderson 1990). In contrast, when the presentation of exemplars was unstructured, subjects who were told to search for rules performed worse than those given more implicit instructions, because the rule-seeking subjects induced nonrepresentative rules. Similarly, Berry & Broadbent (1988) found that intentional instructions helped subjects' performance in the easy "salient" condition but impaired the performance of those in the "nonsalient" condition.

Humans seem to be biased to learn certain types of regularities more readily than others. In category learning, for example, subjects more easily learn categories with unimodal than with bimodal distributions (Flannagan et al 1986). In cue-probability learning, subjects perform better with linear than with nonmonotonic functions; for psychophysical functions, humans find it easiest to learn functions that are linear in logarithmic space (Koh & Meyer 1991). When exposed to stimuli that exhibit nonpreferred regularities, people's judgments early in learning are biased toward the preferred form of regularity; with additional exposure, however, people eventually learn the actual pattern, overcoming their entering bias.

In tasks involving sequence learning, the early items in a series appear to be especially significant, perhaps because they serve to mark the beginning of a new pattern. In the Hebb digits task, changing the first two digits (but not just the first one, or just the last one or two) eliminates learning (Schwartz & Bryden 1971). Servan-Schreiber & Anderson (1990) found that subjects who memorized three-chunk letter strings by building a larger chunk out of the first two smaller chunks were more accurate at grammaticality judgments than subjects who memorized the strings by chunking the second two smaller chunks.

In addition, implicit learning is sometimes affected by the semantic content of the stimuli. The same underlying rule may be learned better when presented with some stimuli than with others. Stanley et al found that subjects performed better in a dynamic systems task when the cover story involved controlling the friendliness of a person rather than the sugar production of a factory, even though the underlying rule was the same in both cases (also Berry & Broadbent 1984). Such content effects suggest that prior knowledge can modulate implicit learning, perhaps by biasing subjects to attend to particular features or to expect particular types of regularities. (We discuss other content effects observed in thinking tasks in a later section.)

Access and Use of Implicit Knowledge

An important characteristic of implicit knowledge, demonstrated in several of the learning studies reviewed above, is that subjects cannot verbalize all of their knowledge. It is clear, however, that at least some useful knowledge is often verbalizable. In an artificial-grammar learning task, Mathews et al (1989) told some subjects that after each test trial they should explain to an "unseen partner" how they made the grammaticality decision. These explanations were later played to another group of subjects as the latter made grammaticality judgments without any prior training or feedback. Such yoked subjects performed better than chance, but not as well as the original subjects, suggesting that some but not all knowledge is verbalizable. Stanley et al (1989) obtained similar results for yoked subjects in a dynamic systems task. Note, however, that the lower performance of the yoked subjects might be attributable either to (*a*) the inability of the original subjects to articulate all of their knowledge, or to (*b*) the inability of the yoked subjects to implement the transmitted knowledge successfully.

In general, it would be difficult to show conclusively that knowledge is inaccessible to consciousness, because the methodology for assessing access to implicit knowledge is inevitably open to challenge. Various methods that do not rely on verbalization of rules have been used to elicit subjects' implicit knowledge; often these methods reveal accessible covariation-based or fragmentary knowledge that can account for much or all of the subjects' performance. In artificial-grammar tasks, classification performance can be accounted for by: subjects' ability to indicate grammatical or ungrammatical parts of letter strings (Dulany et al 1984); knowledge of bigrams or trigams (Perruchet & Pacteau 1990, 1991; Perruchet et al 1992); knowledge of chunks (Servan-Schreiber & Anderson 1990); or knowledge of sequential letter dependencies—i.e. the ability to decide, when presented a string of letters, whether each letter in the grammar, if presented next, would create a grammatical string (Dienes et al 1991). Similar arguments have been levied against the sequence learning tasks (Perruchet et al 1990). Most recently, Perruchet & Amorim (1992) showed that in a sequence-learning task, subjects' conscious ability explicitly to generate and recognize parts of the sequences paralleled their improvement on the reaction-time task. One critique that may be leveled against some methods used to elicit implicit knowledge is that the procedure itself may change the representation of the knowledge from implicit to explicit (Reber et al 1985); certainly knowledge that is initially implicit may eventually be re-represented in some more explicit form (Karmiloff-Smith 1990; also Berry & Broadbent 1988). On the other hand, these studies offer insight into two important and unresolved questions: (*a*) How unconscious is the learning and (*b*) is what is being learned abstract rules or something less complex?

Nonetheless, evidence still suggests that subjects' ability to consciously access implicit knowledge typically lags behind their ability to use it. Rubin et al (1993) had subjects study and recall a series of five highly similar ballads, and then attempt to compose a ballad of their own. These subjects were also asked to write down the rules and generalizations that characterized the ballads they had studied. The ballads by the subjects followed more than half of the objective regularities in the studied ballads, but the subjects could only state about one quarter of these rules. Moreover, the correlation between the implicit regularities observed in the composed ballads and the explicitly stated rules was low and statistically nonsignificant. Further support for a dissociation between implicit and explicit knowledge comes from studies of people solving various types of "insight" problems, which reveal that subjects often reach correct conclusions even though they cither fail to report they are nearing a solution (Metcalfe 1986; Metcalfe & Wiebe 1987) or are unable to verbalize why their conclusion is correct and lack confidence in it (Bowers et al 1990).

Another important issue relevant to accessing and using implicit knowledge concerns the range of related cases to which such knowledge can be applied. Although it is typical to find generalization to new cases drawn from the same basic pool as those used during training, more distant transfer is not readily obtained. Berry & Broadbent (1988) trained subjects on one dynamic systems task and then measured their performance on a second task involving the same underlying rule. The semantic cover story of the transfer task was either superficially similar or dissimilar to that of the learning task. Subjects in the superficially similar condition improved their performance as much as control subjects (who continued to perform the same task); however, subjects in the superficially dissimilar condition showed no such transfer. Furthermore, subjects who were given a hint that the underlying equation in the transfer task was the same as that in the learning task were not helped: In fact, subjects in the similar condition performed *worse* when given such a hint. By contrast, such a hint generally aids transfer in explicit tasks involving analogical transfer across semantic contexts; in the latter case, subjects seem only to need reminding of knowledge available to them (e.g. Gick & Holyoak 1980). Lack of transfer has also been demonstrated for intuitive physics problems: Subjects are accurate when making predictions about familiar problems but do not transfer their knowledge to make correct predictions on an unfamiliar problem with the same underlying structure (Kaiser et al 1986).

There is, however, evidence of remote transfer in studies using the artificial-grammar task: Subjects who are trained on an artificial grammar using one set of letters perform well on test items generated by structurally isomorphic grammars based on new letters (Mathews et al 1989; Reber 1969). Reber (1989) and Mathews (1990) have argued, on the basis of such evidence for remote transfer, that implicit knowledge is abstract in the sense that the person has learned rules about the structure of the stimuli independent of its physical instantiation. It is possible, however, that such transfer effects are due to shared relational features that were associated with grammaticality during learning of the initial grammar (e.g. a run of three "same" letters near the middle of the string, or the presence of certain "fragments" as discussed above, might indicate grammaticality; see Seger 1992). Although such relational features are arguably abstract to some degree (Mathews 1990), learning covariations based on such features might not indicate general sensitivity to relational structure. It is also possible that transfer between isomorphic grammars depends at least in part on analogical reasoning between studied exemplars and transfer items (Brooks & Vokey 1991), or possibly between studied chunks and transfer items. However, exemplars need not be explicitly remembered. Knowlton et al (1992) found that amnesic patients were able to classify letter strings according to the rules of an artificial grammar as well as control subjects, even though the patients' ability to recognize the studied exemplars was impaired.

The transfer issue may well prove central in assessing potential models of implicit learning. In particular, subsymbolic models based on classifier systems (Druhan & Mathews 1989) and on connectionist learning algorithms (Dienes 1992; Kushner et al 1991) have been successful in accounting for many aspects of human performance in learning artificial grammars. The subsymbolic models readily account for the difficulty that people have in fully verbalizing their knowledge of stimulus regularities, which in the models is largely contained in low-level patterns (of strengths of classifier rules, or weights on links in a distributed connectionist network). However, none of these models can account for transfer between isomorphic grammars based on entirely different sets of letters. If such transfer could be explained on the basis of a limited set of relational features, or by additional explicit processes operating in the transfer task, then it might be possible to characterize the limits of implicit learning in terms of the capabilities of current subsymbolic models; if not, more sophisticated learning models will be required. The limits of transfer based on implicit learning clearly warrant further investigation.

The kind of covariational information that can be learned implicitly appears to be statistical in nature. The question then arises as to the conditions under which such information will be accessed and used in tasks that require making intuitive predictions. The classic work of Kahneman & Tversky (see Kahneman et al 1982) produced many compelling illustrations of people's failures to use statistical information of the sort that could plausibly be acquired by implicit learning. In particular, people commonly underutilize base rates and sample frequencies in making judgments about the likelihood of individual events. Hasher & Zacks (1984) argued that the encoding of event frequency is based on an automatic or implicit process that takes place largely without awareness. Although there has been some dispute about the extent to which frequency encoding satisfies various proposed criteria for automaticity, people are generally accurate in picking up such information (e.g. Sanders et al 1987). The evidence for implicit frequency encoding leads to the following puzzle: If the encoding of frequency is automatic, then one would expect the encoding of base rates, which are simply relative frequencies, to be automatic also. Why, then, have countless studies shown that subjects neglect base-rate information when making various inferences?

In fact, the paradox may be more apparent than real. In almost all studies showing base-rate neglect, subjects are provided with summary information about base rates, rather than with an opportunity to learn information about each individual event comprising the set of events. The base-rate information is thus presented explicitly; no implicit learning occurs. Generally, in experiments in which base-rate information is derived from real-life experience (Christensen-Szalanski & Bushyhead 1981), or learned from presentation of exemplars (Manis et al 1980), subjects use that information effectively. The procedure of giving summary information can be contrasted with the typical category-learning experiment in which subjects are presented with individual exemplars of the categories, are asked to make category judgments, and then are given feedback. When subjects learn to predict membership in categories that occur with different frequencies, they learn to use base rates accurately during the study trials (Estes et al 1989; Gluck & Bower 1988; Medin & Edelson 1988). However, subsequent transfer trials, in which subjects are asked either to indicate category membership or to estimate the probability that a category was correct given a cue, often reveal some apparent misuse of base-rate information.

The findings from category learning experiments suggest that base-rate use has two components: acquisition, which might be done implicitly and is quite accurate (perhaps based on learning feature-to-category conditional probabilities); and access, which (depending on the type of test) may well be explicit and under more conscious control. When acquisition and test both tap implicit knowledge (e.g. during learning trials), subjects generally use base rates accurately; however, it seems that when engaged in more explicit tasks, subjects must be "reminded" to use base rates. In such tasks, people tend to focus on the strength or extremity of the individuating evidence about the case, with insufficient regard for its weight or credibility; base rates and sample size are special cases of the latter type of information (Griffin & Tversky 1992). Explicitly presented base rates have greater impact when they have strong causal implications (Ajzen 1977; Bar-Hillel & Fischhoff 1981; Tversky & Kahneman 1980), when people bring a scientific rather than a clinical orientation to a problem (Zukier & Pepitone 1984), when subjects watch a random sampling process or operate in a domain in which revision of base-rate information is common (Gigerenzer et al 1988), or when conversational context suggests that base-rate information is more relevant than individuating information (Krosnick et al 1990; Schwarz et al 1991). The situations in which subjects tend to use base-rate information are similar to those in which subjects are more likely to invoke other elements of appropriate statistical reasoning (Nisbett et al 1983). We discuss further examples of contextual variations in the use of reasoning strategies in a later section.

In the absence of cues of the above types, statistical knowledge is more likely to be evoked when subjects make judgments about an aggregated set of cases rather than individual cases (Tversky & Kahneman 1983). In particular, people are typically overconfident in evaluating the accuracy of their own beliefs taken one at a time, yet quite well calibrated in judging their overall accuracy for a set of beliefs (Gigerenzer et al 1991; Griffin & Tversky 1992). Without some "reminder" cue, people tend to base their decisions on their assessment of individual events, rather than estimates about a population of similar events (Kahneman & Lovallo, 1993; Tversky & Kahneman 1982). It follows that people are often overconfident in their decisions even though at some level they "know better." Such examples support the general possibility of dissociation between the acquisition and use of implicit knowledge.

Evolutionary History and Adaptation

Another theme in recent work on thinking, closely related to evidence for implicit cognition, concerns the use of evolutionary arguments to support cognitive analyses. The evolution of human cognition is, of course, largely a matter of conjecture. As Lewontin (1990) wrote, in a cautionary introduction to the topic, "If it were our purpose ... to say what is actually known about the evolution of human cognition, we could stop at the end of this sentence" (1990:229). Lewontin pointed out that different types of evolutionary arguments must be distinguished. Here we focus on two types: those based on the evolutionary connections among species, and those based on the adaptive significance of cognitive characteristics. Of these, the former has the virtue of being more closely tied to observable evidence.

