

The heuristic-analytic theory of reasoning: Extension and evaluation

JONATHAN ST. B. T. EVANS
University of Plymouth, Plymouth, England

An extensively revised heuristic-analytic theory of reasoning is presented incorporating three principles of hypothetical thinking. The theory assumes that reasoning and judgment are facilitated by the formation of epistemic mental models that are generated one at a time (singularity principle) by pre-conscious heuristic processes that contextualize problems in such a way as to maximize relevance to current goals (relevance principle). Analytic processes evaluate these models but tend to accept them unless there is good reason to reject them (satisficing principle). At a minimum, analytic processing of models is required so as to generate inferences or judgments relevant to the task instructions, but more active intervention may result in modification or replacement of default models generated by the heuristic system. Evidence for this theory is provided by a review of a wide range of literature on thinking and reasoning.

In recent years, a number of researchers in the psychology of thinking and reasoning have advocated *dual-process* theories of cognition (Evans, 2003). It has been suggested that people's behavior reflects the operation of two distinct thinking systems. Moreover, these systems interact in a complex way and often appear to be competing to control behavior. Such theories have become increasingly popular and are now being applied in the field of judgment and decision making (Gilovich & Griffin, 2002; Hammond, 1966; Kahneman & Frederick, 2002), as well as in the psychology of reasoning (Evans & Over, 1996; Sloman, 1996; Stanovich, 1999, 2004). In a largely separated but conceptually connected literature, there is also considerable interest in dual-process models of social cognition (Chaiken & Trope, 1999; Hassin, Uleman, & Bargh, 2005). However, the term *dual process* is a label used for historical reasons that disguises an important distinction, that between dual processes and dual systems. Stronger forms of the theory (e.g., Evans & Over, 1996; Reber, 1993; Stanovich, 2004) propose that the dual processes are rooted in two distinct cognitive systems that have sharply differing evolutionary histories and neurological substrates (see also Goel, 2005).

Dual-system theories are, by necessity, broad scheme and not easily applied to specific cognitive tasks. More-

over, they come with a lot of theoretical baggage that not all authors interested in the dual-process distinction will wish to endorse. For this reason, it seems useful to elaborate a more specific dual-process account of reasoning at an intermediate level of abstraction, from which models of particular tasks can easily be developed. I propose to do this here by major revision and extension of an earlier dual-process account, generally known as the heuristic-analytic theory of reasoning (Evans, 1984, 1989). It is noticeable that a number of contemporary authors prefer to refer to the heuristic-analytic distinction when discussing particular experimental findings on reasoning tasks, rather than the more broadly based dual-system theories (Ball, Lucas, Miles, & Gale, 2003; Klaczynski, 2001; Kokis, MacPherson, Toplak, West, & Stanovich, 2002; Roberts & Newton, 2002; Schroyens, Schaeken, Fias, & d'Ydewalle, 2000). However, in both my own writing and that of other authors, a number of revisions and extensions of the original theory seem to be present by implication.

The heuristic-analytic theory of reasoning as first published (Evans, 1984, 1989) was based on a quite simple idea. The theory was designed to explain the prevalence of cognitive biases in reasoning tasks and the puzzling fact that logical competence demonstrated on one task often failed to be exhibited on another (see Evans, 1989, for a discussion of relevant examples). The heuristic-analytic theory proposed that two kinds of cognitive processes were involved: *heuristic* processes, which generated selective representations of problem content, and *analytic* processes, which derived inferences or judgments from these representations. Biases were accounted for by the proposal that logically relevant information might be omitted or logically irrelevant information included at the heuristic stage. Since analytic reasoning could be applied only to these heuristically formed representations, biases could result. The theory was presented as a simple two-stage sequential model (see Figure 1).

This work was supported by Research Fellowship Award RES-000-27-0184 to the author from the Economic and Social Research Council of the United Kingdom. The author thanks Simon Handley and David Over for their assistance in the development of the principles of hypothetical thinking incorporated within this article. The final version also benefited from critical readings by Shira Elqayam, David Over, Keith Stenning, Christoph Klauer, and an anonymous referee of an earlier draft. All correspondence concerning this article should be addressed to J. St. B. T. Evans, Centre for Thinking and Language, School of Psychology, University of Plymouth, Plymouth PL4 8AA, England (e-mail: jevans@plymouth.ac.uk).

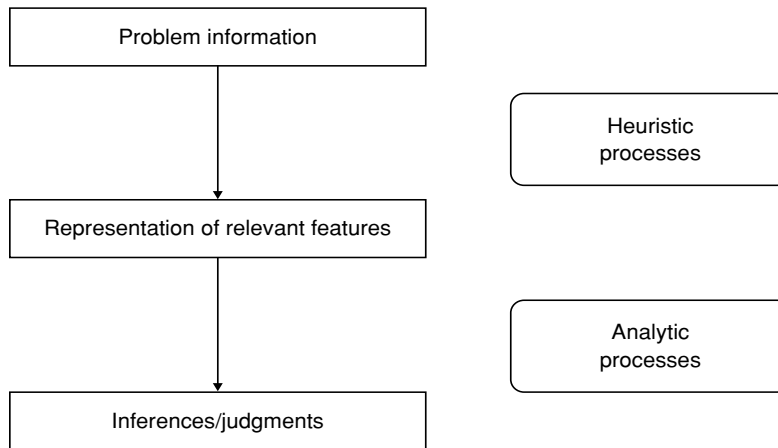


Figure 1. The original heuristic-analytic theory. From *Bias in Human Reasoning: Causes and Consequences* (p. 25), by J. St. B. T. Evans, 1989. London: Erlbaum. Copyright 1989 by Lawrence Erlbaum Associates. Reprinted with permission.

In the revised theory, presented here, the heuristic-analytic terminology is retained, with an attempt to define more precisely the nature of the interaction between the two processes and to assist in the generation of experimental predictions about particular reasoning tasks. At the same time, assumptions about dual systems are kept to a minimum. In this article, I will also show, by a review of recent literature, that there is already considerable empirical support available for this extended heuristic-analytic theory. I shall do this by reference to work on hypothesis testing, mental simulation, deductive reasoning, the Wason selection task, and statistical inference.

The Extended Heuristic-Analytic Account

The present account draws heavily on the theory of hypothetical thinking put forward by Evans, Over, and Handley (2003) in an attempt to gain greater understanding of how the analytic (or explicit) system works and how it interacts with the heuristic (or implicit) system. Evans, Over, and Handley (2003) were attempting to advance in more specific terms the idea proposed by Evans and Over (1996) that the analytic system is involved whenever hypothetical thought is required. Hypothetical thinking involves the imagination of possibilities that go beyond the representation of factual knowledge about the world. Examples include hypothesis testing, forecasting, consequential decision making, and (on certain assumptions) deductive reasoning.

Evans, Over, and Handley (2003) proposed three principles of hypothetical thinking that are shown in Table 1,

together with a processing model illustrated in Figure 2. It was proposed (*singularity principle*) that people construct only one mental model at a time with which to represent a hypothetical situation. This model is pragmatically cued to be the most relevant in the context (*relevance principle*). By default, but by no means by necessity, this model will represent the most probable or believable state of affairs. Finally, such representations are subject to explicit (i.e., analytic) evaluation, which complies with a *satisficing principle*. The singularity principle derives from the consideration that the analytic system is required for hypothetical thinking but has a sharply limited processing capacity. The other two principles reflect the operation of the two systems. The relevance principle is rooted in the heuristic system, whose purpose is to deliver content for analytic processing reflecting relevant knowledge and belief. The satisficing principle reflects a fundamental *bias* in the analytic system to work with the representation it has unless there is good reason to give it up. The principle is quite rational, however, given that, realistically, decisions have to be made without endless analysis of possibilities.

None of these three principles is novel in itself, although I am not aware of any other theory that combines them in the way proposed here. The singularity principle has been expressed, for example, by authors studying *pseudodiagnostic* reasoning as the idea that people can think about only one hypothesis at a time (Mynatt, Doherty, & Dragan, 1993) and by researchers in the mental models tradition who have described *focusing* effects in which people construct only one mental model at a time (Legrenzi, Girotto,

Table 1
The Three Principles of Hypothetical Thinking Proposed by Evans, Over, and Handley (2003)

The singularity principle	People consider a single hypothetical possibility, or mental model, at one time.
The relevance principle	People consider the model which is most relevant (generally the most plausible or probable) in the current context.
The satisficing principle	Models are evaluated with reference to the current goals and accepted if satisfactory.

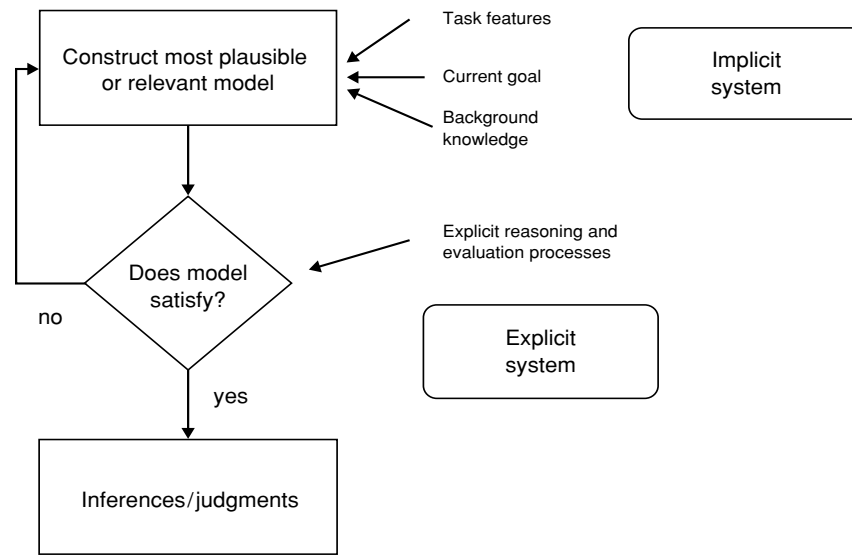


Figure 2. The hypothetical thinking model of Evans, Over, and Handley (2003). From *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (p. 5), by D. Handman and L. Maachi (Eds.), 2003, Chichester, U.K.: Wiley. Copyright 2003 by John Wiley & Sons Ltd. Reprinted with permission.