There are many reasons to suspect that implicit cognition is phylogenetically older than the explicit variety (Rozin 1976; Sherry & Schacter 1987)—certainly, basic covariation detection is within the cognitive capacity of many species, whereas the writing of review articles is practiced, as far as we know, by *Homo sapiens* alone. Based on the assumption that implicit cognition evolved long before high-level consciousness, Reber (1992) argues that certain principles of evolutionary biology—von Baer's pre-Darwinian laws of embryological development and Wimsatt's (1986) "developmental lock" model of evolutionary change—can be used to derive a number of predictions about the general character of implicit cognition. The major predictions Reber derives are that: 1. implicit systems should be more *robust* than explicit systems, operating despite injuries, diseases and other disorders; 2. implicit processes should be more *age independent*, revealing fewer differences than explicit processes in both infancy and old age; 3. implicit processes should be *IQ independent*; 4. more generally, implicit processes should show *lower population variance* than explicit processes; and 5. implicit processes should show *across-species commonalities*. Reber cites empirical evidence in support of each of these predicted characteristics of implicit cognition.

In addition to deriving predictions based on evolutionary history, Reber (1992) and others have proposed reasons it would be adaptive to have two basic cognitive systems—one to passively pick up covariations among significant environmental stimuli, and another to selectively integrate and control information from many different sources. Others, most notably Anderson (1990), have made much more general appeals to arguments based on adaptation. Anderson argues that psychologists can best develop theories of human cognition by making the assumption that human cognition is optimally adapted to the environment. He terms his research program "rational analysis," taking care to distinguish two meanings of the term "rational": 1. "the normative sense, as used in philosophy, in which human behavior is matched against some model that is supposed to represent sound reasoning ... [and 2.] the adaptive sense, as used in economics, in which the behavior is said to be optimized with respect to achieving some evolutionarily relevant goals" (1990:250–51). It is in this latter sense that Anderson argues human behavior is rational.

Developing a theory using Anderson's program of rational analysis involves six steps (Anderson 1990:29): 1. Specify the goals of the cognitive system; that is, what the system is trying to optimize. 2. Develop a formal model of the environment to which the system is adapted; that is, the environment in which the cognitive system evolved. 3. Make minimal assumptions about computational limits and abilities of the system, including the costs incurred in achieving optimal performance. 4. Derive what the optimal behavioral function should be, given Steps 1–3. 5. Examine the empirical literature to see whether the predictions of the behavioral functions are confirmed. 6. If the predictions are off, try reexamining and revising the assumptions in Steps 1–3.

There is considerable potential for slipperiness in executing this program, especially in Steps 1–3, as has been noted both by Anderson himself (Anderson 1990:30) and others (see commentary on Anderson 1991a in *Behavioral* *and Brain Sciences,* Vol. 14; also Lewontin 1990). Step 3 is especially problematic because we know little about the environment(s) in which cognition evolved. Thus if the predictions of rational analysis fail, it may be all too tempting to go back and redefine the environment. It is always possible to invent an environment in which a behavior would be adaptive (Dawkins 1987; Simon 1991).

Anderson (1990) demonstrates the use of his method in analyzing memory (also Anderson & Milson 1989), categorization (also Anderson 1991b), causal inference, and problem solving. In practice, his rational approach involves Bayesian analysis, making various simplifying assumptions to avoid the computational intractibility of unconstrained Bayesian inference (see Pearl 1988 for an artificial-intelligence approach along broadly similar lines). With some minor tinkering, Anderson's analysis of memory predicts many major memory findings; the optimized categorization model also predicts a large number of laboratory phenomena. In both cases Anderson's models are nearly equivalent to more mechanistic models that had been proposed previously, and inherits their weaknesses as well as their strengths (e.g. his categorization model, like other similarity-based models, does not provide any constraints on which features are used to represent objects, nor does it account for theory-based influences on categorization; see Medin 1989).

Anderson's program has so far been less successful in dealing with other topics. In particular, rational analysis provides few insights into problem solving, as Anderson admits. He believes that this failing "is more a comment on the state of the literature on problem solving than on the theory" (1990:229). As Anderson points out, most studies of problem solving have been concerned with games and puzzles, which have little adaptive value when compared to the sorts of problems that people confront in real life. More naturalistic problems might include choosing a birthday card, getting the car fixed, or deciding whether to buy an expensive zoom lens. (Of course, none of the latter examples is any more likely than are puzzle problems to have exerted great selection pressure during human evolution.) There is reason to suspect, however, that the limitations of rational analysis in dealing with general problem solving arise for a more basic reason. Problem-solving performance reveals much more pronounced individual differences (often due to the relative efficacy of different strategies) than do tasks that more specifically tap memory retrieval, categorization, or causal inference. In fact, problem-solving skill seems to meet none of Reber's (1992) predicted criteria for implicit knowledge—it is not robust, and it is quite variable across age, IQ, and within as well as across species. The other cognitive functions that Anderson analyzes, which at least in primitive forms are doubtlessly phylogenetically older than general problem solving, seem to have a more implicit than explicit character (although Anderson does not address the implicit/explicit distinction). It is possible that implicit systems, which have undergone longer evolutionary refinement, are in general better optimized to the environment than is explicit cognition and therefore more amenable to rational analysis (S. Kosslyn and S. J. Gould, personal communication). Indeed, given the greater inherent computational complexity of the functions that explicit cognition attempts to compute (which are almost certainly intractible in the general case), it is unlikely that even a few more billion years of evolution will make much difference.

What Is Explicit Thinking For?

In many ways we have painted a more attractive picture of implicit thinking than of its younger explicit sibling. Implicit knowledge is acquired with little effort and is often accurate, perhaps even optimized in some sense. Explicit knowledge, by contrast, takes hard mental work to achieve and might seem barely worth the effort—it often flouts the dictates of rationality, even wantonly ignoring the accurate statistical knowledge that the implicit system has patiently accumulated. What, then, is explicit thinking for?

One answer (at the risk of seeming facetious) is that it enables us to draw a picture of a person with two heads. If the answer seems absurd, consider: For how long would the implicit system have to absorb covariations passively before its knowledge would enable us to draw such a picture? Longer than a lifetime, no doubt. Drawing a two-headed person requires more than the simple reproduction of patterns observed in the environment; it requires the creation of something we have never seen. We may use our experience, however limited, with drawing other objects, such as people with the more typical quota of heads; but to make the required transformation, the usual procedure for drawing a person must itself be manipulated in the process. And to be manipulated in this way, the procedure must be represented explicitly.

The role of internal representations in creative drawing has been explored by Karmiloff-Smith (1990), from whom we have borrowed our example. As she points out (also Rutkowska 1987), a procedure can have two functions: 1. it can be activated to generate an output; and 2. it can itself be manipulated or reorganized by other procedures. In terms of the implicit vs explicit distinction, only a procedure that has an explicit representation can serve as "data" for another procedure, and hence be transformed. According to Karmiloff-Smith (1986, 1990), it is characteristic of cognitive development that procedures initially represented implicitly, such that they can accomplish routine tasks, must be re-represented at a more explicit level before they can be manipulated to accomplish novel tasks.

Karmiloff-Smith (1990) found evidence of a developmental progression in children's abilities to draw such novel objects as a two-headed man—evidence that supports her theoretical claims. When given this task, 5-year-old children were likely to draw a normal man first, then add a second head with an entire second body attached. By age 8–10, however, children had acquired the ability to systematically interrupt their routine drawing procedure, insert two heads symmetrically tilted away from the upright, and then continue drawing the rest of the man. Whereas for the younger children the drawing procedure seemed to be represented as a fixed sequence, for the older children it seemed to have a more abstract part-whole structure, allowing the routine to be interrupted and novel parts to be inserted.

The re-representation of procedures from an implicit to an explicit form is the opposite of the progression assumed in Anderson's (1983, 1987) conception of knowledge compilation, a process that transforms explicit declarative knowledge into an implicit procedural representation. We return to this contrast below when we discuss expertise and knowledge transfer. For now, we note that Karmiloff-Smith's concept of explicit representation appears to imply a capacity for systematic manipulation of knowledge of much the same sort as we discussed earlier when we considered the relationship between connectionist and symbolic representations.

A more general answer to our opening question, then, is that explicit thinking is required for some important forms of creative thought (Boden 1990). (This is not to deny, however, that implicit processes may also play important roles in creativity.) Such a function would suggest that the explicit system ought *not* to be adapted to the environment in which it evolved—at least not in the sense of Anderson's (1990) rational analysis. Let the implicit system become adapted to the environment; the explicit system can help us adapt the environment to us. Explicit representations of knowledge allow us to imagine what is not the case, but might be, and how we might make it so.

CONTENT IN THINKING, THINKING IN CONTEXT

Our third theme is an extension of a topic we encountered in the previous section: the role played in thinking by prior knowledge and contextual cues. Across a wide variety of tasks, the manner in which individuals reason and solve problems is intimately related to the content of what is being thought about as well as to the context in which the thinking takes place. Indeed, content and context are themselves interwined, since the context of thinking—for example, the actions of other individuals and the goals of the reasoner—directly influences the content of thought. Conversely, the content of thought—for example, the internal representation of a problem situation—may trigger goals that alter the effective context in which the thinker is operating. The implications of content and context effects are a focus of current theoretical debates.

Thinking vs Theorem Proving

Cognitive scientists all agree that thinking is properly construed as computation. At the same time, the most vigorous debates in the field concern the questions of what kinds of computation underlie human thinking, and what kinds of representations are used. Various theorists have championed representations for reasoning and problem solving based on stored cases, schemas, rules of varying degrees of generality, constructed semantic models of individuals, and quasi-spatial or image-like structures. The "classical" view has been that thinking is much like the conventional procedures for proving theorems in a formal system, such as a logic. The derivation of inferences, in this proof-theoretic view, depends on the serial application of exceptionless formal rules of inference to internal "statements" expressed in a "language of thought" (Fodor 1983). The classical account implies that reasoning depends on rules that manipulate internal representations on the basis of their syntactic form, rather than their semantic content.[4] Fodor & Pylyshyn have emphasized the centrality of the proof-theoretic approach to cognitive science: "It would not be unreasonable to describe Classical Cognitive Science as an extended attempt to apply the methods of proof theory to the modelling of thought" (1988:29-30).

It is now clear that the conventional proof-theoretic approach is seriously limited in its ability to account for human thinking (Oaksford et al 1990). Two empirical problems present themselves. First, everyday commonsense reasoning is based on defeasible (i.e. "defeatable") inferences, such that conclusions derived from premises can be overturned by subsequent information. For example, suppose we have a rule-like belief that Tom leaves for work in his car every morning at 8 AM. If we see Tom pulling out of the driveway at 8 AM sharp, we will probably conclude that he is going to work. However, we are likely to retract this conclusion if we later remember that it is actually a holiday or if we hear that Tom's wife delivered a baby at the local hospital at 9:00 AM, and so on. The list of exceptions to most rule-like regularities is indefinitely long, thwarting attempts to code the exceptions exhaustively into rules. Thus although there have been many attempts to formalize commonsense reasoning in versions of nonmonotonic logics (i.e. logics in which inferences may be subtracted, as well as added, with the addition of new premises), these efforts have met with limited success (Minsky 1991; Oaksford et al 1990).

The second problem with the proof-theoretic approach is that everyday reasoning is highly content dependent. As we observed in our earlier discussion of connectionist approaches to thinking, people are often poor at solving deductive problems based on arbitrary content (like syllogisms), yet they easily use knowledge stored in long-term memory to make inferences that

4 The terms "formal" and "syntactic" have multiple usages in cognitive science. As has often been noted (e.g. Rips 1990), any computational system is necessarily "formal" and "syntactic" in the sense that the procedures are specified in terms of the form of the representations over which they operate. In principle, any aspect of semantic or pragmatic content could be used by a theorist to define the forms that determine the range of applicability for inferential procedures. Accordingly, any computational model of reasoning (including, if they are made explicit, those based on pragmatic schemas and mental models; see below) is necessarily formal in this basic sense. This point has often been misunderstood (see Stenning 1992). Those psychological theories that we refer to as based on "formal rules," however, place strong constraints on what aspects of meaning are required to define the form of representation. Such theories postulate that only the meanings of the "logical terms" included in logics that have been proposed by logicians (e.g. connectives, quantifiers, and modals) are required to define the forms to which inference rules apply.

contribute to a coherent interpretation of a situation. That is, 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.

The defeasibility of commonsense inferences is closely connected to their content specificity. For example, studies of reasoning about propositional connectives, such as "if," reveal that the content of premises can affect the conclusions drawn. For example, Byrne (1989) gave one group of subjects premises such as the following:

1. If she has an essay to write then she will study late in the library.
2. She has an essay to write.

Almost all subjects (96%) in this group drew the conclusion supported by the formal inference rule *modus ponens,* namely

3. She will study late in the library.

Another group of subjects also received premises 1 and 2 but in addition received

1′. If the library stays open then she will study late in the library.

Only 38% of the latter subjects drew conclusion 3, indicating that introduction of premise 1′ blocked application of *modus ponens* based on premises 1 and 2.

Although Byrne (1989; Johnson-Laird & Byrne 1991) interpreted her results as indicating that people do not "follow rules" such as *modus ponens* when reasoning, this conclusion is overstated (Politzer & Braine 1991). The most obvious interpretation of Byrne's results, as she herself noted, is simply that the addition of premise 1′ causes subjects to implicitly alter their interpretation of premise 1, reinterpreting the antecedent as a conjunction of two clauses, "she has an essay to write" and "the library is open." Once the premise is tacitly altered, *modus ponens* does not apply (cf Henle 1962). Other results indicate that people are reluctant to apply *modus ponens* to conditional statements when the antecedent is interpreted as a probabilistic cause, such as, "If a person smokes, then that person will get lung cancer" (Cummins et al 1991). Cheng & Nisbett (1993) argue that causal regularities of the form "If *cause* then *effect*" (a major subclass of conditionals) are treated as expressions of probabilistic contingencies. It seems that people generally treat conditional rules not as deterministic and inviolate but as expressions of "default" regularities, assumed to hold unless overridden by other information (cf Oaksford & Chater, 1993). Certain general inference rules, such as *modus ponens*, may, in fact, be used in reasoning, as some theorists argue (e.g. Braine & O'Brien 1991; Rips 1989; Smith et al 1992); however, if the content triggers knowledge that overrides the stated premises, application of the inference rule may be blocked. The theoretical challenge is to provide a model that accounts for people's facility in making plausible but defeasible inferences (e.g. Osherson et al 1990). It seems unlikely that a successful model of everyday reasoning will resemble standard methods for constructing proofs using exceptionless formal rules. [See Stenning & Oaksford (1993) for a discussion of the relationship between logic and reasoning.]

Relevance and Pragmatic Reasoning

A crucial question for theories of thinking concerns *relevance*: How do people access and exploit knowledge relevant to their goals when drawing inferences, making decisions, or solving problems? The problem of determining relevance emerges in many guises. In the area of deductive reasoning, psychological theories based entirely on formal logics have been unable to explain how everyday inference is constrained by intuitions about the relevance of premises to conclusions. For example, an apparent constraint on the use of the English connective "if" is that the antecedent should be relevant to the consequent. But when "if" is interpreted in terms of the material conditional in propositional logic, the sentence "If the moon is made of green cheese, then 13 is a prime number" is considered to be true (because any conditional with either a false antecedent or a true consequent is true). The "schema for Conditional Proof" that Braine & O'Brien (1991) adapt from natural-deduction logic to form a core component of their psychological theory of "if" implies that any sentence of the form "If p then q" is true whenever q is already known to be true, even when p is irrelevant to q. Often, however, people find such irrelevant conditionals peculiar. And even premises that are relevant to making a deductive inference may fail to be relevant to explanation, as an example from Hempel (1965) demonstrates. Suppose we are given the premises

> All members of the Greenbury School Board are bald.
> Horace is a member of the Greenbury School Board.

We can readily deduce that Horace is bald. However, if asked to explain *why* Horace is bald, we are unlikely to say it is because he is a member of the school board. So far, no formal theory has satisfactorily solved the problem of defining relevance.

People's inferential procedures are also influenced by their current beliefs about the content of the premises. In deductive inference tasks, for example, reasoners are more likely to accept an invalid conclusion that is consistent with their beliefs than an invalid conclusion that they do not believe is true (and hence are motivated to refute) (Evans et al 1983; Markovits & Nantel 1989; Oakhill et al 1989). In hypothesis testing, the reasoner's current hypothesis will guide selection of cases chosen to be examined. People have a strong preference for "positive" tests—that is, for examining cases in which either the hypothesized condition for the target outcome, or the target outcome itself, is known to hold. As Klayman & Ha (1987, 1989) have argued, positive testing need not indicate an irrational bias toward confirmation of one's hypotheses. In many realistic cases, positive testing actually maximizes the possibility that the tested hypothesis will be disconfirmed. More generally, focus on confirmation vs disconfirmation may vary dynamically as the reasoner collects evidence over time. Successful hypothesis testing, in studies both of scientific inference and of medical diagnosis, often involves an initial focus on confirmation followed by more critical examination of "loose ends" or apparent anomalies, which may lead to hypothesis revision (Dunbar, 1993; Mynatt et al 1978; Patel & Groen 1991). In cases of "pathological science," however, scientists with strong attachment to their hypotheses may actively avoid collecting or recognizing disconfirming evidence (Rousseau 1992). In a review of evidence for motivated reasoning, Kunda (1990) argues that motivation to arrive at particular conclusions enhances the use of strategies (such as biased memory search) likely to lead to the desired conclusion.

Content effects and intuitions of relevance have been studied extensively over the past quarter century using Wason's (1966) "selection task" (see Evans 1989 for a review). The selection task involves giving subjects a conditional rule in the form "If *p* then *q*." Subjects are shown one side of each of four cards, which respectively show the cases corresponding to *p*, *not-p*, *q*, and *not-q*. They are told that the cards show the value of *p* (i.e. *p* or *not-p*) on one side and the value of *q* on the other. Their task is to decide which of the cards must be turned over to check whether the rule is false. The "correct" choice, according to standard propositional logic, is to select the *p* card (which might have *not-q* on its back) and the *not-q* card (which might have *p* on its back), because these are the only two possibilities that would falsify the rule. Subjects seldom make the correct choice when the conditional rule has arbitrary content (e.g. "If a card has an 'A' on one side, then it must have a '4' on the other"). Rather, they tend to make various errors, of which the most common is to select the cards corresponding to *p* and *q* (i.e. "A" and "4"). In contrast, for certain comparable rules that can be interpreted as expressing permission or obligation relations, such as "If a person is to drink alcohol, then they must be at least 21 years old," the *p* and *not-q* cases are selected much more frequently (see, for example, D'Andrade 1982; Cheng & Holyoak 1985; Cosmides 1989; Girotto et al 1989b; Johnson-Laird et al 1972; Light et al 1990; Manktelow & Over 1991; Politzer & Nguyen-Xuan 1992).

To explain the influence of content on reasoning in the selection task and other tasks (such as linguistic rephrasing) involving inference with conditionals, Cheng & Holyoak (1985) suggested that thematic content evokes pragmatic reasoning schemas: sets of rules that deal with situations defined in terms of recurring classes of goals and relationships to these goals. Pragmatic reasoning schemas fall into broad categories, of which prominent varieties are those dealing with causal inferences (Cheng & Nisbett, 1993; Tversky & Kahneman 1980) and those dealing with inferences based on the concepts of permission and obligation (Cheng & Holyoak 1985; Cheng et al 1986). The pragmatic schema theory predicts that performance on the selection task will depend on which schema (if any) is evoked by the content and context of the stated rule. For example, the "drinking age" rule just mentioned will tend to evoke a "permission schema," which applies when a precondition must be satisfied if a regulated action is to be taken. If a rule is interpreted as a conditional permission, the schema will focus attention on the case in which the action is taken (e.g. someone who drinks alcohol should be checked to be sure the age precondition has been met) and that in which the precondition has not been met (e.g. someone who is under age should be checked to be sure they are not drinking alcohol), because these two cases might reveal a violation. These are in fact the *p* and *not-q* cases—the selections dictated by standard logic. Accordingly, problems that evoke the permission schema show a dramatically greater frequency of these "correct" selections.

In addition to explaining patterns of facilitation for rules with concrete thematic content, Cheng & Holyoak (1985) demonstrated that facilitation could be obtained even for an abstract permission rule, "If one is to take action 'A', then one must first satisfy precondition 'P'" (also see Cheng & Holyoak 1989; Girotto, et al, 1992). The fact that people can reason reliably about rules with novel content or abstract content fulfills two of the major empirical criteria for rule use proposed by Smith et al (1992). In addition, the ability to reason about regulations has been demonstrated in children as young as 6 years old (Girotto et al 1988; Legrenzi & Murino 1974; Light et al 1989).

As Cheng & Holyoak (1985) noted, evocation of a pragmatic schema will not necessarily lead to selection of the "logically correct" cases. The perceived relevance of cases may vary both across schemas (because different schemas may encourage different inferences) and within schemas (because the context may alter the mapping of the elements of the stated conditional onto components of the schema). Moreover, a single conditional may be potentially mapped onto multiple schemas. Subjects who are encouraged to take different perspectives on a rule may interpret the rule in terms of different schemas, each of which yields a distinct response pattern (Politzer & Nguyen-Xuan 1992). Politzer & Nguyen-Xuan's analysis can account for other demonstrations that subjects' perspectives guide their evaluation of conditional regulations (Cosmides 1989; Gigerenzer & Hug 1992; Manktelow & Over 1991).

It has been suggested that people only have one special case of the permission schema—that in which someone who receives a rationed benefit must pay a cost (Cosmides 1989). In fact, however, many findings are inconsistent with this restriction (e.g. Cheng & Holyoak 1989; Girotto et al 1989a; Manktelow & Over 1990; Politzer & Nguyen-Xuan 1992). For example, Manktelow & Over (1990) found facilitation in the selection task for the conditional precaution, "If you clean up spilt blood, then you must wear rubber gloves," where there was no suggestion that cleaning up spilt blood was a rationed benefit for which one must pay a cost. Although Cosmides (1989) reported failing to obtain facilitation with some conditional rules, none of them was unambiguously cast as a permission situation for subjects (Cheng & Holyoak 1989; Pollard 1990).

The influence of content on reasoning extends well beyond the domain of social regulations. Work on causal reasoning—both inductive and deductive—reveals that people have inference procedures that are to some extent specialized for reasoning about cause and effect relations (e.g. Cheng & Nisbett, 1993; Cheng & Novick 1990, 1991, 1992; Hilton & Slugoski 1986; Kahneman & Miller 1986; Tversky & Kahneman 1980). One line of research has investigated the conditions under which people base causal judgments on the contingency between potential causal factors and an effect, where the contingency is defined as the proportion of events for which the effect occurs when a factor is present vs absent. Contingency is therefore sensitive not only to information about what occurs when the causal factor is present, but also to information about what happens in its absence. For example, in evaluating whether smoking causes cancer, information about nonsmokers who do not develop cancer is relevant. Cheng & Novick (1990) found that people's causal attributions could be reliably predicted from a contingency computation—as long as the set of events over which contingency is computed was taken into account (also Novick, Fratianne & Cheng, 1992). A number of apparent biases in causal attribution, such as a bias to attribute effects to a person rather than a situation, can be attributed to the fact that experimenters have not been fully aware of the information their subjects were using to compute contingency (Cheng & Novick 1992). Often people do not compute contingency over all possible cases, but rather some subset of cases—the *focal set* (Cheng & Novick 1990)—that they consider pragmatically relevant in the context. Variations in focal sets have been shown to account for people's intuitions about the distinction between causes and enabling conditions (Cheng & Novick 1991). For example, people will typically perceive a lightning strike as the cause of a forest fire, but they will view the presence of oxygen as merely an enabling condition, even though the lightning and the oxygen (along with other factors, such as the presence of combustible material) were individually necessary and jointly sufficient to yield the fire. In another context, for example that of a special oxygen-free laboratory, oxygen will be considered the cause of a fire that breaks out when it seeps into the lab. The distinction between causes and conditions thus depends on pragmatic contextual influences, rather than simply on the formal properties of necessity and sufficiency.

The application of causal schemas is also constrained by factors other than contingency, most notably temporal directionality: People assume that causes must precede their effects (Bullock et al 1982; Tversky & Kahneman 1980). Waldmann & Holyoak (1992) have shown that when a causal context is imposed on a task of classification learning, the pattern of performance differs radically across a predictive context, in which the cues are interpreted as possible causes of a common effect, and a diagnostic context, in which the cues are interpreted as possible effects of a common cause. In particular, competition among cues during learning is reduced or eliminated when they are perceived as joint effects of a common cause, rather than as alternative causes of a common effect. These differing patterns of cue competition suggest that people have a natural tendency to induce contingencies from causal factors to effects, rather than the reverse, even when the order in which information is presented is "effect followed by cause."

Taken as a whole, work on pragmatic reasoning indicates that thinking is heavily constrained by semantic and pragmatic content, and that the effects of broad classes of content are interpretable in terms of schemas that are relatively abstract, although less so than rules of formal logic. The forms of inference generated by the two classes of schemas that have received closest scrutiny—causal and regulation schemas—are very different, reflecting the differing goals associated with these domains. Whereas causal schemas serve to guide informative prediction, diagnosis, and explanation, the permission and obligation schemas govern assessment of conformity with contractual agreements and maintenance of freedom of choice for individuals within the limits of established regulations. Each schema provides a unique set of inference rules that embodies relevance relations appropriate for the pertinent goals.

The Context of Learning and the Content of Transfer

The issue of relevance arises again in connection with the transfer of knowledge from the context of learning to other related situations. Essentially by definition, transfer is based on the perception that prior knowledge is relevant to the current context. Transfer is in turn intimately related to the nature of expertise: We typically think of an "expert" as someone who is particularly good at recognizing the relevance of domain knowledge to new problems.