& Johnson-Laird, 1993). However, mental model theorists are not consistent on this point. For example, the model theory of propositional reasoning involves the idea that people may sometimes initially represent a statement with two explicit mental models, plus having the ability to flesh out full sets of as many as three models (Johnson-Laird & Byrne, 2002). In her theory of counterfactual thinking, Byrne (2005) repeatedly suggested that people engage in such thought for situations that lead them to consider *two* possibilities, rather than one. I will comment on this apparent conflict with the singularity principle later in this article.

The concept of satisficing derives from the bounded rationality approach of Simon (1982), expressing the idea that we could not practically aim to optimize choices in the real world, since we would remain forever lost in thought. The principle of rational choice prescribed by decision theory cannot be applied in an unbounded way to real-world decision making, since, for example, there is no way to know how far ahead in time people should grow their decision trees before folding them back to justify a choice (see Evans & Over, 1996, chap. 2). The notion of bounded rationality has been applied to human reasoning by Oaksford and Chater (e.g., 1995), who have argued that logic fails as a model of real-world reasoning that contains a nontrivial number of premises, since consistency checking involves a combinatorial explosion. Bounded rationality has also been proposed by those who have argued that we possess very simple but effective heuristics for making us intelligent (Gigerenzer & Todd, 1999). However, the use of the term *satisficing principle* here is more specific. It combines with the singularity principle to convey the idea that in hypothetical thinking, we consider one hypothesis at a time and maintain it until we find a good

reason to give it up. Note that whereas the heuristic system generally supplies hypotheses, it is the task of the analytic system critically to evaluate them and, if need be, modify or replace them.

The relevance principle concerns the generation of mental models and hypotheses by the heuristic system. It refers to the powerful tendency to contextualize all problems with reference to prior knowledge elicited by contextual cues and the current goals that are being pursued. This has been described as the *fundamental computational bias* by Stanovich (1999), although the term *bias* should certainly not be taken here in a pejorative sense. Given the notorious *frame problem* of artificial intelligence, we might describe the fundamental computational bias in computers as the *failure* to contextualize problems. What Stanovich (1999) is getting at is the fact that we need, in a modern technological society, to be capable also of abstract, decontextualized reasoning, which he believes the analytic system can achieve. Note that the relevance principle contrasts with the *principle of truth* in the mental model theory (Johnson-Laird & Byrne, 2002), in which it is proposed that people represent only true possibilities. By default, I assume that people represent what is believable or plausible (*true* is too strong a term) but also that this default can be altered according to context. Our attention can easily be focused on hypotheses that are improbable (buying health insurance to cover emergencies on a particular vacation) or most improbable (thinking about the consequences of life being discovered on Mars).

The heuristic-analytic theory does not offer an original or profound solution to the problem of *how* relevant knowledge is delivered by the heuristic system. However, in our proposals about mental representations, we have drawn on the notion that implicatures may be added to our

mental models (Evans & Over, 2004). Within the field of pragmatics and communication, I am generally attracted to the relevance theory put forward by Sperber and Wilson (1995) and, also, to the way in which it has been applied to explicit reasoning tasks (Sperber, Cara, & Girotto, 1995). Sperber’s scheme agrees with the present one in that most of the inferencing that undoubtedly is involved in the determination of relevance in discourse goes on at an automatic and preconscious level.

The term *mental model* has, I believe, at least three different meanings. One important idea is that of mental *simulation* (Gentner & Stevens, 1983). We may perform mental simulations in order to generate forecasts or to plan future activities (Kahneman & Tversky, 1982b), and we may also need to simulate possibilities, in order to decide whether we believe conditional statements (Evans & Over, 2004), or to engage in counterfactual thinking about what might have been (Byrne, 2005; Mandel, Hilton, & Catellani, 2005; Roese, 1997). Mental simulations are one of the main ways in which the analytic system evaluates current hypotheses but are not the mental models referred to in Figure 3. We use the term here, as do Johnson-Laird and his collaborators, to refer to a mental representation of a hypothetical state of affairs. However, what Johnson-Laird (1983; Johnson-Laird & Byrne, 1991) means by a mental model is often no more than a structural analogue of situations in the world. I call such mental models *semantic*, since they have truth-verifiable meaning. The term *mental model* in Figures 2 and 3 refers to neither a mental simulation nor a semantic mental model. Rather, these models are *epistemic*, representing states of belief and knowledge. The difference is that semantic models can only represent states of the world, whereas epistemic models can additionally encode propositional attitudes

toward those world states—in philosophical terms, their intentionality.

Consider, for example, the future possibility that global warming continues and London is flooded. A semantic model of this state of affairs, as in Johnson-Laird’s theory, can represent this only as a possible state of affairs:

Global warming continues London is flooded.

Although it is evident that one would have degrees of belief about this possibility, they cannot be represented in Johnson-Laird’s system, since he has no means to deal with nonextensional probabilities (Girotto & Johnson-Laird, 2004; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999; Over, 2004). An epistemic mental model might, however, look like this:

Global warming continues \rightarrow (.30) London is flooded.

This model encodes a degree of conditional belief (see Evans & Over, 2004, chap. 8) that one event will lead to the other. The model may go further and represent the fact that this belief is rooted in a causal linkage. (The discussion of causal mental models is beyond the scope of this article, but see Sloman, 2005.) In the case of hypothetical thinking, where suppositions about possible world states are concerned, an epistemic model crucially also encodes a further propositional attitude—namely, that the model represents a hypothetical, rather than an actual, state of affairs. I have argued elsewhere that the need for explicit representation of hypotheses is the reason that analytic thinking is invariably involved in hypothetical thinking (Evans & Over, 1999).

In the case of Evans, Over, and Handley’s (2003) theory (Figure 2) and the revised heuristic-analytic theory presented here (Figure 3), mental models should be taken to

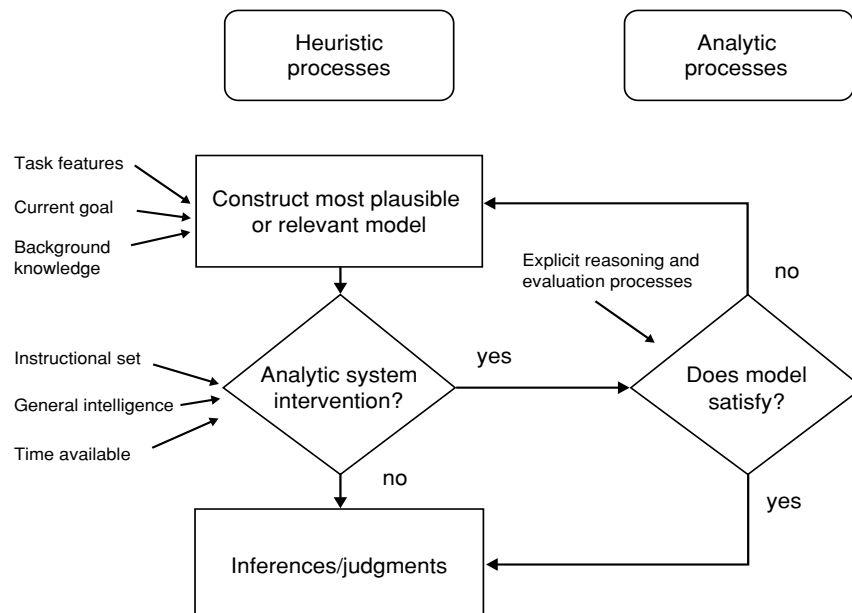


Figure 3. The revised and extended heuristic-analytic theory.

be not only epistemic, but also explicitly hypothetical or suppositional in nature. In all the applications considered in this article, people are representing hypothetical possibilities. Since Evans, Over, and Handley (2003) were specifically discussing hypothetical thinking, they considered only situations in which analytic processing was involved in controlling behavior. However, we know that heuristic processes can generate responses with little or no intervention by analytic processes. Indeed, a strong feature of the writing on contemporary dual-systems theory is the emphasis on the ability of the analytic system to inhibit default heuristic responses, a key concept in the explanation of why higher IQ individuals more often come up with normatively correct solutions to reasoning and judgment problems (Stanovich, 1999). Hence, the proposed extended heuristic-analytic theory appears as is shown in Figure 3, explicitly representing the fact that the analytic system may or may not intervene. The more theoretically neutral terms *heuristic/analytic* have also been substituted for the original terms *implicit/explicit*.

What do we know about the analytic thinking system and its operation? It appears to be slow and sequential in nature, controlled rather than automatic, and responsive to verbal instructions (Evans & Over, 1996; Stanovich, 1999). In contrast with implicit and automatic forms of cognitive processes, it also appears to be capable of domain-general reasoning. Also, in contrast with implicit forms of cognition, analytic thinking is undoubtedly linked to general cognitive ability, as has been proposed by a number of authors, including Reber (1993) and Stanovich (1999). Individual differences in working memory capacity, reasoning ability, and general intelligence scores are all very closely intercorrelated (Kyllonen & Christal, 1990).

Those who propose that there is a logic built into the mind (Braine & O'Brien, 1998a; Rips, 1994; Sperber, 2000) would doubtless associate this with the analytic system. Although I do not wish to engage in debate in this article about whether there is some form of mental logic, it is clear that the analytic system would have to be very much more than this. First, those of higher general intelligence, in whom the system seems to function more effectively, can provide normative solutions to a wide range of reasoning and decision-making problems whose solutions are not provided simply by a natural deduction system (Stanovich, 1999; Stanovich & West, 2000). Second, there are clearly links between proposals made by dual-process theorists and by working memory researchers that go beyond capacity considerations and involve such notions as inhibition and executive control. It is specifically proposed, for example, that higher ability people cannot just reason more accurately, but also can inhibit belief-based responding when instructed to reason logically (Stanovich & West, 1997). Of course, you might expect higher accuracy in more able groups simply because they are better reasoners, but there are reasons to believe that inhibition of belief bias is a separate facility. For example, the ability to resolve belief–logic conflict problems in favor of logic declines sharply with age (Gilinsky & Judd, 1994). Also, most participants, regardless of abil-

ity, will tend to give belief-based responses to reasoning tasks unless specifically instructed to reason logically (see Evans & Over, 2004). Finally, in neural-imaging studies, when belief–logic conflict is resolved in favor of logic, brain areas in the prefrontal cortex associated with executive control studies are recruited, whereas only language areas are activated during abstract deductive reasoning (Goel, 2005; Goel & Dolan, 2003). We should not, however, assume that the analytic system is restricted to reasoning in a way that is abstract and decontextualized. Recent evidence suggests that slow analytic processes may compete with fast heuristic processes within contextualized reasoning (Verschueren, Schaeken, & d'Ydewalle, 2005a, 2005b).