Expertise, however, may come in two qualitatively distinct varieties, only one of which promotes transfer across contexts. Hatano & Inagaki (1986; Hatano 1988) have drawn a distinction between routine and adaptive expertise; Salomon & Perkins (1989) elucidate a related distinction between "low-road" and "high-road" mechanisms of transfer. Routine expertise is characterized by rapid and accurate solution of well-practiced types of problems; adaptive expertise is characterized by flexible transfer of knowledge to novel types of problems and the ability to invent new procedures derived from expert knowledge. In terms of the distinctions drawn earlier, routine expertise may be based on implicit knowledge of procedures, whereas adaptive expertise may depend on more explicit and abstract representations. Of course, it is possible for a single individual to demonstrate both forms of expertise. Current production-system models of learning, with their emphasis on the acquisition of more specialized production rules through knowledge compilation, can be characterized as attempts to explain routine expertise (e.g. Anderson 1987; Rosenbloom et al 1991). These models have been directed primarily at accounting for stable superior performance on representative tasks, for which reproductive methods and specific knowledge are in fact central. Indeed, such theories are often described as models of "skill acquisition," which as Wenger (1987) has pointed out, is not coextensive with expertise: "Whereas skill acquisition can be tested by straightforward performance measures, expertise is a much more subtle notion .... [It] must also be evaluated by the capacity to handle novel situations, to reconsider and explain the validity of rules, and to reason about the domain from first principles ... " (p. 302).

Hatano & Inagaki (1986) suggest that the key to adaptive expertise—which involves facility in the recognition of relevance relations across contexts—is the development of a deeper and more explicit understanding of the target domain (cf Karmiloff-Smith 1990). Such understanding is heavily dependent on the conditions under which learning takes place. Understanding is more likely to result when the task is variable and in some degree unpredictable rather than stereotyped, and when the task is explored freely without heavy pressure to achieve an immediate goal (Sweller 1988). Understanding can result from sensitivity to internally generated feedback, such as surprise at a predictive failure, perplexity at noticing alternative explanations of a phenomenon, and discoordination due to lack of explanatory links between pieces of knowledge that apparently should be related. Understanding is also fostered by social support and encouragement of deeper comprehension, and by efforts to explain a task to others or to oneself. For example, Chi et al (1989) found that better students of physics, as measured by transfer performance, took a more active approach to learning from worked examples of word problems than did weaker students. The better students continually tried to explain *why* the steps of the illustrated solutions were required.

Analogical transfer—transfer of structural knowledge between specific situations—demonstrates an important bridge between context-bound and abstract knowledge. Theories of analogical thinking must attempt to explain when and how a novel target situation will evoke potentially useful source analogs stored in memory. The issue of how relevance relations can be recognized is thus central to analogical transfer. As we noted earlier, one general proposal (Thagard et al 1990) is that analog retrieval is governed by three types of constraints on the mappings between elements of the target and those of potential source analogs: semantic similarity (i.e. preference for mappings between taxonomically related concepts), isomorphism (i.e. preference for one-to-one mappings in which corresponding elements consistently fill parallel roles), and pragmatic centrality (i.e. preference for mappings involving elements deemed to be especially important to goal attainment). Empirical evidence suggests that for novices in a knowledge domain, retrieval is dominated by semantic similarity but that isomorphism also plays a role (e.g. Holyoak & Koh 1987; Keane 1988; Ratterman & Gentner 1987; Ross 1987, 1989; Seifert et al 1986; Wharton et al 1992). Access is improved if the source and target have similar goal structures. Schank (1982) placed particular emphasis on the importance to the reminding process of encodings that are influenced by goal failures. Recent evidence indicates that an initial goal failure experienced in connection with a source problem increases the likelihood that it will be retrieved in the context of a subsequent target problem in which an analogous impasse is reached (Gick & McGarry 1992; Read & Cesa 1991).

An important component of the development of expertise appears to be the induction of more abstract knowledge structures, such as rules and schemas, that serve to "highlight" problem-relevant aspects of situations, including less salient relations that are crucial to finding solutions (e.g. Chi et al 1981; Sweller 1988). In a kind of bootstrapping, analogical reasoning between problem examples fosters schema induction, schematic knowledge yields more expert transfer across superficially different content domains, and expertise permits more effective processing of novel analogs (Brown 1989; Catrambone & Holyoak 1989; Gick & Holyoak 1983; Novick 1988; Novick & Holyoak 1991; Ross & Kennedy 1990). Transfer of problem-solving procedures appears to be limited by the diversity of the content represented in the learning context, and by the structural parallels between the concepts in the acquired schema and the concepts that the learner uses to represent the target domain (Bassok 1990).

In addition to laboratory studies of learning and transfer, related lines of research have examined these processes in naturalistic contexts, investigating the roles of social and cultural contexts in guiding thinking. The work on cultural constraints in learning includes detailed studies of apprenticeship learning (Lave & Wenger 1991). Other research examines the differences between skills (such as mathematical strategies) as they emerge from formal instruction vs informal cultural practices (e.g. Carraher et al 1985, 1988; Lave 1988; Saxe 1982, 1988, 1991; Stevenson & Stigler 1992).

The "cultural practice" model offered by the Laboratory of Comparative Human Cognition (LCHC 1983) fostered the extreme view that thinking is simply a collection of cognitive skills, each independently acquired in a specific social context and inextricable from that context. This view, a version of what is sometimes termed "situated cognition," has led some people to conclude that it is not possible to understand thinking in terms of the individual's cognitive processes [see Vera & Simon (1993) for a critical discussion]. Early evidence from laboratory studies indicating that spontaneous cross-domain analogical transfer is difficult to obtain with novices (e.g. Gick & Holyoak 1980) was taken as evidence that transfer inevitably depends on the social organization of experience: "*Transfer is arranged by the social and cultural environment.* This shift of focus does not so much solve the transfer problem as it dissolves it" (LCHC 1983:341; italics in original).

Such extreme "situationism" provides an overly restrictive picture of the impact of culture and context on thinking. There are in fact important variations in the degree to which culturally embedded learning impedes or promotes flexible transfer. Hatano (1988) exemplifies the distinction between routine and adaptive expertise with a cross-cultural contrast between two forms of mathematical calculation skills: use of the abacus in Japan and other Asian cultures (e.g. Hatano & Osawa 1983), and the "street math" of Brazilian children working as vendors. Expertise in use of the abacus leads to extremely rapid calculations and to increased digit span; however, such knowledge cannot be readily generalized to repair "buggy" pencil-and-paper arithmetic procedures (Amaiwa 1987) or to use nonconventional abacuses with different base values. In contrast, unschooled Brazilian children who acquire arithmetic skills in the context of selling merchandise on the street can adapt general components of their procedures, such as decomposition and regrouping, to solve novel problems both on the street and in classroom mathematics (Saxe 1991). Hatano suggests that the primary difference between the two skills is that representations of number relations on the abacus are impoverished in meaning, whereas those used in street math are semantically transparent, analogous to actual activities dealing with goods and money. In addition, abacus use is basically a solitary skill in which speed and accuracy are the dominant goals, whereas street math is a social enterprise in which transparency to the customer is more important than speed.

In general, as Guberman & Greenfield (1991) have argued, sociocultural studies of everyday cognition provide evidence that dovetails nicely with laboratory research on learning and transfer. The extent of transfer varies enormously as a function of the content and context of learning. In both formal and informal settings, degree of transfer depends on the induction of abstract schemas, which is fostered by such factors as diversity of learning contexts, free exploration of the results produced by applying problem-solving operators, and perceived similarity of goal structures across examples.

Vivid Representations for Reasoning

The role of content in reasoning is intimately related to debates about the nature of the representational systems available for human thinking. Few if any issues in psychology or cognitive science have been debated as vigorously as the question of whether (or when) people think in "images" or "propositions." The debate peaked in the 1970s (Kosslyn et al 1979; Kosslyn & Pomerantz 1977; Pylyshyn 1973), stimulating a great deal of important empirical research but also, owing to the lack of theoretical clarity, causing some psychologists to become pessimistic about the very possibility of evaluating competing cognitive models (Anderson 1978). After a relative lull of a decade or so, basic questions about distinctions among types of human representations are now being reopened. Several lines of research bear upon these issues, including studies of imagery and the use of diagrams in problem solving, logical analyses of graphical representations, and theoretical proposals about "mental models."

We use the term *vivid representations* to refer to representations of the general sort just mentioned. The psychological character of these representations remains a matter for investigation, and so we use the term informally. It is inspired, however, by Levesque's (1986) characterization of "vivid knowledge" in logical systems—a conception not tied to spatial representations. The key idea is that reasoning is often facilitated when it can make use of representations in which information is definite rather than vague (Stenning & Oberlander 1992). In a vivid representation, a finite number of objects are represented and each is associated with definite values for all relevant properties and relations.

Representational systems vary in the degree of vagueness they permit. Imagine a simple world consisting solely of three animals (a fox, a pig, and a hen) standing in three positions, ordered from left to right. We might describe the current state of this world using sentences such as "The fox is left of the pig" and "The pig is left of the hen." Note the proposition expressed by the first sentence leaves the location of the hen unspecified, while the second is vague about the fox. From this sentential representation we can draw no further inferences in the absence of explicit inference rules. Such paucity of immediate inferences characterizes what we might term *pallid* representations.

In contrast, to illustrate a vivid representation let us represent this same simple world using a system quite different from sentences: an imagined horizontal line with three positions on it. Suppose we establish correspondences between the positions of the animals in the world and locations on the line. Each complete assignment of animal positions to line locations will constitute a model of this simple world. Now we can assess whether various propositions are true in particular models. Let the letters F, H and P stand for the positions of the fox, hen, and pig, respectively. There will be three models in which it is true that the fox is left of the pig (F P H, F H P, H F P) and three models in which it is true that the pig is left of the hen (P H F, P F H, F P H). The two sets of models have only one model in common, F P H; i.e. there is only one model in which both propositions are true. (Finding this unique model need not require exhaustive search of all the possibilities; rather, one might first represent the animals as described by the initial premise, and then add the third animal introduced in the second premise in a way that maintains consistency with the first premise.) Thus from two propositions, each of which fixes the relative position of two objects, we can deduce the absolute position of all three, as well as an additional relative ordering (the fox is left of the hen).

Note these inferences do not require use of any explicit rules: The deductions follow from the basic structure of the represented world (the finite number of animals in a restricted set of possible arrangements). If there is a unique model in which both premises are true, and we succeed in identifying it, we receive the inferences about the absolute locations of the objects as a "bonus." Such inferences will remain implicit in the model until some "read-out" procedure interprets it. Read-out procedures for models bear some resemblance to inference rules; however, there is one highly significant difference.

Whereas inference rules operate on given premises to yield conclusions, read-out procedures operating on a model completely blur the distinction between "premises" and "conclusions." Thus once the model F P H has been established as the representation of the stated propositions (and assuming the initial sentential representation of the premises has been lost), then the same read-out procedures would be required to derive from the model the fact that the fox is left of the pig (an initial premise) and the fact that the fox is left of the hen (an inferred conclusion). In fact, empirical studies of reaction time to make judgments of relative order for items in a memorized linear series have revealed that some valid inferences (based on relations between items far apart in the series) can be judged to be true more quickly than premises based on relations between adjacent items (e.g. Potts 1974). Such evidence that conclusions can be more accessible than premises is difficult to explain in terms of the operation of inference rules but is consistent with read-out mechanisms applied to imagined arrays. [See McGonigle & Chalmers (1986) for a discussion of the conditions under which a distinction in memory between premises and conclusions is maintained or lost.]

In general, a vivid representational system is one that *compels* specification of certain information and that specifies interdependencies between properties and relations so that a partially specified input can yield a definite model. In such a system, to think a certain thought may not only entail that you *can* think some other related thought (as in Fodor & Pylyshyn's 1988 concept of systematicity) but that you *are* thinking it. To form a model, it is necessary to have enough information to assign each represented object to a unique symbol in the model and to establish values for objects with respect to all relevant predicates. A wide range of evidence indicates that indeterminancy is in fact often highly detrimental to both memory and inference (Mani & Johnson-Laird 1982; McGonigle & Chalmers 1986). In addition, capacity limits on the number of objects that can be maintained in working memory imply that humans can only reason with vivid representations based on very small numbers of objects.

The fact that a vivid representational system promotes definiteness (by requiring that all variables be bound to values) does not prevent it from allowing abstraction (Stenning & Oberlander 1992). Abstraction may arise in several ways. First, a crucial aspect of the general characterization of a representational system is that it involves specifying which aspects of the represented world are relevant. Thus if only a subset of the objects and properties in the world is selected as relevant to the model, the model is allowed to be abstract—it need only be definite with respect to the selected subset of possible information. Second, vagueness can be represented by forming multiple models, each of which corresponds to a single determinate state of affairs, such that the set of models exhausts the possibilities. Third, vagueness of a property might be represented within a single model by introducing probability distributions of property values. Finally, the system may be augmented by explicit sentential statements: Representational vividness can be viewed as a continuum rather than a strict categorical distinction. Levesque succinctly captures the inherent trade-off between pallid and vivid representational systems: "... [T]he representational expressiveness of a language ... is not so much in what it allows you to say, but in what it allows you to *leave unsaid*. ... The more that is left unsaid, the more possibilities are allowed by what is said" (1988:370; italics in original). We would add the following: when more possibilities are allowed, fewer inferences can be made immediately.