A key issue in dual-process theories is whether systems are parallel, sequential, or interactive (see Evans & Over, 1996). I do not subscribe to the view that heuristic and analytic processes are parallel thinking styles that can be adopted at will, in the manner of the proposed distinctions between *rational* and *experiential* thinking (Epstein & Pacini, 1999) or between *holistic* and *analytic* thinking styles that may be culturally determined (Nisbett, Peng, Choi, & Norenzayan, 2001). In general thinking, the two systems are interdependent, since preconscious heuristic (or pragmatic) processes supply content continuously to consciousness for analytic processing. However, I cannot retain the simple sequential structure of the original heuristic-analytic theory either. From a functional point of view, heuristic and analytic processes often seem to *compete* for control of behavior, as has been the case in studies of belief bias reviewed later in this article.

The architectural solution that I am proposing is that shown in Figure 3. Although heuristic processes may still bias and shape analytic thinking by the nature of the contextualized representations they generate, they do not compete as parallel processes. Rather, what is proposed is that heuristic process often cue *default* mental models that imply—with only shallow analytic processing of the task requirements—default responses, inferences, or decisions. Analytic processes may or may not *intervene* in order to revise or replace such default models and to inhibit default heuristic responding. The factors known to influence the likelihood of such intervention are shown in Figure 3. These include two that have already been mentioned: cognitive ability (or working memory capacity) and the use of instructions requiring abstract or logical reasoning. Another is the time available for this more effortful and reflective form of thinking. There are several studies in the literature showing that when participants are required to respond quickly to reasoning problems, default heuristics dominate responding. For example, on the Wason selection task (see below), *matching bias* is accentuated (Roberts & Newton, 2002) when participants are allowed only a short time to respond. The same manipulation has been shown to make belief bias more dominant in syllogistic reasoning (Evans & Curtis-Holmes, 2005) and is associated with more frequent endorsement of fallacious inferences in abstract conditional reasoning (Schroyens, Schaeken, & Handley, 2003).

The characteristics that determine analytic system intervention, other than cognitive ability, are *dispositional*. People may *choose* to engage in effortful analytic thinking because they are inclined to do so by strong deductive reasoning instructions (Evans, 2002) or, perhaps, because they have personal motivation. It is a commonplace observation that people spend much more time and effort thinking about important decisions that carry strong implications for their personal goals, whereas less important judgments may be made quickly and intuitively. Equally, when people are limited by cognitive ability or the available time, they may fail to apply this form of reasoning. This dispositional notion also allows us to reconcile the present architectural proposals about heuristic and analytic systems with the undoubted evidence of individual differences in thinking styles. Intuitive thinkers, for example, may be predisposed by personality or by cultural context to accept uncritically default judgments that are generated heuristically, whereas analytic thinkers may be more inclined to check them out with explicit reasoning. This dispositional aspect of the analytic system also provides encouragement to those who believe that our educational systems can and should encourage people to think in a more “rational,” abstract, and decontextualized manner (Baron, 1985; Stanovich, 1999).

The revised heuristic-analytic theory can continue to account for some cognitive biases in a way similar to that of the original theory (Evans, 1989). That is, biases can arise because the heuristic system fails to represent logically relevant features of the problem or represents features that are logically irrelevant to the problem. The evidence suggests that such heuristically generated biases can be inhibited, at least to some extent, by analytic system intervention, which is now proposed to include the ability to reset default epistemic mental models. However, the analytic system is also prone to biases of its own, especially due to the operation of the satisficing principle. I will demonstrate in the literature review that follows that a key feature of hypothetical thinking is the tendency of the analytic system to hold on to representations that are merely *good enough*, leading to such characteristic features as the endorsement of fallacious inferences in deductive reasoning and apparent confirmation biases in hypothesis testing. The perceived role of analytic, as well as heuristic, processes in cognitive biases is one of the major consequences of the revised theory presented here.

To understand how the satisficing principle can operate in decision making, consider the case of deciding what to do with a day's holiday. Rather than comparing options and maximizing choice, the idea is that pragmatically cued possibilities occur in order of maximum relevance. Say one has a friend visiting who is interested in historical buildings. One may think of a local historic site to visit that one knows he has not seen and will choose this unless it turns out that it is closed on the day in question. In that event, one generates another possibility for consideration. Perhaps he also likes country walks; one may think of a lovely place one has discovered since his last visit. That may become one's choice but will fail to satisfy if one then

learns of an adverse weather forecast. In that case, another pragmatically cued option is required, and so on (for evidence that people consider only one option at a time in everyday decision making, see Klein, 1999; Legrenzi et al., 1993). In the case of reasoning, the principle implies that defeasible inference is the norm in everyday reasoning, just as some authors have argued (Oaksford & Chater, 1998). That is to say, hypothetical thinking involves drawing inferences that are good enough—plausible in the context of current representations—but that are always subject to revision and rejection in the light of new evidence.

Hypothesis Testing

The psychological literature on thinking includes two broadly distinct paradigms for the study of hypothesis testing. One, deriving from the traditional literature on inductive reasoning and concept learning, deals with the learning of rules from exemplars. Participants are encouraged to discover the correct rule that classifies the cases given. In effect, they have to consider hypotheses in an all-or-none manner, deciding that they are true or false. I will consider studies of this kind in the present section. The other main paradigms for hypothesis testing provide people with statistical information and invite them to make probabilistic judgments or inferences. I will consider this kind of hypothesis testing later in the article.

Although working in one of the statistical traditions (the study of pseudodiagnostic reasoning) Mynatt et al. (1993) described the interaction of what I am calling the singularity and satisficing principles in a way that applies to hypothesis testing in general. With references to hypotheses, they state, “We propose that the number of objects that can be maintained and operated upon in working memory is one. . . . A corollary of this . . . assumption is that subjects will continue to test the hypothesis in working memory, unless prompted to change it.” Traditional studies of concept learning have shown precisely this kind of process. For example, in their classic studies of how people induce categorical rules from positive and negative instances, Bruner, Goodnow, and Austin (1956) reported that it was very common for participants to focus on one specific hypothesis at a time. When falsifying evidence was encountered, however, people gave up their hypotheses and formed new ones that were consistent, as far as they could remember, with the information already presented. Later studies in which simpler, one-dimensional rules that allowed hypotheses to be inferred without need for verbal protocols were used, confirmed the one-hypothesis-at-a-time strategy (Levine, 1966).

Mynatt et al. (1998) also added a principle of *confirmation bias* in hypothesis testing, which I believe to be redundant and conceptually mistaken. There has been a considerable debate in the literature about whether there is a general confirmation bias in human hypothesis testing (Evans, 1989; Klayman, 1995; Klayman & Ha, 1987; Poletiek, 2001). The term *confirmation bias* implies that people are, in some way, motivated to seek evidence that will confirm, rather than refute, their own hypotheses.

The idea was popularized by Peter Wason (Wason, 1960, 1968) in his early studies of the “2 4 6” problem. In this task, people are told that the experimenter has in mind a rule that classifies groups of three whole numbers, which I will describe as triads. They are also told that an example of a triad that conforms to this rule is 2 4 6. The participants are then invited to try to discover the rule by generating triads of their own. In each case, the experimenter provides feedback, saying whether or not the generated triad conforms to his rule. The participants normally keep track of their current hypotheses and are allowed to announce the rule only when they are highly confident that they have the correct answer. In the original paradigm, people announcing a wrong rule were allowed to continue with further testing of triads and to try again.

The experimenter’s rule was actually *any ascending sequence*. The biased exemplar 2 4 6 typically elicited an initial hypothesis that defined a subset of positive exemplars of the experimenter’s rule, such as *ascending with equal intervals* or *ascending even numbers*. This had the unfortunate consequence for the participants that any positive tests of their hypothesis would always lead to confirmatory feedback. Only a negative test could disconfirm a hypothesis such as *ascending with equal intervals*. For example, if the participant said “1 2 5,” this would conform to the experimenter’s rule even though it should be a negative instance according to the hypothesis. Participants make few such negative tests and, consequently, get stuck on incorrect hypotheses that receive continual confirmation. In consequence, they often announce several wrong rules, differing only in their wording. Wason (1960, 1968) claimed that this was evidence for a confirmation bias and argued that, in general, people were bad Popperians.

The problem with Wason’s (1960, 1968) argument is that all that is really being shown in these experiments is a positive testing bias, as was first demonstrated by Wetherick (1962), even though his critique was largely ignored in the literature prior to later reviews by Klayman and Ha (1987) and Evans (1989). Nothing in the repeated testing of positive cases demonstrates a confirmatory attitude. Interestingly, Poletiek (2001) has recently shown that an attitude of verification or falsification makes no essential difference to how people should test hypotheses: From either perspective, the best test is the one that is most diagnostic. A diagnostic test is one that makes a *risky prediction* (Popper, 1959) of something that is much more likely to hold given the hypothesis than given its alternatives. Psychological experiments that have strongly instructed participants to take a falsification approach to the 2 4 6 task have, in fact, had little effect in improving performance on the problem (Poletiek, 1996; Tweney, Doherty, Warner, et al., 1980).

Does the 2 4 6 task show evidence of cognitive bias at all? Klayman and Ha (1987) showed that in most real-world and scientific situations, positive testing will generally be more effective than negative testing, contrary to the special situation created by Wason’s 2 4 6 task. It is also arguable, in any case, that a Bayesian, rather than a Popperian, philosophy is more appropriate, especially in sta-

tistical sciences such as psychology (Evans, 2005a; Howson & Urbach, 1993). Even the so-called *exact* sciences have uncertainty in their experimental findings (due to possible errors in methodology, human observation, data analysis, etc.), which makes the strictly logical Popperian approach impractical.

What do real scientists do when confronted with falsifying evidence? In a recent study, Fugelsang, Stein, Green, and Dunbar (2004) carried out an observational study of a number of research groups working the field of molecular biology. Of 447 experiments studied, more than half (!) failed to conform with theoretical predictions. In only 31% of these cases did the scientists conclude that there was a fault with theory; in the other cases, they questioned the methodology. When (in 154 cases) the experiments were repeated with improved methodology, the majority (84) replicated the problematic finding. In the majority of these cases, the scientists were prepared to abandon or revise their hypotheses. This shows that real scientists behave more like Bayesians than like Popperians, revising belief in theories in a gradual, rather than an all-or-none, manner and factoring in uncertainty of both hypotheses and evidence.