The kinds of relation structures often assumed to underlie knowledge-based human thinking (e.g. schemas) are generally viewed as having properties that foster vivid representations. A schema specifies the relevant properties to be modeled, thus guiding relevant abstraction. Once a situation has been mapped into a schema, each relevant object will typically take on a definite value—either a determinate value or one generated as a default by the schema itself—with respect to each schema-relevant property. As Stenning & Oberlander (1992) point out, connectionist-style constraint satisfaction tends to naturally generate vivid representations in which a partially specified input may "complete itself," thus automatically performing a kind of default reasoning.

The fact that general representational concepts, such as schemas and connectionist networks, provide properties associated with vivid representations implies that vivid representations are not strictly tied to spatial imagery alone. Indeed, Stenning & Oberlander (1992) argue that the contrast between "images" and "propositions" is fundamentally misleading. In psychological discussions, "propositions" are often equated with sentences—if not natural-language sentences, then sentences in a language of the predicate-calculus or LISP-string style. But from the point of view of logic, propositions are abstract objects that can have truth values with respect to a represented world, and are not tied to any specific representational system. Under this usage, the information conveyed by a graph is just as propositional as that conveyed by a sentence. Thus the central issues for cognitive psychologists do not concern whether imagery is somehow nonpropositional but whether representational systems differ in the range of propositions they can express and the nature of the procedures they provide for drawing inferences. (Of course, imagistic and sentential representations may have different neural underpinnings.)

Although vivid representations are not necessarily imagistic, quasi-spatial representations certainly provide prominent examples of the vivid variety. A great deal of research indicates that people can manipulate visuospatial representations to make certain types of inferences, such as judgments of the similarity of rotated objects (Shepard & Metzler 1971), or of the shape formed by the two-dimensional projection of a rotated three-dimensional object (Pinker & Finke 1980; for a recent review see Finke 1989). Image-based inferences may also play a role in various forms of creative thinking (Finke 1990). There is considerable evidence that graphical and imagistic representations provide expressive and inferential power that differs from that afforded by sentential representations; for certain types of problems, the former type of system conveys distinct advantages. Larkin & Simon (1987) discuss the case of external diagrams of the sort used to capture relationships in physical space (e.g. pulley problems), in an ideal space (e.g. geometry problems), or in nonspatial domains that nonetheless can be mapped onto spatial displays (e.g. supply and demand functions in economics). They emphasize that diagrams organize information by location in a plane, and that information required for inferencing is therefore often present and explicit at a single location. Both recognition of meaningful elements and control of search (i.e. matching of elements to inference rules) are likely to be enhanced when using vivid visuospatial rather than pallid sentential representations. Recognition benefits from the highly specialized procedures available to the human visual system; search benefits because multiple elements that must be matched to a rule are typically found in close spatial proximity.

Because external diagrams are effective for certain types of problem solving, it is natural to expect that internal memory representations of a quasi-spatial nature would also prove useful. Koedinger & Anderson (1990) describe a simulation model of expert theorem proving in geometry based on quasi-spatial schemas. Each schema is a cluster of geometry facts associated with a prototypical geometric image. Schemas enable efficient forward search from given information to the goal, a central characteristic of expert performance in this domain. Koedinger & Anderson found that the schema-based model could account for the steps that experts skipped mentioning in verbal protocols. The main steps experts did tend to mention corresponded to "whole-statements" (the overall conclusion supported by the configuration, such as the fact that two triangles with certain properties are congruent). Thus schemas seem to serve as "macro-operators," shortening the number of steps required to execute a procedure. It is interesting, however, that current production-system models of the formation of macro-operators (Anderson 1987; Rosenbloom et al 1991) were unable to account for how such geometry macro-operators could be learned. In these production-system models, macro-operators are formed by conjoining consecutive production rules that are applied to achieve the same goal. In contrast, the geometry schemas involve macro-operators organized around objects and aggregations of objects in the domain. Other evidence suggests that induction of problem schemas is better fostered by free exploration of the search space (e.g. investigating the effects of applicable operators on objects) than by direct pursuit of specific goals (Sweller 1988; Sweller et al 1983).

The work of Johnson-Laird and his colleagues on "mental models" illustrates how vivid representations in working memory can be used in reasoning (e.g. Johnson-Laird 1983; Johnson-Laird & Byrne 1991; Johnson-Laird, Byrne, et al, 1992; Johnson-Laird et al 1989). In the mental-models framework, deductive reasoning is viewed as the construction and manipulation of models derived from the premises. The most compelling examples of the approach involve spatial reasoning. Consider the following two sets of premises:

Problem I Premises	Problem II Premises
A is on the right of B.	B is on the right of A.
C is on the left of B.	C is on the left of B.
D is in front of C.	D is in front of C.
E is in front of B.	E is in front of B.

For both sets of premises, one can ask the same question: What is the relation between D and E? For Problem I, it is possible to construct a single determinate model:

C	B	A
D	E	

yielding the conclusion that D is left of E. In contrast, Problem II requires two distinct models:

C	A	B	and	A	C	B
D		E			D	E

but both models support the conclusion that D is left of E. The two problems not only have the same answer, they are identical except for the first premise in each set. Moreover, the initial premise is in both cases irrelevant to the conclusion, because the relation between D and E depends only on the relation between C and B and not on the location of A. Although the problems are closely matched in form and content, the mental-models theory predicts that Problem I should be easier than Problem II because the first premise leads to multiple models for the latter problem. A study by Byrne & Johnson-Laird (1989) provided support for this prediction. Subjects were significantly more accurate for cases such as Problem I that had a valid conclusion based on a single model than for cases such as Problem II that had a valid conclusion based on multiple models (61% vs 50% correct). Indeterminate problems with no valid conclusion, which should theoretically require exhaustive scrutiny of multiple models, were much more difficult (18% correct). For such spatial deductive tasks it would be difficult to devise a proof-theoretic model that could account for the observed patterns of human reasoning performance.

Johnson-Laird and his colleagues have also proposed that people construct models in order to reason with syllogisms and propositional connectives. More complex assumptions are required to apply the mental-model approach to these more abstract tasks. It is assumed that sets of individuals or situations are mapped onto tokens in the model, thus eliminating variables; and that people tend to establish initial "default" models for the various logical terms (e.g. the quantifiers "all" and "some," and connectives such as "and" and "if"). Under some circumstances, it is assumed that people will "flesh out" their initial representation by constructing further possible models. Various symbolic devices are also postulated—e.g. special symbols to indicate negations, to denote which sets have been exhaustively considered, and to signify whether further models remain to be explored. The most general prediction of the approach is that problem difficulty will increase with the number of models that must be considered to arrive at a logically correct conclusion.

The overall success of the mental-models approach in this domain is so far mixed. On the positive side, several novel predictions of the model were confirmed in the Johnson-Laird et al (1992) study (e.g. deductions from exclusive disjunctions proved easier than those from inclusive disjunctions, as expected given the assumption that fewer models are required in the former case). Other inference phenomena are accounted for by auxillary assumptions. For example, to explain why people do not usually restate premises when drawing conclusions, Johnson-Laird et al assume that people keep track of the stated premises. Since a mental model (like vivid representations in general) does not preserve the identity of premises, some other representational system must presumably be helping out.

Other phenomena place greater strain on the theory. If mental models are intended to be vivid representations in the sense we have discussed, then the application of the approach to propositional reasoning is based on questionable assumptions. Johnson-Laird et al introduce a notation that allows for objects that are not fully specified—what they call an "implicit model." Of course, for vivid representations the notion of an "implicit model" is an oxymoron: The basic requirement for a vivid model is that it be fully definite. Given such problems, Johnson-Laird et al's account of propositional reasoning does not convincingly refute the view that formal inference rules such as *modus ponens* play some role in explaining human inference patterns. In any case, such a conclusion seems extremely unlikely given the highly evolved linguistic abilities of humans. Humans are presumably capable of reasoning both with pallid sentential representations, for which inference rules are well suited, and with more vivid representations.

Moreover, the mental-models approach to deductive reasoning is itself fundamentally based on formal procedures for representing arbitrary situations involving logical terms, just as are formal-rule theories. The procedures are directed at the manipulation of tokens rather than syntactic rules, but the two approaches are equally formal (in the sense of making minimal reference to semantic content; see footnote 4). Accordingly, theories based on mental models, like those based solely on formal rules, are unable to account for content effects in reasoning except by adding auxillary assumptions (often unacknowledged) about when people retrieve relevant counterexamples, schematic knowledge about types of situations, and so on (e.g. Johnson-Laird & Byrne 1992).

Thus although vivid representations undoubtedly play an important role in human thinking, an adequate theory must specify much more than procedures for manipulating symbolic tokens in mental models. The construction of mental models in working memory must be guided by retrieval of relevant knowledge structures from long-term memory, and the use of models requires read-out procedures that can make use of information implicit in the models. Moreover, as Stenning & Oaksford (1993) have argued, it would be desirable to show how initial default models in working memory emerge as a consequence of more primitive computational mechanisms, such as procedures for dynamic binding of objects to roles (e.g. Shastri & Ajjanagadde, 1993). In a constraint-satisfaction model of binding, only one set of consistent bindings of objects to roles (i.e. one vivid model) can be maintained at one time. This basic processing limitation offers an explanation of why it is that if multiple models must be considered to solve a deductive problem, each must be considered serially, yielding a concomitant increment in problem difficulty.

CONCLUSION

There can be no real "conclusion" to a review of a field in progress, any more than to a mid-career biography. What conclusions we have reached were largely laid out at the beginning of the chapter in our selection of central themes—the confluence of symbolic and connectionist perspectives on thinking, the relationship between implicit and explicit cognition, and the theoretical implications of the impact of content and context on thinking. These themes amount to our "best guess" about the trajectory of current research on thinking. There are hopeful signs that topics previously investigated in isolation from one another—for example, dynamic binding and vivid representations, implicit learning and neuropsychological organization, and sociocultural learning contexts and knowledge transfer—may prove deeply related. How these confluences will change the study of thinking remains to be seen.

ACKNOWLEDGMENTS

Preparation of this article was supported by Contract MDA 903-89-K-0179 from the Army Research Institute. We are grateful to the many researchers who kindly provided us with material to review, and we apologize for the fact that space limitations and the vagaries of the topics we selected precluded treatment of many significant papers. We thank Patricia Cheng, John Hummel, Carol Seger, Keith Stenning, and Michael Waldmann for helpful discussions as well as comments on earlier drafts. Requests for reprints should be sent to K. J. Holyoak, Department of Psychology, University of California, Los Angeles, CA 90024-1563.

Literature Cited

Ajzen, I. 1977. Intuitive theories of events and the effects of base rate information on prediction. *J. Pers. Soc. Psychol.* 35:303–14

Amaiwa, S. 1987. Transfer of substraction procedures from abacus to paper and pencil. *Jpn. J. Educ. Psychol.* 35:41–48 (In Japanese with English summary)

Anderson, J. R. 1978. Arguments concerning representations for mental imagery. *Psychol. Rev.* 85:249–77

Anderson, J. R. 1983. *The Architecture of Cognition.* Cambridge, MA: Harvard Univ.

Anderson, J. R. 1987. Skill acquisition: compilation of weak-method problem solutions. *Psychol. Rev.* 94:192–210

Anderson, J. R. 1990. *The Adaptive Character of Thought.* Hillsdale, NJ: Erlbaum

Anderson, J. R. 1991a. Is human cognition adaptive? *Behav. Brain Sci.* 14:471–517

Anderson, J. R. 1991b. The adaptive nature of human categorization. *Psychol. Rev.* 98:409–29

Anderson, J. R., Milson, R. 1989. Human memory: an adaptive perspective. *Psychol. Rev.* 96:703–19

Bar-Hillel, M., Fischhoff, B. 1981. When do base rates affect predictions? *J. Pers. Soc. Psychol.* 41:671–81

Barsalou, L. W., Hale, C. R. 1992. Components of conceptual representation: from feature lists to recursive frames. In *Categories and Concepts: Theoretical Views and Inductive Data Analysis,* ed. I. Van Mechelen, J. Hampton, R. Michalski, P. Theuns. San Diego, CA: Academic.