Since the present theory provides for representation of hypotheses as *epistemic* mental models, people’s degree of belief, confidence, or uncertainty can be encoded directly into such models. Although the singularity principle would not support the normative Bayesian procedure of directly comparing alternative specific hypotheses, it would allow a favored hypothesis to be continually reevaluated against evidence until belief in it fell to a level at which it no longer satisfied. At this point, the hypothesis would be modified or replaced. The theory can also give an account of positive testing bias without recourse to an additional principle of confirmation bias, as proposed by Mynatt et al. (1993). In constructing mental simulations to evaluate suppositions, the analytic system continues to be influenced in the construction of content by the heuristic system, with its built-in relevance principle. As was stated earlier, this will normally lead to the generation of content that is plausible or probable, given the supposition or hypothesis under current consideration. It is natural for us to construct plausible scenarios of how the world would be (rather than not be) if our supposition holds.

The singularity principle implies that any task requiring simultaneous consideration of alternative hypotheses will be very difficult. There is evidence that thinking about disjunctive problems is very hard for people, especially when the disjunction is exclusive, indicating that either *p* or else *q* is the case, but not both.

For example, Bruner et al. (1956) extended their studies of concept learning to such exclusively disjunctive rules. These tasks were associated with very high error rates and a common intuitive error: People found it very hard to learn exclusive disjunctive rules from exemplars, often falsely assuming that they must contain common elements. A very similar pattern of results has been found with Wason’s THOG problem (Wason & Brooks, 1979), which requires consideration of mutually exclusive hy-

potheses for its solution. In the THOG problem, one is told that a figure conforms to the experimenter's rule if it has either the color or the shape that he or she has written down, *but not both*. One is then shown the following shapes:

white diamond white circle black diamond black circle.

One is told that the white diamond is a THOG and is asked what can be inferred about the other figures. The logically correct answer, which few participants find, is that the black circle must be a THOG and that the other two figures *cannot* be THOGs. The task requires consideration of two hypotheses for its solution. If the experimenter wrote down *white* he or she must have written *circle*, or the white diamond could not be a THOG (no THOG can have both the color and the shape written down). Similarly, if he or she wrote *diamond*, he or she must have written *black*. In either case, however, the black circle must also be a THOG. If we now consider the white circle, it cannot be a THOG, since if *white and circle* was written down, it has both features, and if *black and diamond* was written down, it has neither feature. Similar reasoning leads to the conclusion that the black diamond cannot be a THOG either. What makes the problem doubly fiendish is that each of the alternative rules is itself an exclusive disjunctive classifying together cases with no common elements.

A different kind of problem that involves consideration of mutually exclusive hypotheses is known as *meta-deduction* and was first introduced to the psychological literature by Rips (1989), although it has since received attention mostly by those working in the mental model tradition (e.g., Byrne & Handley, 1997). These problems are normally set in the land of knights and knaves, where knights always tell the truth and knaves always lie. They look similar, however, and can be distinguished only by what they say. These problems are difficult, since they have an inherently disjunctive nature. When you meet someone in this land and evaluate his utterances, you must consider both a knight and a knave hypothesis. Suppose you meet two such creatures, knowing that one is a knight and one a knave but not which is which. One says to you, "I am a knight and he is a knave." What could you conclude? Well, the speaker may be a knight, in which case he tells the truth, so he is indeed a knight and his friend is a knave. Now suppose that the speaker is a knave. He always lies, so he will describe himself as a knight. When he calls his friend a knave, however, he is also lying, as in this case his friend is a knight. Hence, you are no better off than you were to start with! It is likely, however, that many people will conclude that the speaker is a knight, in line with what Elqayam (2006) calls the *collapse illusion*. The collapse illusion arises because participants assume that the speaker is a knight unless they have evidence to the contrary. As Elqayam points out, this finding is fully consistent with the principles of hypothetical-thinking theory summarized in Table 1. The evidence indicates that people first try the knight hypothesis, and if it satisfies, they stick with it. Similar evidence of satisficing in suppositional

reasoning, on a different task, has been reported by Handley and Evans (2000).

In this section, I have considered a wide variety of tasks that involve what we might call logical (as opposed to statistical) hypothesis testing. That is to say, in all cases, people have to decide whether hypotheses are true or false, rather than to assign them probabilities. These studies provide strong evidence for the singularity and satisficing principles. In studies of concept learning and inductive reasoning, it is clear that people formulate one hypothesis at a time and test it positively until or unless evidence is encountered that forces them to give it up. Real scientists appear to eschew the logicism of Popper in favor of something more closely approximating a Bayesian philosophy, in which hypotheses are held with degrees of subjective probability, rather than being considered true until proven false. Studies that require people to think disjunctively, in defiance of the singularity principle, are some of the most notoriously difficult to be found in the literature, including Wason's THOG problem and the metaductive problem of knights and knaves. On such a task, the evidence suggests that people try to work with single mental models, which leads them into characteristic intuitive errors.

Mental Simulations and the Suppositional Conditional

Inferences, decisions, and judgments can be based on epistemic mental models that are directly retrieved from semantic memory by the heuristic system. This will frequently be the case when we are dealing with situations that are familiar and about which we have well-established beliefs. As was noted earlier, there is a powerful tendency for the heuristic system to contextualize problems and decisions with respect to prior knowledge and to provide default responses. However, when we are dealing with novel problems or decisions that are of particular significance for us, we may engage in more protracted mental simulations. Such simulations strongly involve the analytic system and reflect its characteristics as a slow, sequential, and capacity-limited system. Simulations are also cognitively expensive, since they tie up central working memory resources.

Mental simulations are required whenever we need to foresee the future. As Kahneman and Tversky (1982b) have pointed out, we use mental simulations to forecast future events, to judge the probability of uncertain events, to assess causal relationships, and to engage in counterfactual thinking that may "undo" past actions and events. The last of these topics has been the focus of much recent research, and we now know a good deal, at least, about the conditions under which people will engage in counterfactual thinking (Byrne, 2005; Mandel et al., 2005). As Kahneman and Tversky (1982b) also pointed out, the use of mental simulations is prone to cognitive biases. The major problem, in my view, is that we try to encode complex scenarios into single simulations. For example, people consistently underestimate how long it will take to perform some planned task, a phenomenon known as the *planning*

fallacy (Bueler, Griffin, & Ross, 2002). For some curious reason, people seem to prefer to compute such estimates by mental simulations than by the much more reliable method of considering how long similar tasks have taken in the past. Thus, for example, academics will chronically underestimate the time it takes to perform a familiar task, such as writing a research paper. This comes about because they simulate an idealized mental scenario in which they fail to take into account all of the problems and distractions that will actually slow down the process.

One of the most fascinating aspects of human cognition is our ability to entertain *suppositions*—that is, temporary beliefs that form the basis of a mental simulation of a possible scenario. It is essential that such suppositions be represented within epistemic mental models that encode their hypothetical nature. Suppositions must be *decoupled* from semantic memory (Cosmides & Tooby, 2000), so that we do not confuse them with actual beliefs about the real world. In my opinion, the use of the word *if* in everyday language is the primary linguistic trigger that we use to elicit mental simulations in our listeners (Evans, 2005b; Evans & Over, 2004). Until recently, psychological research on conditionals has not properly recognized the suppositional nature of *if*, which has been more clearly developed by the work of philosophical logicians (Bennett, 2003; Edgington, 1995).

What we term the suppositional theory of conditionals (Evans & Over, 2004; Evans, Over, & Handley, 2005) can be linked directly to the principles of hypothetical thinking. The key notion is that a statement of the form *if p then q* deals only in *p*-possibilities and invites the listener to engage in mental simulation in which some consequent event *q* is considered under the supposition of some antecedent condition *p*. In this regard, the suppositional conditional complies with what philosophers call the *Ramsey test* (Ramsey, 1931), although philosophers have not developed the psychological implications of this hypothesis. What is clear from the philosophical work is that the *material conditional* is most unsatisfactory as a description of the ordinary conditional of everyday discourse (Bennett, 2003; Edgington, 1995), sanctioning unacceptably paradoxical inferences, because the material conditional has exactly the same meaning as *not-p or q*. In our view, the popular psychological theory of conditionals based on mental model theory (Johnson-Laird & Byrne, 2002) unfortunately embodies the material conditional in its definition of the core meaning of the statement (see Evans & Over, 2004; Evans, Over, & Handley, 2005; for a response, see Schroyens & Schaeken, 2004). Nor is the mental model theory able to account for two key forms of evidence, reviewed below, that support the suppositional account: the assignment of defective truth tables to conditionals and the relation of belief in conditionals with the conditional probability, $P(q|p)$ (Evans, Over, & Handley, 2005).

The suppositional conditional invites you to consider just one supposition, *p* (singularity principle), and to construct a mental simulation whose relevance is determined

by the goal of evaluating *q* in that context. It follows that people have a *truth value gap* for *not-p* cases (Adams, 1975) and, hence, should have what is known as a *defective truth table* for a conditional statement (Wason, 1966). That is, when asked to evaluate the effect of different possibilities on the truth of a conditional statement, most people say that it will be true when we have *p* and *q* and false when we have *p* and *not-q*. However, contrary to the material conditional and consistent with the defective truth table, people most often describe *not-p* cases as *irrelevant* (rather than true; see Evans, Newstead, & Byrne, 1993, and Evans & Over, 2004, for reviews of the relevant studies).

What happens when people are asked to judge the probability or believability of a conditional sentence? They suppose that *p* is the case, introducing relevant knowledge such as pragmatic implicatures into their epistemic mental models (Evans & Over, 2004). The simulation also has a clear analytic system goal: the evaluation of *q* within this model. Consequently, we should expect that the subjective conditional probability of *q* given *p* will be the probability assigned to the conditional statement as a whole, something broadly confirmed in recent research on the judged probability of both abstract (Evans, Handley, & Over, 2003; Oberauer & Wilhelm, 2003) and thematic (Over, Hadjichristidis, Evans, Handley, & Sloman, in press) conditionals. The study of Over et al. concerned everyday causal conditionals referring to real-world events. In a typical experiment, participants were asked to judge the probability of such statements as

If the price of petrol increases then traffic congestion will improve.

Over et al. asked people in a separate task to judge the probability of the four truth table cases corresponding to such statements:

- The price of petrol increases and traffic congestion improves
- The price of petrol increases and traffic congestion does not improve
- The price of petrol does not increase and traffic congestion improves
- The price of petrol does not increase and traffic congestion does not improve.

In this *probabilistic truth table task*, participants assign a percentage probability to each possibility, making them add to 100%. From these ratings, we can compute various relevant probabilities. Multiple regression analyses showed that by far the strongest predictor of the probability of the conditional statement was $P(q|p)$, in line with the suppositional theory. Hence, people focus on the *p*-hypothesis—states of the world in which the petrol price increases—and then decide on the relative extent to which they believe that traffic congestion would or would not improve. If the relevant beliefs are not directly stored, analytic processes may be involved in simulating the above possibilities, as well as in the evaluation of the corresponding conditional sentence presented in a separate task. The key

prediction is that evaluating the conditional requires consideration only of the first two possibilities above, those in which the antecedent condition holds. Belief in these two possibilities is sufficient to determine the subjective conditional probability, $P(q|p)$, which is the probability generally assigned to the conditional statement.

As was mentioned earlier, researchers in the mental models tradition sometimes propose that people form representations with two or more (semantic) mental models. This may appear to contradict the singularity principle, but the two systems are not necessarily in conflict, since epistemic mental models are richer in format and can include added pragmatic implicatures (Evans & Over, 2004) that would require more than one semantic type of mental model to capture. A case in point is counterfactual conditionals, which Johnson-Laird and Byrne (2002) have treated differently from indicatives. Normally, they have proposed that the *initial* representation of a conditional involves only one explicit mental model. For example, the statement

If the Democrats are elected then taxes will go up

would have the models

Democrats elected	taxes go up
...	

where ... is an implicit mental model corresponding to the states of the world in which the Democrats are not elected—formally defined as true possibilities for Johnson-Laird and Byrne's conditional. However, the counterfactual conditional

If the Democrats had been elected then taxes would have gone up

is represented by *two* explicit models in their theory (Byrne, 2005; Johnson-Laird & Byrne, 2002):

<i>Fact:</i> Democrats were not elected	Taxes did not go up
<i>Counterfactual possibility:</i>	
Democrats were elected	Taxes did go up

As Evans and Over (2004) have pointed out, it is odd to describe the presumed actual state of the world here as a *fact*, since it is only pragmatically, and not logically, implied by the counterfactual conditional. Although such a pragmatic implicature would be available to a person thinking about this conditional, only *one* mental simulation would be suggested, in line with the singularity principle. This would be the simulation of the counterfactual possibility that taxes would have gone up if the Democrats had been elected. Hence, evaluation of a counterfactual conditional should involve exactly the same mental processes as those for an indicative. To test this, Over et al. (in press, Experiment 3) extended their methodology to the consideration of counterfactual conditionals. In this experiment, participants judged the probability of past tense counterfactuals that referred to causally linked events that might have occurred in the past 5 years. In the corresponding probabilistic truth table task, they were asked to judge the likelihood that the four relevant possibilities would have occurred subsequent to a point in time 5 years ear-

lier. The results of this experiment were indistinguishable from those run on indicative causal conditionals. People again set their belief in the conditional primarily on the basis of $P(q|p)$. For example, belief in the statement "If the Democrats had been elected then taxes would have gone up" would be determined only by beliefs concerning possibilities of the Democrats being elected within the past 5 years, taking no account of beliefs in the possibilities where they are not elected.

The evidence clearly supports the view that conditionals are processed by consideration of a single hypothesis (the p -possibility) that may require analytic involvement in a single mental simulation. Does research on conditionals provide evidence also of the satisficing principle? The answer is clearly yes. First, as with syllogistic reasoning, reviewed below, people endorse many fallacious conditional inferences, such as the affirmation of the consequent (AC): Given q , infer p . Some studies of university students have shown very high rates of fallacies when abstract materials have been used (see Evans, Clibbens, & Rood, 1995; Schroyens, Schaeken, & d'Ydewalle, 2001). The prevalence of fallacies suggests that people tend to use biconditional representations that go beyond the information given and satisfy by default. Fallacies can easily be blocked, however, by use of realistic content in which it is evident that p is sufficient but not necessary for q (see, e.g., Thompson, 1994, 2000). For example, given "if it is an elephant then it has a tail," few people would infer that the presence of a tail implies an elephant. They would be much more likely to make an AC, however, if *tusks* were substituted for *a tail*. My interpretation is that relevant knowledge cued pragmatically by the heuristic system ensures that the model underlying such inferences will no longer satisfy.

In general, the hypothetical thinking principles imply that ordinary reasoning should be *defeasible*, as has been argued by Oaksford and Chater (1993). That is to say, inferences may be made and then later withdrawn in light of new information, something that is beyond explanation in standard mental logic accounts. As was mentioned earlier, the satisficing principle is broadly in line with the bounded rationality approach to human reasoning advocated by these authors (Oaksford & Chater, 1998). In the case of conditional statements, people may decline to draw valid, as well as fallacious, inferences under some circumstances, including the apparently obvious modus ponens (MP) inference: Given p , conclude q . For example, people resist the MP when they disbelieve the conditional statement or indicate a degree of confidence in the conclusion related to their degree of belief in the conditional (George, 1995, 1997; Liu, Lo, & Wu, 1996; Stevenson & Over, 1995). As Byrne (1989) was the first to show, the MP can also be defeated by adding a second conditional that undermines the first by suggestion of a condition that could disable the link from p to q .

The Wason Selection Task

Many psychologists would argue that the Wason selection task (Wason, 1966) has been greatly overresearched

and is, in any case, a poor test of deductive reasoning. I do not disagree but, nevertheless, believe that it has provided a lot of interesting evidence about the nature of hypothetical thought (Evans & Over, 2004, chap. 5). It has certainly been the focus for much discussion of heuristic and analytic processes. In a typical version of the abstract and indicative version of the task, participants are told that a conditional statement applies to four cards and may be true or false. The cards are known to have a capital letter on one side and a single digit number on the other. The rule may be stated as

If there is an A on one side of the card, then there is a 3 on the other side of the card,

and the displayed values of the four cards as

A D 3 7.

The instruction is to choose those cards and only those cards that need to be turned over in order to decide whether the statement is true or false. Most participants choose A or A and 3. However, the generally agreed logically correct choice is A and 7, because only a card that has an A on one side and does not have a 3 on the other side would disprove the statement. Both the material and the suppositional theories of conditionals agree on this point. Whether the statement could be proved *true*, as opposed to false, as suggested by the instruction, is more questionable. Wason's (1966) assumption was presumably that the rule was true if not proved false, since it applied only to the four cards presented. However, this assumes a bivalent logic in which all propositions are true or false, a principle not accepted in the suppositional account. The D card, for example, can prove the conditional neither true nor false on the suppositional theory, since it is irrelevant to the statement.

If the abstract and indicative version of the selection task above requires analytic reasoning for its solution, then according to dual-process theory, ability to solve it should be related to general intelligence. In general, the evidence supports the view that solvers of abstract selection tasks are unusually high in general intelligence (Stanovich & West, 1998; see also Evans et al., 1993), although this may be harder to observe if the populations sampled are low in high-ability individuals (Newstead, Handley, Harley, Wright, & Farelly, 2004). Given the very low solution rates typically observed (10%–20%), does this mean that the majority of people do not engage in analytic reasoning on the task?

In fact, a matter of some debate has been whether selection task responses can be explained entirely by heuristic processes. Evans (1989) gave an account of selections that seemed to require operation only of two heuristics: an *if* heuristic and a *not*, or *matching*, heuristic (Evans, 1998). The *if* heuristic would cause people to attend to the A and ignore the D, since *if* powerfully induces a process of hypothetical thinking on the supposition that the antecedent holds (Evans & Over, 2004). The matching heuristic causes items named explicitly in the statement to be perceived as relevant. This would also favor A over D but also

3 over 7. When negated components are introduced into the conditional statement (as was first shown by Evans & Lynch, 1973), people tend to switch logical choices to maintain matching responses, a phenomenon known as *matching bias* (Evans, 1998). For example, if the rule is phrased as “If there is an A one side of the card then there is *not* a 3 on the other side of the card,” most participants will choose the A and 3 cards, which are now the logically correct choices. The choice of consequent cards seems almost entirely determined by matching, whereas antecedent card choices continue to reflect also a preference for true antecedent over false antecedent. Hence, an additive combination of *if* and matching heuristics seems adequate to account for typical data sets (see Evans & Over, 2004, chap. 5, for detailed analysis and argument).

In spite of appearances, I believe that for most participants, the analytic system is actively engaged on this task. First, there are good a priori reasons why it should be. The task is of an abstract nature, which may encourage activation of analytic reasoning as a factor additional to those shown in Figure 3. The instructions refer to truth and falsity and, by implication, necessity (choose *only* those cards . . .). The population tested—generally, undergraduate students—are also bright enough for more than the 10% or so who solve the abstract problem to be engaging in analytic reasoning. Certainly, the same population can provide much higher normative solution rates on other deductive reasoning tasks studied in the literature.

There are actually several forms of evidence suggesting that analytic reasoning *is* engaged on the abstract selection task. When verbal protocols are examined, participants make reference not only to the visible sides of the card, but also to possible values on their hidden sides (Evans, 1995; Wason & Evans, 1975) and refer to the effect these combinations have on the truth value of the statement. The next form of evidence comes from card inspection times. Evans (1996), using a mouse-pointing method, predicted and found that participants spent much longer considering cards that they would end up selecting than those that they would reject. This is consistent with the idea that nonselected cards are ignored because attention to them is not cued by the *if* or the matching heuristic. However, it is hard to explain why lengthy periods of consideration should be given to cards prior to their selection, unless people are trying to justify their selection by analytic reasoning. Although the mouse-pointing methodology has been criticized (Roberts, 1998), all the original predictions of Evans (1996) have been confirmed recently, using the superior method of eye movement tracking (Ball et al., 2003). For other recent evidence purporting to show analytic processing on the selection task, see Feeney and Handley (2000; Handley, Feeney, & Harper, 2002).