Bassok, M. 1990. Transfer of domain-specific problem-solving procedures. *J. Exp. Psychol.: Learn., Mem., Cognit.* 16:522–33

Berry, D. C., Broadbent, D. E. 1984. On the relationship between task performance and associated verbalizable knowledge. *Q. J. Exp. Psychol.* 36A:209–31

Berry, D. C., Broadbent, D. E. 1988. Interactive tasks and the implicit-explicit distinction. *Br. J. Psychol.* 79:251–72

Boden, M. A. 1990. *The Creative Mind: Myths and Mechanisms.* New York: Basic Books

Bowers, K. S., Regehr, G., Balthazard, C., Parker, K. 1990. Intuition in the context of discovery. *Cognit. Psychol.* 22:72–110

Braine, M. D. S., O'Brien, D. P. 1991. A theory of *if:* a lexical entry, reasoning program, and pragmatic principles. *Psychol. Rev.* 98:182–203

Broadbent, D. E., FitzGerald, P., Broadbent, M. H. P. 1986. Implicit and explicit knowledge in the control of complex systems. *Br. J. Psychol.* 77:33–50

Brooks, L. R., Vokey, F. R. 1991. Abstract analogies and abstracted grammars: comments on Reber (1989) and Mathews et al. 1989. *J. Exp. Psychol.: Gen.* 120:316–23

Brown, A. L. 1989. Analogical learning and transfer: What develops? In *Similarity and Analogical Reasoning,* ed. S. Vosniadou, A. Ortony, pp. 369–412. Cambridge: Cambridge Univ. Press

Bullock, M., Gelman, R., Baillargeon, R. 1982. The development of causal reasoning. In *The Developmental Psychology of Time,* ed. W. J. Friedman, pp. 209–53. San Diego: Academic

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

Byrne, R. M. J., Johnson-Laird, P. N. 1989. Spatial reasoning. *J. Mem. Lang.* 28:564–75

Carraher, T. N., Carraher, D. W., Schliemann, A. D. 1985. Mathematics in the streets and in the schools. *Br. J. Dev. Psychol.* 3:21–29

Carraher, T. N., Carraher, D. W., Schliemann, A. D. 1988. Mathematical concepts in everyday life. In *Children's Mathematics,* ed. G. B. Saxe, M. Gearhart, pp. 71–88. San Francisco: Jossey Bass

Catrambone, R., Holyoak, K. J. 1989. Overcoming contextual limitations on problem-solving transfer. *J. Exp. Psychol.: Learn., Mem., Cognit.* 15:1147–56

Chase, W. G., Simon, H. A. 1973. Perception in chess. *Cognit. Psychol.* 4:55–81

Cheng, P. W., Nisbett, R. E. 1993. A pragmatic constraint on causal deduction. In *Rules for Reasoning: Implications for Cognitive Science and Education,,* ed. R. E. Nisbett. Hillsdale, NJ: Erlbaum. In press

Cheng, P. W., Holyoak, K. J. 1985. Pragmatic reasoning schemas. *Cognit. Psychol.* 17:391–416

Cheng, P. W., Holyoak, K. J. 1989. On the natural selection of reasoning theories. *Cognition* 33:285–313

Cheng, P. W., Novick, L. R. 1990. A probabilistic contrast model of causal induction. *J. Pers. Soc. Psychol.* 58:545–67

Cheng, P. W., Novick, L. R. 1991. Causes versus enabling conditions. *Cognition* 40:83–120

Cheng, P. W., Novick, L. R. 1992. Covariation in natural causal induction. *Psychol. Rev.* 99:365–82

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., Oliver, L. M. 1986. Pragmatic versus syntactic approaches to training deductive reasoning. *Cognit. Psychol.* 18:293–328

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P. Glaser,, R. 1989. Self-explanations: how students study and use examples in learning to solve problems. *Cognit. Sci.* 13:145–82

Chi, M. T. H., Feltovich, P. J., Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. *Cognit. Sci.* 5:121–52

Christensen-Szalanski, J. J. J., Bushyhead, J. B. 1981. Physicians' use of probabilistic information in a real clinical setting. *J. Exp. Psychol.: Hum. Percept. Perform.* 7:928–35

Cleeremans, A., McClelland, J. L. 1991. Learning the structure of event sequences. *J. Exp. Psychol.: Gen.* 120:235–53

Cohen, A., Ivry, R. I., Keele, S. W. 1990. Attention and structure in sequence learning. *J. Exp. Psychol.: Learn., Mem., Cognit.* 16:17–30

Cosmides, L. 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187–276

Cummins, D. D., Lubart, T., Alksnis, O., Rist, R. 1991. Conditional reasoning and causation. *Mem. Cognit.* 19:274–82

D'Andrade, R. 1982. Reason versus logic. Paper presented at the Symp. Ecol. Cognit.: Biol., Cult., Hist. Perspect., Greensboro, NC

Dawkins, R. 1987. *The Blind Watchmaker.* New York: Norton

Dienes, Z. 1992. Connectionist and memory-array models of artificial grammar learning. *Cognit. Sci.* 16:41–80

Dienes, Z., Broadbent, D., Berry, D. 1991. Implicit and explicit knowledge bases in artificial grammar learning. *J. Exp. Psychol.: Learn., Mem., Cognit.* 17:875–87

Druhan, B., Mathews, R. 1989. THIYOS: a classifier system model of implicit knowledge of artificial grammars. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society,* pp. 66–73. Hillsdale, NJ: Erlbaum

Dulany, D. E., Carlson, R. A., Dewey, G. I. 1984. A case of syntactical learning and judgment: how conscious and how abstract? *J. Exp. Psychol.: Gen.* 113:541–55

Dunbar, K. 1993. Concept discovery in a scientific domain. *Cognit. Sci.* In press

Duncker, K. 1945. On problem solving. *Psychol. Monogr.* 58(270)

Dyer, M. G. 1991. Symbolic neuroengineering for natural language processing: a multilevel research approach. In *Advances in Connectionist and Neural Computation Theory.* Vol. 1: *High Level Connectionist Models,* J. Barnden, J. Pollack, pp. 32–86. Norwood, NJ: Ablex

Estes, W. K. 1991. Cognitive architectures from the standpoint of an experimental psychologist. *Annu. Rev. Psychol.* 42:1–28

Estes, W. K., Campbell, J. A., Hatsopoulos, N., Hurwitz, J. B. 1989. Base-rate effects in category learning: a comparison of parallel network and memory storage-retrieval models. *J. Exp. Psychol.: Learn., Mem., Cognit.* 15:556–71

Evans, J. St. B. T. 1989. *Bias in Human Reasoning: Causes and Consequences.* Hove, UK: Erlbaum

Evans, J. St. B. T., Barnston, J. L., Pollard, P. 1983. On the conflict between logic and belief in syllogistic reasoning. *Mem. Cognit.* 11:295–306

Falkenhainer, B., Forbus, K. D., Gentner, D. 1989. The structure-mapping engine: algorithm and examples. *Artif. Intell.* 41:1–63

Farah, M. J., McClelland, J. L. 1991. A computational model of semantic memory impairment: modality specificity and emergent category specificity. *J. Exp. Psychol.: Gen.* 120:339–57

Finke, R. A. 1989. *Principles of Mental Imagery.* Cambridge, MA: MIT Press

Finke, R. A. 1990. *Creative Imagery: Discoveries and Inventions in Visualization.* Hillsdale, NJ: Erlbaum

Flannagan, M. J., Fried, L. S., Holyoak, K. J. 1986. Distributional expectations and the induction of category structure. *J. Exp. Psychol.: Learn., Mem., Cognit.* 12:241–56

Fodor, J. A. 1983. *The Modularity of Mind.* Cambridge, MA: MIT Press

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

Fodor, J. A., Pylyshyn, Z. W. 1988. Connectionism and cognitive architecture: a critical analysis. In *Connections and Symbols,* ed. S. Pinker, J. Mehler, pp. 3–71. Cambridge, MA: MIT Press

Gick, M. L., Holyoak, K. J. 1980. Analogical problem solving. *Cognit. Psychol.* 12:306–55

Gick, M. L., Holyoak, K. J. 1983. Schema induction and analogical transfer. *Cognit. Psychol.* 15:1–38

Gick, M. L., McGarry, S. J. 1992. Learning from mistakes: Inducing analogous solution failures to a source problem produces later successes in analogical transfer. *J. Exp. Psychol.: Learn., Mem., Cognit.* 18:623–39

Gigerenzer, G., Hug, K. 1992. Domain-specific reasoning: social contracts, cheating, and perspective change. *Cognition* 43:127–71

Gigerenzer, G., Hell, W., Blank, H. 1988. Presentation and content: the use of base rates as a continuous variable. *J. Exp. Psychol.: Hum. Percept. Perform.* 14:513–25

Gigerenzer, G., Hoffrage, U., Kleinbölting, H. 1991. Probabilistic mental models: a Brunswikian theory of confidence. *Psychol. Rev.* 98:506–28

Girotto, V., Blaye, A., Farioli, F. 1989a. A reason to reason: pragmatic basis of children's search for counterexamples. *Euro. Bull. Cognit. Psychol.* 9:227–321

Girotto, V., Gilly, M., Blaye, A., Light, P. H. 1989b. Children's performance in the selection task: plausibility and familiarity. *Br. J. Psychol.* 80:79–95

Girotto, V., Light, P. H., Colburn, C. J. 1988. Pragmatic schemas and conditional reasoning in children. *Q. J. Exp. Psychol.* 40A:469–82

Girotto, V., Mazzocco, A., Cherubini, P. 1992. Judgments of relevance in deonic reasoning: a reply to Jackson and Griggs. *Q. J. Exp. Psychol.* In press

Gluck, M. A., Bower, G. H. 1988. From conditioning to category learning: an adaptive network model. *J. Exp. Psychol.: Gen.* 117:227–47

Goldstone, R. L., Medin, D. L. 1993. Similarity, interactive activation, and mapping: an overview. See Holyoak & Barnden 1993. In press

Goldstone, R. L., Medin, D. L., Gentner, D. 1991. Relational similarity and the nonindependence of features in similarity judgments. *Cognit. Psychol.* 23:222–62

Griffin, D., Tversky, A. 1992. The weighing of evidence and the determinants of confidence. *Cognit. Psychol.* 24:411–35

Guberman, S. R., Greenfield, P. M. 1991. Learning and transfer in everyday cognition. *Cognit. Dev.* 6:233–60

Halford, G. S., Wilson, W. H. 1980. A category theory approach to cognitive development. *Cognit. Psychol.* 12:356–441

Halford, G. S., Wilson, W. H., Guo, J., Wiles, J., Stewart, J. E. M. 1993. Connectionist implications for processing capacity limitations in analogies. See Holyoak & Barnden 1993. In press

Hasher, L., Zacks, R. T. 1984. Automatic processing of fundamental information: the case of frequency of occurrence. *Am. Psychol.* 39:1372–88

Hatano, G. 1988. Social and motivational bases for mathematical understanding. In *Children's Mathematics,* ed. G. B. Saxe, M. Gearhart, pp. 55–70. San Francisco: Jossey-Bass

Hatano, G., Inagaki, K. 1986. Two courses of expertise. In *Child Development and Education in Japan,* ed. H. Stevenson, H. Azuma, K. Hakuta, pp. 262–72. San Francisco: Freeman

Hatano, G., Osawa, K. 1983. Digit memory of grand masters in abacus-derived mental calculation. *Cognition* 15:95–110

Hayes, N. A., Broadbent, D. E. 1988. Two modes of learning for interactive tasks. *Cognition* 23:80–108

Hebb, D. O. 1961. Distinctive features of learning in the higher animal. In *Brain Mechanisms and Learning,* ed. J. F. Delafresnaye. London: Oxford Univ.

Hempel, C. 1965. *Aspects of Scientific Explanation.* New York: Free Press

Henle, M. 1962. On the relation between logic and thinking. *Psychol. Rev.* 69:366–78

Hilton, D. J., Slugoski, B. R. 1986. Knowledge-based causal attribution: the abnormal conditions focus model. *Psychol. Rev.* 93:75–88

Hofstadter, D. R. 1984. *The Copycat Project: An Experiment in Nondeterministic and Creative Analogies.* Cambridge, MA: MIT A. I. Lab. Memo 755

Hofstadter, D. R., Mitchell, M. 1993. An overview of the Copycat project. See Holyoak & Barnden 1993. In press

Holland, J. H., Holyoak, K. J., Nisbett, R. E., Thagard, P. R. 1986. *Induction: Processes of Inference, Learning, and Discovery.* Cambridge, MA: MIT Press

Holyoak, K. J. 1991. Symbolic connectionism: toward third-generation theories of expertise. In *Toward a General Theory of Expertise: Prospects and Limits,* ed. K. A. Ericsson, J. Smith, pp. 301–35. Cambridge: Cambridge Univ. Press

Holyoak, K. J., Barnden, J. A., eds. 1993. *Advances in Connectionist and Neural Computation Theory.* Vol. 2: *Analogical Connections.* Norwood, NJ: Ablex. In press

Holyoak, K. J., Koh, K. 1987. Surface and structural similarity in analogical transfer. *Mem. Cognit.* 15:332–40

Holyoak, K. J., Thagard, P. 1989. Analogical mapping by constraint satisfaction. *Cognit. Sci.* 13:295–355

Holyoak, K. J., Novick, L. R., Melz, E. R. 1993. Component processes in analogical transfer: mapping, pattern completion, and adaptation. See Holyoak & Barnden 1993. In press

Hummel, J. E., Biederman, I. 1992. Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* 99:480–517

Hummel, J. E., Holyoak, K. J. 1992. Indirect analogical mapping. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society,* pp. 516–21. Hillsdale, NJ: Erlbaum

Jacobs, R. A., Jordan, M. I., Barto, A. G. 1991. Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. *Cogn. Sci.* 15:219–50

Johnson-Laird, P. N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge, MA: Harvard Univ. Press