If analytic reasoning is involved in selection task choices, why do heuristics seem to account so strongly for the selections made? The answer lies in the satisficing principle and shows again how cognitive biases can occur in the analytic, as well as the heuristic, system. Most people (except those of very high cognitive ability) treat verification and falsification as though they were sym-

metrical. Participants will happily justify a choice of a matching card combination on the grounds that it will prove the rule true or prove it false, as was originally shown by Wason and Evans (1975). Since the standard instruction refers to discovering whether the statement is true or false, the analytic system satisfices (accepts the heuristically cued choice) whenever it can find a verification or falsification justification. In practice, this means that heuristically cued choices on the selection task will nearly always be accepted.

This discussion of the abstract selection task illustrates all the main features of the extended heuristic-analytic theory. Potential choices are typically restricted to those that are heuristically cued. The analytic system is engaged to scrutinize the choices but, due to a built-in cognitive bias in most people, will satisfice by accepting cases that could be true without confirming their logical necessity, hence explaining the common choice of the 3 card. Those very high in intelligence, however, can successfully inhibit the heuristically generated choices and pay attention to the 7 card. Consistent with this claim, a recent study in my own laboratory has shown that about 20% of participants with the highest cognitive ability are able largely to inhibit matching bias on a closely related reasoning problem (Evans, Handley, Neilens, & Over, 2005).

Syllogistic Reasoning and Belief Bias

Reasoning with categorical syllogisms has been extensively studied by experimental psychologists. Such syllogisms (dating from Aristotle's early logical system) relate three terms, which I will call A, B, and C. Syllogisms consist of two premises and a conclusion, all of which take one of four forms:

A	Universal affirmative	All A are B
E	Universal negative	No A are B
I	Particular affirmative	Some A are B
O	Particular negative	Some A are not B

The combination of forms used in the three statements determines the mood of the syllogism. For example, AAI would mean that both premises took the A form but the conclusion the I form. There are also four possible figures for syllogisms. Assuming that conclusions have the form A–C, the premises may reference the terms as A–B, B–C; A–B, C–B; B–A, B–C; or B–A, C–B. Hence, there are $4^3 = 64$ possible moods that can appear in each of four figures, making 256 logically distinct syllogisms. The great majority of these are logically invalid; that is, their conclusions do not necessarily follow from their premises.

Many experiments have been conducted on syllogistic reasoning, and a wide range of theoretical accounts have been proposed (see Evans et al., 1993, for a detailed review). The most comprehensive experimental study appears to be that of Evans, Handley, Harper, and Johnson-Laird (1999). To my knowledge, this is the only study to have presented all 256 possible syllogisms for evaluation. Moreover, this was done under two different instruction conditions. One group was asked to judge whether the con-

clusion was necessary, given the premises; the other group was asked to judge whether the conclusion was possible.

General findings with syllogistic reasoning studies are the following. First, people make many logical errors. In particular, they endorse the conclusions of many invalid syllogisms—also known as *fallacies*—despite instructions that normally require them to judge whether the conclusion necessarily follows. Second, they are apparently biased by the mood of the syllogism, preferring conclusions that are similar in mood to the premises (Woodworth & Sells, 1935) or that are less informative (Chater & Oaksford, 1999). They are also biased by the figure of the syllogism, preferring to draw conclusions that correspond with the order in which terms are mentioned in the premises (Johnson-Laird & Bara, 1984). However, there are numerous different theoretical accounts of these findings, some based on quite different principles (see Evans et al., 1993).

The original mental model account of syllogistic reasoning (Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991) was based on the idea that people form mental models of the premises that may or may not include a putative conclusion. It was argued that people search for counterexample cases—that is, models of the premises that do not include a putative conclusion—in order to validate the inference. Where there are multiple models consistent with the premises, people will be more prone to error than when there is only one. In particular, people may endorse fallacies if they fail to find a counterexample that exists. Recent research has, however, cast considerable doubt on whether people will normally search for counterexamples at all (Evans et al., 1999; Newstead, Handley, & Buck, 1999).

It is a feature of the revised heuristic-analytic theory that satisficing in the analytic system can account for some cognitive biases. We have already seen that the tendency to hold on to models that satisfy can account for some of the phenomena associated with hypothesis-testing behavior and the Wason selection task. Syllogistic reasoning performance illustrates a closely related bias in the analytic system. People will tend to accept conclusions that *could* be true, given their premises, but do not have to be true. I suggest that so fundamental is the satisficing principle in ordinary reasoning and decision making that the instruction to seek necessary conclusions (as in the standard deduction paradigm) is very hard for most participants to follow.

Evans et al. (1999), by giving all possible syllogisms for evaluation, discovered a new phenomenon. Conclusions to syllogisms can be necessary (valid inferences), possible, or impossible (the negation of the conclusion is a valid inference). Many fall into the *possible* category, where the conclusion could be true, given the premises, but need not be. In mental model terms, this means that there is at least one model of the premises that includes the conclusion and at least one that excludes it. What Evans et al. (1999) discovered was that there are two different classes of *possible* syllogisms. In one case, the conclu-

sion is nearly always endorsed; they called these *possible strong*. In the other case, conclusions are rarely endorsed, called *possible weak*. In fact, they showed that endorsement rates for possible strong syllogisms were as high as those for valid syllogisms and that rates for possible weak syllogisms were as low as those for impossible syllogisms (those whose conclusions are determinately false). This suggests that participants construct only *one* model of the premises. If this model includes the conclusion, they accept it (valid and possible strong cases), and if not, they reject it (impossible and possible weak cases).

This analysis fits well with the extended heuristic-analytic theory. If the first model considered excludes the conclusion, the argument is rejected straight away as invalid; clearly, the model does not satisfy. If the model includes the conclusion, it does satisfy, even though no check for counterexamples has taken place. This is our fundamental mode of hypothetical thinking: We accept a hypothesis (model) until there is reason to give it up. Although this accounts for the most striking finding on syllogistic reasoning (high rates of acceptance of fallacies), it can by no means account for all the errors and biases observed. There is good reason to think that there are a number of *response bias* effects. Choosing conclusions that match the mood or figure of the premises, for example, appears to be a response bias unaffected by analytic processing (but for a mental model account of figural bias, see Johnson-Laird & Bara, 1984).

The other main reason to suspect response bias arises from the study of how prior belief affects reasoning with syllogisms. The classic method for studying belief biases (Evans, Barston, & Pollard, 1983) involves presenting syllogisms that have conclusions that are either valid or invalid and are either believable or unbelievable. There are three typical findings: (1) People endorse more valid conclusions than they do invalid ones; (2) people endorse more believable conclusions than they do unbelievable ones; and (3) the two factors interact. The belief bias effect is usually significantly stronger for invalid than for valid syllogisms. The mental model theory of syllogistic reasoning accounts for the strong belief bias on invalid arguments by proposing that people are more motivated to search for counterexamples if they disbelieve the conclusion (Oakhill & Johnson-Laird, 1985; Oakhill, Johnson-Laird, & Garnham, 1989). However, the theory has great difficulty in accounting for the weaker, but still present, belief bias on valid arguments. Oakhill et al. suggested a *conclusion filter*: a mechanism external to the model theory and, in essence, a response bias.

Two recent studies of belief bias, conducted independently, came to very similar theoretical conclusions (Evans, Handley, & Harper, 2001; Klauer, Musch, & Naumer, 2000). Consistent with the heuristic-analytic theory, both sets of authors argued that people, in fact, construct only one model of the premises but that this model is motivated by the believability of the conclusion. That is, people attempt to find a model that includes or excludes the conclusion presented, depending on whether it is believable. This explains the major belief bias effect on invalid syllogisms,

where both kinds of models exist. However, like Oakhill et al. (1989), both sets of authors concluded that the main effect of belief, including that on valid syllogisms, must reflect some kind of response bias to simply accept believable and reject unbelievable conclusions. Evidence that such a bias is heuristic in origin comes from a recent study in which performance on the belief bias paradigm was compared under speeded and free-time instructions (Evans & Curtis-Holmes, 2005). Under speeded conditions, belief bias is significantly more marked, and the usual belief by logic interaction disappears.

In conclusion, research on syllogistic reasoning and belief bias effects is interpreted here in a way that is entirely compatible with the extended heuristic-analytic theory. Note that this account differs significantly from that of the mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). According to the latter, a fundamental principle is that people are deductively competent in understanding the basic semantic principle of validity—namely, that a conclusion is valid if there are no counterexamples to it. The evidence supports, in contrast, the satisficing principle of the heuristic-analytic theory: People generally accept conclusions that are consistent with a current model without validating them. The heuristic-analytic account further proposes that with pragmatically rich content, model generation is biased (in line with the relevance principle) to generate models that favor prior belief. The evidence suggests that people build single mental models of the premises that support believable conclusions but oppose unbelievable ones.

Statistical Inference

Intuitive statistical judgment and inference have been most famously studied within the *heuristics and biases* tradition of Amos Tversky and Danny Kahneman (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). The idea is that people assess uncertainty by use of simple heuristics that are often effective but can lead to cognitive biases. For example, we may judge frequency, in accordance with the *availability heuristic*, by basing our judgments on the ease with which examples come to mind (Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973). If our recollections are biased by the organization of our memory systems, this bias will be reflected in the judgment of probability made. Within the present framework, I propose that (1) the availability of relevant instances is determined by the heuristic system as it retrieves information from long-term memory in accord with the relevance principle and (2) the conversion of this information into a judgment of probability is an analytic process. As we have seen with reasoning research, such analytic involvement can be minimal. However, in principle, the analytic system could intervene actively, deciding that our memories are biased or that some explicit statistical reasoning should be applied instead.

Recently, Kahneman and Frederick (2002) adopted the dual-process framework and identified most of their heuristics as automatic processes in *System 1* (heuristic

system), which may be overridden by explicit reasoning in *System 2* (analytic system). Originally, some of the biases identified in this literature were portrayed in an extreme way, but these claims have not stood the test of time. For example, a bias to *ignore* sample size in statistical judgments (Kahneman & Tversky, 1972) was later reassessed as a tendency to *underweight* sample size (Evans, 1989; Kahneman & Tversky, 1982a). Similarly, a phenomenon widely reported at one time as the complete neglect of base rate information in Bayesian reasoning (Kahneman & Tversky, 1973) is seen, in the light of extensive subsequent research, to be a tendency to underweight base rate information (Koehler, 1996). Thus, biases of statistical intuition do appear to compete with a tendency to make normatively correct choices, in a manner parallel to the observations, in deductive reasoning research, that matching and belief heuristics compete with logically correct responses. The parallel is also observed with regard to cognitive ability, with more normative solutions being offered by high-ability participants in statistical, as well as logical, reasoning (Stanovich, 1999).