Johnson-Laird, P. N., Byrne, R. M. J. 1991. *Deduction.* Hillsdale, NJ: Erlbaum

Johnson-Laird, P. N., Byrne, R. M. J. 1992. Modal reasoning, models, and Manktelow and Over. *Cognition* 43:173–82

Johnson-Laird, P. N., Byrne, R. M. J., Schaeken, W. 1992. Propositional reasoning by model. *Psychol. Rev.* 99:418–39

Johnson-Laird, P. N., Byrne, R. M. J., Tabossi, P. 1989. Reasoning by model: the case of multiple quantification. *Psychol. Rev.* 96:658–73

Johnson-Laird, P. N., Legrenzi, P., Legrenzi, S. M. 1972. Reasoning and a sense of reality. *Br. J. Psychol.* 63:395–400

Jordan, M. I., Rosenbaum, D. A. 1989. Action. In *Foundations of Cognitive Science,* ed. M. I. Posner, pp. 727–67. Cambridge, MA: MIT Press

Just, M. A., Carpenter, P. A. 1992. A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99:122–49

Kahneman, D., Lovallo, D. 1993. Timid choices and bold forecasts: a cognitive perspective on risk taking. *Manage. Sci.*

Kahneman, D., Slovic, P., Tversky, A. 1982. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge Univ. Press

Kahneman, D., Miller, D. T. 1986. Norm theory: comparing reality to its alternatives. *Psychol. Rev.* 93:136–53

Kaiser, M. K., Proffitt, D. R. 1984. The development of sensitivity to causally-relevant dynamic information. *Child Dev.* 55:1614–

24

Kaiser, M. K., Proffitt, D. R., Anderson, K. 1985. Judgments of natural and anomalous trajectories in the presence and absence of motion. *J. Exp. Psychol.: Learn., Mem., Cognit.* 11:795–803

Kaiser, M. K., Jonides, J., Alexander, J. 1986. Intuitive reasoning about abstract and familiar physics problems. *Mem. Cognit.* 14:308–12

Kaplan, C. A., Simon, H. A. 1990. In search of insight. *Cognit. Psychol.* 22:374–419

Karmiloff-Smith, A. 1986. From meta-processess to conscious access: evidence from children's metalinguistic and repair data. *Cognition* 23:95–147

Karmiloff-Smith, A. 1990. Constraints on representational change: evidence from children's drawing. *Cognition* 34:57–83

Keane, M. 1988. *Analogical Problem Solving.* Chichester: Ellis Horwood

Kihlstrom, J. F. 1987. The cognitive unconscious. *Science* 237:1445–52

Kintsch, W. 1988. The role of knowledge in discourse comprehension: a construction-integration model. *Psychol. Rev.* 95:163–82

Klayman, J., Ha, Y. 1987. Confirmation, disconfirmation, and information in hypothesis testing. *Psychol. Rev.* 94:211–28

Klayman, J., Ha, Y. 1989. Hypothesis testing in rule discovery: strategy, structure and content. *J. Exp. Psychol.: Learn., Mem., Cognit.* 15:596–604

Knowlton, B. J., Ramus, S. J., Squire, L. R. 1992. Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances. *Psychol. Sci.* 3:172–79

Koedinger, K. R., Anderson, J. R. 1990. Abstract planning and perceptual chunks: elements of expertise in geometry. *Cognit. Sci.* 14:511–50

Koh, K., Meyer, D. E. 1991. Function learning: induction of continuous stimulus-response relations. *J. Exp. Psychol.: Learn., Mem., Cognit.* 17:811–36

Kosslyn, S. M., Pomerantz, J. R. 1977. Imagery, propositions, and the form of internal representations. *Cognit. Psychol.* 9:52–76

Kosslyn, S. M., Pinker, S., Smith, G. E., Schwartz, S. P. 1979. On the demystification of mental imagery. *Behav. Brain Sci.* 2:535–81

Krosnick, J. A., Li, F., Lehman, D. R. 1990. Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *J. Pers. Soc. Psychol.* 59:1140–52

Kunda, Z. 1990. The case for motivated reasoning. *Psychol. Bull.* 108:480–98

Kushner, M., Cleeremans, A., Reber, A. 1991. Implicit detection of event interdependencies and a PDP model of the process. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society,* pp. 215–20. Hillsdale, NJ: Erlbaum

Laboratory of Comparative Human Cognition. 1983. Culture and cognitive development. In *Handbook of Child Psychology:* Vol. 1. *History, Theories and Methods,* ed. P. H. Mussen, pp. 295–356. New York: Wiley

Lakoff, G. 1993. What is metaphor? See Holyoak & Barnden 1993. In press

Larkin, J., Simon, H. A. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognit. Sci.* 11:65–100

Lave, J. 1988. *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life.* New York: Cambridge Univ. Press

Lave, J., Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation.* New York: Cambridge Univ.

Legrenzi, P., Murino, M. 1974. Falsification at the pre-operational level. *Ital. J. Psychol.* 1:361–68

Levesque, H. 1986. Making believers out of computers. *Artif. Intell.* 30:81–108

Levesque, H. 1988. Logic and the complexity of reasoning. *J. Philos. Logic* 17:355–89

Lewicki, P., Hill, T., Bizot, E. 1988. Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognit. Psychol.* 20:24–37

Lewontin, R. C. 1990. The evolution of cognition. In *Invitation to Cognitive Science: Thinking,* 3:239–46. Cambridge, MA: MIT

Light, P. H., Blaye, A., Gilly, M., Girotto, G. 1989. Children's reasoning on conditional promises and permissions. *Cognit. Dev.* 5:369–83

Light, P. H., Girotto, V., Legrenzi, P. 1990. Children's reasoning on conditional promises and permissions. *Cognit. Dev.* 5:369–83

Mandler, J. M. 1992. How to build a baby: II. Conceptual primitives. *Psychol. Rev.* In press

Mani, K., Johnson-Laird, P. N. 1982. The mental representation of spatial descriptions. *Mem. Cognit.* 10:81–87

Manis, M., Dovenalina, I., Avis, N. E., Cardoze, S. 1980. Base rates can affect individual predictions. *J. Pers. Soc. Psychol.* 38:287–98

Manktelow, K. I., Over, D. E. 1990. Deontic thought and the selection task. In *Lines of Thought: Reflections on the Psychology of Thinking,* ed. K. J. Gilhooly, M. Keane, R. Logie, G. Erdos, pp. 153–64. Chichester: Wiley

Manktelow, K. I., Over, D. E. 1991. Social roles and utilities in reasoning with deontic conditionals. *Cognition* 39:85–105

Mannes, S. M., Kintsch, W. 1991. Routine computing tasks: planning as understanding. *Cognit. Sci.* 15:305–42

Markovits, H., Nantel, G. 1989. The belief-bias effect in the production and evaluation of logical conclusions. *Mem. Cognit.* 17:11–

17

Mathews, R. C. 1990. Abstractness of implicit grammar knowledge: comments on Perruchet and Pacteau's analysis of synthetic grammar learning. *J. Exp. Psychol.: Gen.* 119:412–16

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J., Druhan, B. 1989. Role of implicit and explicit processing in learning from examples: a synergistic effect. *J. Exp. Psychol.: Learn., Mem., Cognit.* 15:1083–1100

McCarthy, J. 1988. Epistemological challenges for connectionism. Commentary to "On the proper treatment of connectionism" by P. Smolensky. *Behav. Brain Sci.* 11:44

McCloskey, M., Kohl, D. 1983. Naive physics: the curvilinear impetus principle and its role in interactions with moving objects. *J. Exp. Psychol.: Learn., Mem., Cognit.* 9:146–56

McGonigle, B., Chalmers, M. 1986. Representations and strategies during inference. In *Reasoning and Discourse Processes,* ed. T. F. Myers, E. K. Brown, B. McGonigle. London: Academic

McKelvie, S. J. 1987. Learning and awareness in the Hebb digits task. *J. Gen. Psychol.* 114:75–88

Medin, D. L. 1989. Concepts and conceptual structure. *Am. Psychol.* 44:1469–81

Medin, D. L., Edelson, S. M. 1988. Problem structure and the use of base-rate information from experience. *J. Exp. Psychol.: Gen.* 117:68–85

Medin, D. L., Smith, E. E. 1984. Concepts and concept formation. *Annu. Rev. Psychol.* 35:113–38

Metcalfe, J. 1986. Premonitions of insight predict impending error. *J. Exp. Psychol.: Learn., Mem., Cognit.* 12:623–34

Metcalfe, J., Wiebe, D. 1987. Intuition in insight and non-insight problem solving. *Mem. Cognit.* 15:238–46

Miller, G. A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63:81–97

Minsky, M. 1975. A framework for representing knowledge. In *Metaphor and Thought,* ed. P. H. Winston. Cambridge: Cambridge Univ. Press

Minsky, M. 1991. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Mag.* 12(2):34–51

Mynatt, C. R., Doherty, M. E., Tweney, R. D. 1978. Consequences of confirmation and disconfirmation in a simulated research environment. *Q. J. Exp. Psychol.* 30:395–406

Newell, A. 1973. Production systems: models of control structures. In *Visual Information Processing,* ed. W. G. Chase, pp. 463–526. New York: Academic

Newell, A. 1990. *Unified Theories of Cognition.* Cambridge, MA: Harvard Univ. Press

Newell, A., Simon, H. A. 1972. *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall

Nisbett, R. E., Krantz, D. H., Jepson, C., Kunda, Z. 1983. The use of statistical heuristics in everyday inductive reasoning. *Psychol. Rev.* 90:339–63

Nissen, M. J., Bullemer, P. 1987. Attentional requirements of learning: evidence from performance measures. *Cognit. Psychol.* 19:1–32

Novick, L. R. 1988. Analogical transfer, problem similarity, and expertise. *J. Exp. Psychol.: Learn., Mem., Cognit.* 14:510–20

Novick, L. R., Holyoak, K. J. 1991. Mathematical problem solving by analogy. *J. Exp. Psychol.: Learn., Mem., Cognit.* 17:398–415

Novick, L. R., Fratianne, A., Cheng, P. W. 1992. Knowledge-based assumptions in causal attribution. *Soc. Cognit.* 10:299–332

Oakhill, J., Johnson-Laird, P. N., Garnham, A. 1989. Believability and syllogistic reasoning. *Cognition* 31:117–40

Oaksford, M., Chater, N. 1993. Reasoning theories and bounded rationality. In *Rationality,* ed. K. I. Manktelow, D. E. Over. London: Routledge . In Press

Oaksford, M., Chater, N., Stenning, K. 1990. Connectionism, classical cognitive science and experimental psychology. *AI Soc.* 4:73–90

Oden, G. C. 1987. Concept, knowledge and thought. *Annu. Rev. Psychol.* 38:203–28

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., Shafir, E. 1990. Category-based induction. *Psychol. Rev.* 97:185–200

Palmer, S. 1978. Fundamental aspects of cognitive representation. In *Cognition and Categorization,* ed. E. Rosch, B. B. Lloyd, pp. 259–302. Hillsdale, NJ: Erlbaum

Patel, V. L., Groen, G. J. 1991. The general and specific nature of medical expertise: a critical look. In *Toward a General Theory of Expertise: Prospects and Limits,* ed. K. A. Ericsson, J. Smith, pp. 93–125. Cambridge: Cambridge Univ. Press

Payne, J. W., Bettman, J. R., Johnson, E. J. 1992. Behavioral decision research: a constructive processing perspective. *Annu. Rev. Psychol.* 43:87–131

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufmann

Perruchet, P., Amorim, M.-A. 1992. Conscious knowledge and changes in performance in sequence learning: evidence against dissociation. *J. Exp. Psychol.: Learn., Mem., Cogn.* 18:785–800

Perruchet, P., Gallego, J., Pacteau, C. 1992. A reinterpretation of some earlier evidence for abstractiveness of implicitly acquired knowledge. *Q. J. Exp. Psychol.* 44A:193–

210

Perruchet, P., Gallego, J., Savy, I. 1990. A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognit Psychol.* 22:493–516

Perruchet, P., Pacteau, C. 1990. Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *J. Exp. Psychol.: Gen.* 119:264–75

Perruchet, P., Pacteau, C. 1991. The implicit acquisition of abstract knowledge about artificial grammar: some methodological and conceptual issues. *J. Exp. Psychol.: Gen.* 120:112–16

Pew, R. W. 1974. Levels of analysis in motor control. *Brain Res.* 71:393–400

Pinker, S., Finke, R. A. 1980. Emergent two-dimensional patterns in images rotated in depth. *J. Exp. Psychol.: Hum. Percept. Perform.* 6:244–64

Politzer, G., Braine, M. D. S. 1991. Responses to inconsistent premises cannot count as suppression of valid inferences: comment on Byrne. *Cognition* 38:103–8

Politzer, G., Nguyen-Xuan, A. 1992. Reasoning about promises and warnings: Darwinian algorithms, mental models, relevance judgments or pragmatic schemas? *Q. J. Exp. Psychol.* 44A:401–21

Pollard, P. 1990. Natural selection for the selection task: limits to the social exchange theory. *Cognition* 36:195–204

Potts, G. R. 1974. Storing and retrieving information about ordered relationships. *J. Exp. Psychol.* 103:431–39

Proffitt, D. R., Kaiser, M. K., Whelan, S. M. 1990. Understanding wheel dynamics. *Cognit. Psychol.* 22:342–73

Pylyshyn, Z. 1973. What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychol. Bull.* 80:1–24

Quine, W. V. O. 1961. *From a Logical Point of View.* New York: Harper Torchbooks. 2nd ed.