In Bayesian reasoning, people are required to combine prior probabilities with diagnostic information to generate posterior probabilities. As an example, consider the medical diagnosis problem of Casscells, Schoenberger, and Graboys (1978), given to medical students and staff at Harvard Medical School:

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?

The most common answer given to this question was 95%, whereas the correct answer is approximately 2%. It is much easier to understand why if the problem is presented as follows (adapted from Cosmides & Tooby, 1996):

One out of every 1,000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But out of 1,000 people who are perfectly healthy, 50 of them will test positive for the disease. Out of 1,000 Americans drawn at random, how many who test positive will have the disease?

In the second version (formally equivalent), we can easily construct a mental model that combines information about false positive rates (which affect the diagnosticity of the evidence) and base rates. We imagine 1,000 people, of whom 50 have false positive results and of whom 1 has the disease. So we judge that 1 in 50, or 2%, is the correct answer (actually, it is 1.96%; the false positive rate applies only to the 999 who do not have the disease). In the first version, people cannot combine the information and are inclined to think that since the test has an error rate of 5%, the chance of a correct identification (positive means disease present) is 95%—shallow and unfounded reasoning.

The original explanation of base rate neglect in terms of the representativeness heuristic (Kahneman & Tversky, 1973) was based on experiments in which the diagnostic

information had a perceived degree of similarity to the target category and not to the base rate. This does not apply to the Casscells et al. (1978) version, where base rate neglect still occurs. However, people do not necessarily give the 95% response when they go wrong on this problem. A finding that has received little attention in the literature is that people may also give the *base rate* as the answer, neglecting the diagnostic information (Cosmides & Tooby, 1996; Evans, Handley, Perham, Over, & Thompson, 2000). This suggests that the real difficulty lies in the ability to find a way to integrate the two pieces of information.

Why does the second version facilitate correct responding? At one time, it was claimed, on evolutionary grounds, that we have a built-in cognitive ability to process frequencies but not one-case probabilities (Gigerenzer, 1993). We now know that this is not right, since researchers have succeeded in developing hard frequency and easy probability versions of the problem (Evans et al., 2000; Girotto & Gonzalez, 2001; Sloman, Over, & Slovack, 2003). All this research suggests that what makes Bayesian inference easy are problems that provide direct cues to the *nested-set* relationships involved (Tversky & Kahneman, 1983), a position also incorporated into recent analyses in terms of ecologically valid *natural sampling* (Gigerenzer & Hoffrage, 1995; Hoffrage, Gigerenzer, Krauss, & Martignon, 2002).

It appears that heuristic processes cannot lead to correct integration of diagnostic and base rate information, and so Bayesian problems can only be solved analytically. This being the case, problem formats that cue construction of a single mental model that integrates the information in the form of nested sets appears to be critical. However, there is an apparent point of contrast in the base rate neglect literature and the studies of belief bias in reasoning reviewed earlier. The latter literature suggests that prior *belief* should have a large influence on people's thinking. The experimental literature on Bayesian reasoning, however, has almost always presented such tasks with prior probabilities as base rate *statistics*. What happens if real prior beliefs are used to supply the prior probabilities?

This issue was addressed in a recent study by Evans, Handley, Over, and Perham (2002). A series of experiments was run in which either diagnostic information was supplied by prior belief and prior probabilities by statistics or the other way around. Overall, in line with the literature, diagnostic information was more heavily weighted than were prior probabilities. However, information of either kind had more effect on responding when supplied by prior belief, and in one experiment (5), posterior probability estimates were more strongly influenced by (personal) base rates than by diagnostic information. In our laboratory, we have also shown in recent studies that presence of genuine prior beliefs can influence statistical reasoning on other tasks, including choices of diagnostic or pseudodiagnostic information (Evans, Feeney, & Venn, 2001; Feeney, Evans, & Clibbens, 2000) and multicue probability learning (Evans, Clibbens, & Harris, 2005). Whether these effects of belief are seen as biasing or de-

biasing within the tasks set is neither here nor there. What these studies show is that prior beliefs operating through the heuristic system strongly influence problem representation and observed responses on statistical, as well as logical, reasoning tasks.

In contrast with studies of deductive reasoning, researchers in the field of statistical inference have only quite recently approached the topic with an explicit dual-processing framework. However, many of the data do seem amenable to this kind of treatment (Kahneman & Frederick, 2002). Heuristics may make information appear relevant to statistical reasoners for the kinds of reasons proposed by Kahneman and Tversky (availability from memory, similarity, etc.) or simply because the responses are cued by strongly held prior beliefs, resulting in the range of biases observed in this literature. The presence of normatively correct solutions, especially from participants of high cognitive ability, indicates that such heuristic system biases are competing with analytic reasoning processes. Where statistical information must be integrated, as in Bayesian reasoning, only manipulations that make single mental models combining the information available for analytic processing are likely to produce correct responding.

Conclusions

As has been indicated in the review presented here, evidence has been amassing for dual-process effects in the psychology of thinking and reasoning. Broadly, it seems that preconscious *heuristic* processes both focus our attention on selective aspects of presented information and rapidly retrieve and apply relevant prior knowledge and belief. Sometimes our judgments and inferences are determined mostly by these heuristic processes, with any analytic thinking doing no more than translating pragmatically cued beliefs into responses that relate to the experimental instructions given. In other cases, people may use the analytic system actively to inhibit default representations and responses cued by the heuristic system and to engage in conscious strategic thinking. Since publication of the original heuristic-analytic theory (Evans, 1984, 1989), we have learned a great deal more about these processes. For example, we know that analytic reasoning is slow and sequential, requires central working memory resources, can inhibit some influences of heuristic processes by conscious effort, and is responsive to verbal instructions, whereas the opposite seems to be true in all cases of heuristic processes.

Some theorists have tried to develop from this basis the notion that there are two distinct cognitive systems present in the brain with differing evolutionary histories (see Evans, 2003; Stanovich, 2004). Interesting though such speculations are, they may seem to have little immediate relevance to thinking and reasoning researchers attempting to account for the results of their experiments. Hence, in this article, I have offered a revision and extension of the heuristic-analytic theory that incorporates the knowledge gained about dual processes in reasoning and judgment tasks over the past 20 years or so, with the minimum

necessary assumptions about underlying mechanisms. In the process, I have incorporated three principles of hypothetical thinking proposed by Evans, Over, and Handley (2003). I have argued that people construct a single epistemic mental model (hypothesis) at one time and are strongly influenced by preconscious pragmatic processes in the construction of this model. In support of the singularity principle, we have seen that people have great difficulty in dealing with any form of disjunction that requires consideration of two different possibilities.

In contrast with the popular mental models account of Phil Johnson-Laird (1983; Johnson-Laird & Byrne, 1991), I propose that deductive reasoning is not the basic mode of thought. Instead, people satisfice by sticking with a relevant (plausible, believable) model, unless there is good reason to give it up and replace it. The satisficing principle explains by far the most pervasive finding in the psychology of reasoning: that people draw inferences (fallacies) that go beyond the premises given, despite typically being instructed to draw only necessary conclusions (Evans, 2002). The relevance principle explains the other major finding in the field: the pervasive influence of prior knowledge and belief—for example, in the manifestation of belief biases that are apparently impossible to suppress by the use of deductive reasoning instructions and are more strongly marked in their absence.

For many years, the major rival to the Johnson-Laird and Byrne mental model theory of reasoning was cast as one based on rule-based mental logics (Braine & O'Brien, 1998a; Rips, 1994). In common with some other authors (Oaksford & Chater, 1995), however, I see this as a somewhat false and misleading debate (see also Evans & Over, 1997). Both theories are really forms of mental logic that put deduction at center stage and feed off the now controversial deduction paradigm (Evans, 2002). For example, each sees the account of deductive competence as its core task, giving secondary place to the explanation of how pragmatic factors influence reasoning. Mental rule theorists argue that valid inferences follow from rules built into a mental logic, whereas fallacies result from quite separate effects of pragmatic implicature (Braine & O'Brien, 1998b). Similarly, mental model theorists propose that mental models act as extensional logical representations that are subject to pragmatic modulation in the light of real-world knowledge (Evans, Over, & Handley, 2005; Johnson-Laird & Byrne, 2002).

The heuristic-analytic theory does not accept this common agenda of model and rule theories. By contrast, pragmatic inference is seen as the default mode. It is both normal and adaptive to take into account all relevant beliefs and to draw inferences with a degree of confidence or probability. Far from being the central intellectual activity (Rips, 1984), deductive reasoning may be seen as no more than an analytic-level strategy that bright people can be persuaded to adopt by the use of special instructions (Evans, 2000). Deductive effort, when made, attempts to modify pragmatic processes, and not the other way around. In this respect, I agree with Stanovich's (2004) argument that the brain has been optimized by evolution for distrib-

uted processing and is relatively inefficient when applied to sequential and logical forms of thinking.

The conclusion above should give pause for thought to those who still see the psychology of reasoning as somehow apart from other research in cognitive psychology. It is not, although it is understandable that the obsession with normative logic and rationality has created this impression for many in the past. As is apparent in all areas of cognitive psychology (including social cognition), people's performance on cognitive tasks reflects a combination of automatic and effortful processing. I believe that the dual-processing concepts developed within the psychology of reasoning have broad implications for cognitive and neuroscience as a whole, which we are only just beginning to appreciate. I hope that the clarification, revision, and extension of the heuristic-analytic theory of reasoning will not simply assist in the interpretation of reasoning experiments but will point to methodological and theoretical developments of relevance to cognition as a whole.