Ranney, M. 1993. Explorations in explanatory coherence. In *Designing Intelligent Learning Environments: From Cognitive Analysis to Computer Implementation,* ed. E. Bar-On, B. Eylon, Z. Schertz. Norwood, NJ: Ablex. In press

Ratcliff, R., McKoon, G. 1989. Similarity information versus relational information: differences in the time course of retrieval. *Cognit. Psychol.* 21:139–55

Ratterman, M. J., Gentner, D. 1987. Analogy and similarity: determinants of accessibility and inferential soundness. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society,* pp. 22–34. Hillsdale, NJ: Erlbaum

Read, S. J., Cesa, I. L. 1991. This reminds me of the time when ...: expectation failures in reminding and explanation. *J. Exp. Soc. Psychol.* 27:1-25

Read, S. J., Marcus-Newhall, A. 1993. The role of explanatory coherence in social explanations. *J. Pers. Soc. Psychol.* In press

Reber, A. S. 1967. Implicit learning of artificial grammars. *J. Verb. Learn. Verb. Behav.* 6:855–63

Reber, A. S. 1969. Transfer of syntactic structure in synthetic languages. *J. Exp. Psychol.* 81:115–19

Reber, A. S. 1989. Implicit learning and tacit knowledge. *J. Exp. Psychol.: Gen.* 118:219–35

Reber, A. S. 1992. The cognitive unconscious: an evolutionary perspective. *Consciousness Cognit.* 1:93–133

Reber, A. S., Allen, R., Regan, S. 1985. Syntactical learning and judgment, still unconscious and still abstract: comment on Dulaney, Carlson, and Dewey. *J. Exp. Psychol.: Gen.* 114:17–24

Reber, A. S., Kassin, S. M., Lewis, S., Cantor, G. 1980. On the relationship between implicit and explicit modes in the learning of a complex rule structure. *J. Exp. Psychol.: Hum. Learn. Mem.* 6:492–514

Richardson-Klavehn, A., Bjork, R. A. 1988. Measures of memory. *Annu. Rev. Psychol.* 39:475–543

Rips, L. J. 1989. The psychology of knights and knaves. *Cognition* 31:85–116

Rips, L. J. 1990. Reasoning. *Annu. Rev. Psychol.* 41:321–53

Rips, L. J., Conrad, F. G. 1989. Folk psychology of mental activities. *Psychol. Rev.* 96:187–207

Rock, I., Palmer, S. 1990. The legacy of Gestalt psychology. *Sci. Am.* 263(6):84–90

Roediger, H. L. III. 1990. Implicit memory: retention without remembering. *Am. Psychol.* 45:1043–56

Rosenbloom, P. S., Laird, J. E., Newell, A., McCarl, R. 1991. A preliminary analysis of the Soar architecture as a basis for general intelligence. *Artif. Intell.* 47:289–325

Ross, B. H. 1987. This is like that: the use of earlier problems and the separation of similarity effects. *J. Exp. Psychol.: Learn., Mem., Cognit.* 13:629–39

Ross, B. H. 1989. Distinguishing types of superficial similarities: different effects on the access and use of earlier problems. *J. Exp. Psychol.: Learn., Mem., Cognit.* 15:456–68

Ross, B. H., Kennedy, P. T. 1990. Generalizing from the use of earlier examples in problem solving. *J. Exp. Psychol.: Learn., Mem., Cognit.* 16:42–56

Rousseau, D. L. 1992. Case studies in pathological science. *Am. Sci.* 80:54–63

Rozin, P. 1976. The evolution of intelligence and access to the cognitive unconsious. *Progr. Psychobiol. Physiol. Psychol.* 6:245–80

Rubin, D. C., Wallace, W. T., Houston, B. C.

1993. The beginnings of expertise for ballads. *Cogn.it Sci.* In press
Rumelhart, D. E., Smolensky, P., McClelland, J., Hinton, G. E. 1986. Schemata and sequential thought processes in PDP models. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* ed. J. L. McClelland, D. E. Rumelhart, and the PDP Research Group, 2:7–57. Cambridge, MA: MIT Press
Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, eds. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Cambridge, MA: MIT Press. 2 vols.
Rutkowska, J. C. 1987. Computational models and developmental psychology. In *Computation and Development,* pp. 187–215. Chichester: Wiley
Salomon, G., Perkins, D. N. 1989. Rocky roads to transfer: rethinking mechanisms of a neglected phenomenon. *Educ. Psychol.* 24:113–42
Sanders, R. E., Gonzalez, E. G., Murphy, M. D., Liddle, C. L., Vitina, J. R. 1987. Frequency of occurrence and the criteria for automatic processing. *J. Exp. Psychol.: Learn., Mem., Cognit.* 13:241–50
Saxe, G. B. 1982. Developing forms of arithmetic operations among the Oksapmin of Papua New Guinea. *Dev. Psychol.* 18:583–94
Saxe, G. B. 1991. *Culture and Cognitive Development: Studies in Mathematical Understanding.* Hillsdale, NJ: Erlbaum
Schacter, D. L. 1987. Implicit memory: history and current status. *J. Exp. Psychol.: Learn., Mem., Cognit.* 13:501–18
Schacter, D. L., Chiu, C.-Y. P., Ochsner, K. N. 1993. Implicit memory: a selective review. *Annu. Rev. Neurosci.* 16:159–82
Schank, P., Ranney, M. 1991. The psychological fidelity of ECHO: modeling an experimental study of explanatory coherence. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society,* pp. 892–97. Hillsdale, NJ: Erlbaum
Schank, P., Ranney, M. 1992. Assessing explanatory coherence: a new method for integrating verbal data with models of on-line belief revision. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society,* pp. 599–604. Hillsdale, NJ: Erlbaum
Schank, R. C. 1982. *Dynamic Memory.* New York: Cambridge Univ. Press
Schwartz, M., Bryden, M. P. 1971. Coding factors in the learning of repeated digit sequences. *J. Exp. Psychol.* 87:331–34
Schwarz, N., Strack, F., Hilton, D., Naderer, G. 1991. Base-rates, representativeness, and the logic of conversation: the contextual relevance of "irrelevant" information. *Soc. Cognit.* 9:67–84
Seger, C. A. 1992. Implicit learning. Tech. Rep. UCLA-CSRP-92-3, Cogn. Sci. Res. Progr., Univ. Calif., Los Angeles
Seifert, C. M., McKoon, G., Abelson, R. P., Ratcliff, R. 1986. Memory connections between thematically similar episodes. *J. Exp. Psychol.: Hum. Learn. Mem.* 12:220–31
Servan-Schreiber, E., Anderson, J. R. 1990. Learning artificial grammars with competitive chunking. *J. Exp. Psychol.: Learn., Mem., Cognit.* 16:592–608
Shanon, B. 1976. Aristotelianism, Newtonianism and the physics of the layman. *Perception* 5:241–43
Shastri, L., Ajjanagadde, V. 1993. From simple associations to systematic reasoning: a connectionist representation of rules, variables and dynamic bindings. *Behav. Brain Sci.* 16
Shepard, R. N., Metzler, J. 1971. Mental rotation of three-dimensional objects. *Science* 171:701–3
Sherry, D. F., Schacter, D. L. 1987. The evolution of multiple memory systems. *Psychol. Rev.* 94:439–54
Simon, H. A. 1986. The information processing explanation of Gestalt phenomena. *Comput. Hum. Behav.* 2:241–55
Simon, H. A. 1990. Invariants of human behavior. *Annu. Rev. Psychol.* 41:1–19
Simon, H. A. 1991. Cognitive architectures and rational analysis: comment. In *Architectures for Intelligence: The 22nd Carnegie Mellon Symposium on Cognition,* ed. K. VanLehn. Hillsdale, NJ: Erlbaum
Smith, E. E., Langston, C., Nisbett, R. E. 1992. The case for rules in reasoning. *Cognit. Sci.* 13:145–82
Smolensky, P. 1988. On the proper treatment of connectionism. *Behav. Brain Sci.* 11:1–74
Smolensky, P. 1990. Tensor product variable binding and the representation of symbolic structure in connectionist systems. *Artif. Intell.* 46(1–2):159–216
Spellman, B. A., Holyoak, K. J. 1992. If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *J. Pers. Soc. Psychol.* 62:913–33
Stanley, W. B., Mathews, R. C., Buss, R. R., Kotler-Cope, S. 1989. Insight without awareness: on the interaction of verbalization, instruction and practice in a simulated process control task. *Q. J. Exp. Psychol.* 41A:553–78
Stadler, M. A. 1989. On learning complex procedural knowledge. *J. Exp. Psychol.: Learn., Mem., Cognit.* 15:1061–69
Stenning, K. 1992. Distinguishing conceptual and empirical issues about mental models. In *Models in the Mind,* ed. Y. Rogers, A. Rutherford, P. Biddy. San Diego: Academic
Stenning, K., Levy, J. 1988. Knowledge-rich solutions to the binding problem: a simulation of some human computational mechanisms. *Knowledge Based Systems* 1:143–52

Stenning, K., Oaksford, M. 1993. Rational reasoning and human implementations of logics. In *Rationality,* ed. K. I. Manktelow, D. E. Over. London: Routledge. In Press

Stenning, K., Oberlander, J. 1992. A cognitive theory of graphical and linguistic reasoning: logic and implementation. Res. Pap. HCRC/RP-20, Hum. Commun. Res. Cent., Univ. Edinburgh

Stenning, K., Shepherd, M., Levy, J. 1988. On the construction of representations for individuals from descriptions in text. *Lang. Cognit. Proc.* 2:129–64

Stevenson, H. W., Stigler, J. W. 1992. *The Learning Gap: Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education.* New York: Summit

Sweller, J. 1988. Cognitive load during problem solving: effects on learning. *Cognit. Sci.* 12:257–86

Sweller, J., Mawer, R., Ward, M. 1983. Development of expertise in mathematical problem solving. *J. Exp. Psychol.: Gen.* 112:639–61

Thagard, P. 1989. Explanatory coherence. *Behav. Brain Sci.* 12:435–67

Thagard, P. 1992. *Conceptual Revolutions.* Princeton: Princeton Univ. Press

Thagard, P., Holyoak, K. J., Nelson, G., Gochfeld, D. 1990. Analog retrieval by constraint satisfaction. *Artif. Intell.* 46:259–310

Tversky, A. 1977. Features of similarity. *Psychol. Rev.* 84:327–52

Tversky, A., Kahneman, D. 1980. Causal schemas in judgments under uncertainty. In *Progress in Social Psychology,* pp. 49–77. Hillsdale, NJ: Erlbaum

Tversky, A., Kahneman, K. 1980. Causal schemas in judgments under uncertainty. See Tversky & Kahneman 1980, pp. 49–72

Tversky, A., Kahneman, K. 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90:293–315

Tweney, R. D. 1990. Five questions for computationalists. In *Computational Models of Scientific Discovery and Theory Formation,* ed. J. Shrager, P. Langley, pp. 471–84. San Mateo, CA: Morgan Kaufmann

Vera, A. H., Simon, H. A. 1993. Situated action: a symbolic interpretation. *Cognit. Sci.* In press

von der Malsburg, C. 1981. The correlation theory of brain function. Internal Rep. 81–2. Dep. Neurobiol., Max-Planck-Inst. Biophys. Chem., Göttingen

Waldmann, M. R., Holyoak, K. J. 1992. Predictive and diagnostic learning within causal models: asymmetries in cue competition. *J. Exp. Psychol.: Gen.* 121:222–36

Wason, P. C. 1966. Reasoning. In *New Horizons in Psychology,* ed. B. M. Foss, Vol. 1. Harmondsworth, UK: Penguin

Wellman, H. M., Gelman, S. A. 1992. Cognitive development: foundational theories of core domains. *Annu. Rev. Psychol.* 43:337–75

Wenger, E. 1987. *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge.* Los Altos, CA: Morgan Kaufmann

Wertheimer, M. 1945. *Productive Thinking.* New York: Harper Row

Wharton, C. M., Holyoak, K. J., Downing, P. E., Lange, T. E., Wickens, T. D. 1992. The story with reminding: Memory retrieval is influenced by analogical similarity. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society,* pp. 588–93. Hillsdale, NJ: Erlbaum

Willingham, D. B., Nissen, M. J., Bullemer, P. 1989. On the development of procedural knowledge. *J. Exp. Psychol.: Learn., Mem., Cognit.* 15:1047–60

Wimsatt, W. C. 1986. Developmental constraints, generative entrenchment, and the innate-acquired distinction. In *Integrating Scientific Disciplines,* ed. W. Bechtel, pp. 185–208. Dordrecht: Martinus-Nijhoff

Zukier, H., Pepitone, A. 1984. Social roles and strategies in predicting: some determinants of the use of base rate information. *J. Pers. Soc. Psychol.* 47:349–60