REFERENCES

- ADAMS, E. (1975). *The logic of conditionals: An application of probability to deductive logic*. Dordrecht: Reidel.
- BALL, L. J., LUCAS, E. J., MILES, J. N. V., & GALE, A. G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology*, **56A**, 1053-1077.
- BARON, J. (1985). *Rationality and intelligence*. Cambridge: Cambridge University Press.
- BENNETT, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
- BRAINE, M. D. S., & O'BRIEN, D. P. (EDS.) (1998a). *Mental logic*. Mahwah, NJ: Erlbaum.
- BRAINE, M. D. S., & O'BRIEN, D. P. (1998b). The theory of mental-propositional logic: Description and illustration. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 79-89). Mahwah, NJ: Erlbaum.
- BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. (1956). *A study of thinking*. New York: Wiley.
- BUELER, R., GRIFFIN, D., & ROSS, M. (2002). Inside the planning fallacy: The causes and consequences of optimistic time predictions. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 250-270). Cambridge: Cambridge University Press.
- BYRNE, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, **31**, 61-83.
- BYRNE, R. M. J. (2005). *The rational imagination*. Cambridge, MA: MIT Press.
- BYRNE, R. M. J., & HANDLEY, S. J. (1997). Reasoning strategies for suppositional deductions. *Cognition*, **62**, 1-49.
- CASSCELLS, W., SCHOENBERGER, A., & GRABOYS, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, **299**, 999-1001.
- CHAIKEN, S., & TROPE, Y. (EDS.) (1999). *Dual-process theories in social psychology*. New York: Guilford.
- CHATER, N., & OAKSFORD, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, **38**, 191-258.
- COSMIDES, L., & TOOBY, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, **58**, 1-73.
- COSMIDES, L., & TOOBY, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations* (pp. 53-115). Oxford: Oxford University Press.
- EDGINGTON, D. (1995). On conditionals. *Mind*, **104**, 235-329.
- ELQAYAM, S. (2006). The collapse illusion: A semantic illusion of truth and paradox. *Thinking & Reasoning*, **12**, 144-180.
- EPSTEIN, S., & PACINI, R. (1999). Some basic issues regarding dual-process theories from the perspective of cognitive-experiential theory. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 462-482). New York: Guilford.
- EVANS, J. ST. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, **75**, 451-468.
- EVANS, J. ST. B. T. (1989). *Bias in human reasoning: Causes and consequences*. London: Erlbaum.
- EVANS, J. ST. B. T. (1995). Relevance and reasoning. In S. E. Newstead & J. S. B. T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp. 147-172). Hove, U.K.: Erlbaum.
- EVANS, J. ST. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, **87**, 223-240.
- EVANS, J. ST. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking & Reasoning*, **4**, 45-82.
- EVANS, J. ST. B. T. (2000). What could and could not be a strategy in reasoning? In W. Schaeken, G. De Vooght, & A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 1-22). Mahwah, NJ: Erlbaum.
- EVANS, J. ST. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, **128**, 978-996.
- EVANS, J. ST. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences*, **7**, 454-459.
- EVANS, J. ST. B. T. (2005a). *How to do research: A psychologist's guide*. Hove, U.K.: Psychology Press.
- EVANS, J. ST. B. T. (2005b). The social and communicative function of conditionals. *Mind & Society*, **4**, 97-114.
- EVANS, J. ST. B. T., BARSTON, J. L., & POLLARD, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, **11**, 295-306.
- EVANS, J. ST. B. T., CLIBBENS, J., & HARRIS, A. (2005). Prior belief and polarity in multicue learning. *Quarterly Journal of Experimental Psychology*, **58A**, 651-666.
- EVANS, J. ST. B. T., CLIBBENS, J., & ROOD, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. *Quarterly Journal of Experimental Psychology*, **48A**, 644-670.
- EVANS, J. ST. B. T., & CURTIS-HOLMES, J. (2005). Rapid responding increases belief bias: Evidence for the dual process theory of reasoning. *Thinking & Reasoning*, **11**, 382-389.
- EVANS, J. ST. B. T., FEENEY, A., & VENN, S. (2001). Implicit and explicit processes in an hypothesis evaluation task. *British Journal of Psychology*, **93**, 31-46.
- EVANS, J. ST. B. T., HANDLEY, S. J., & HARPER, C. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, **54A**, 935-958.
- EVANS, J. ST. B. T., HANDLEY, S. J., HARPER, C., & JOHNSON-LAIRD, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1495-1513.
- EVANS, J. ST. B. T., HANDLEY, S. J., NEILENS, H., & OVER, D. E. (2005). *The mental representation of abstract conditional sentences: A study of qualitative individual differences*. Unpublished manuscript, University of Plymouth.
- EVANS, J. ST. B. T., HANDLEY, S. J., & OVER, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 321-355.
- EVANS, J. ST. B. T., HANDLEY, S. J., OVER, D. E., & PERHAM, N. (2002). Background beliefs in Bayesian inference. *Memory & Cognition*, **30**, 179-190.
- EVANS, J. ST. B. T., HANDLEY, S. J., PERHAM, N., OVER, D. E., & THOMPSON, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, **77**, 197-213.
- EVANS, J. ST. B. T., & LYNCH, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391-397.
- EVANS, J. ST. B. T., NEWSTEAD, S. E., & BYRNE, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, U.K.: Erlbaum.
- EVANS, J. ST. B. T., & OVER, D. E. (1996). *Rationality and reasoning*. Hove, U.K.: Psychology Press.
- EVANS, J. ST. B. T., & OVER, D. E. (1997). Rationality in reasoning: The case of deductive competence. *Current Psychology of Cognition*, **16**, 3-38.

- EVANS, J. ST. B. T., & OVER, D. E. (1999). Explicit representations in hypothetical thinking. *Behavioral & Brain Sciences*, **22**, 763-764.
- EVANS, J. ST. B. T., & OVER, D. E. (2004). *If*. Oxford: Oxford University Press.
- EVANS, J. ST. B. T., OVER, D. E., & HANDLEY, S. J. (2003). A theory of hypothetical thinking. In D. Hardman & L. Maachi (Eds.), *Thinking: Psychological perspectives on reasoning, judgement and decision making* (pp. 3-22). Chichester, U.K.: Wiley.
- EVANS, J. ST. B. T., OVER, D. E., & HANDLEY, S. J. (2005). Supposition, extensionality and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, **112**, 1040-1052.
- FEENEY, A., EVANS, J. ST. B. T., & CLIBBENS, J. (2000). Background beliefs and evidence interpretation. *Thinking & Reasoning*, **6**, 193-272.
- FEENEY, A., & HANDLEY, S. J. (2000). The suppression of q card selections: Evidence for deductive inference in Wason's selection task. *Quarterly Journal of Experimental Psychology*, **53A**, 1224-1243.
- FUGELSANG, J. A., STEIN, C. B., GREEN, A. E., & DUNBAR, K. N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and cognitive laboratory. *Canadian Journal of Experimental Psychology*, **58**, 86-95.
- GENTNER, D., & STEVENS, A. L. (1983). *Mental models*. Hillsdale, NJ: Erlbaum.
- GEORGE, C. (1995). The endorsement of the premises: Assumption based or belief-based reasoning. *British Journal of Psychology*, **86**, 93-111.
- GEORGE, C. (1997). Reasoning from uncertain premises. *Thinking & Reasoning*, **3**, 161-190.
- GIGERENZER, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 284-313). London: Routledge.
- GIGERENZER, G., & HOFFRAGE, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, **102**, 684-704.
- GIGERENZER, G., & TODD, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- GILINSKY, A. S., & JUDD, B. B. (1994). Working memory and bias in reasoning across the life-span. *Psychology & Aging*, **9**, 356-371.
- GILOVICH, T., & GRIFFIN, D. (2002). Introduction—Heuristics and biases: Then and now. In T. Gilovich, D. Griffin, & A. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 1-18). Cambridge: Cambridge University Press.
- GILOVICH, T., GRIFFIN, D., & KAHNEMAN, D. (2002). *Heuristics and biases: The psychology of intuitive judgement*. Cambridge: Cambridge University Press.
- GIROTTO, V., & GONZALEZ, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition*, **78**, 247-276.
- GIROTTO, V., & JOHNSON-LAIRD, P. N. (2004). The probability of conditionals. *Psychologia*, **47**, 207-225.
- GOEL, V. (2005). Cognitive neuroscience of deductive reasoning. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 475-492). Cambridge: Cambridge University Press.
- GOEL, V., & DOLAN, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, **87**, B11-B22.
- HAMMOND, K. R. (1966). *The psychology of Egon Brunswik*. New York: Holt, Rinehart & Winston.
- HANDLEY, S. J., & EVANS, J. ST. B. T. (2000). Supposition and representation in human reasoning. *Thinking & Reasoning*, **6**, 273-312.
- HANDLEY, S. J., FEENEY, A., & HARPER, C. (2002). Alternative antecedents, probabilities and the suppression of fallacies on Wason's selection task. *Quarterly Journal of Experimental Psychology*, **55A**, 799-813.
- HASSIN, R. R., ULEMAN, J. S., & BARGH, J. A. (Eds.) (2005). *The new unconscious*. Oxford: Oxford University Press.
- HOFFRAGE, U., GIGERENZER, G., KRAUSS, S., & MARTIGON, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition*, **84**, 343-352.
- HOWSON, C., & URBACH, P. (1993). *Scientific reasoning*. (2nd ed.). Chicago: Open Court.
- JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- JOHNSON-LAIRD, P. N., & BARA, B. G. (1984). Syllogistic inference. *Cognition*, **16**, 1-61.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics and inference. *Psychological Review*, **109**, 646-678.
- JOHNSON-LAIRD, P. N., LEGRENZI, P., GIROTTO, V., LEGRENZI, M. S., & CAVERNI, J. P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, **106**, 62-88.
- KAHNEMAN, D., & FREDERICK, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgement* (pp. 49-81). Cambridge: Cambridge University Press.
- KAHNEMAN, D., SLOVIC, P., & TVERSKY, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- KAHNEMAN, D., & TVERSKY, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430-454.
- KAHNEMAN, D., & TVERSKY, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237-251.
- KAHNEMAN, D., & TVERSKY, A. (1982a). On the study of statistical intuition. *Cognition*, **12**, 325-326.
- KAHNEMAN, D., & TVERSKY, A. (1982b). The simulation heuristic. In A. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-210). Cambridge: Cambridge University Press.
- KLACZYNSKI, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision-making. *Child Development*, **72**, 844-861.
- KLAUER, K. C., MUSCH, J., & NAUMER, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, **107**, 852-884.
- KLAYMAN, J. (1995). Varieties of confirmation bias. *Psychology of Learning & Motivation*, **32**, 385-417.
- KLAYMAN, J., & HA, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 211-228.
- KLEIN, G. (1999). *Sources of power*. Cambridge, MA: MIT Press.
- KOEHLER, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral & Brain Sciences*, **19**, 1-53.
- KOKIS, J. V., MACPHERSON, R., TOPLAK, M. E., WEST, R. F., & STANOVICH, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, **83**, 26-52.
- KYLLONEN, P., & CHRISTAL, R. E. (1990). Reasoning ability is (little more than) working memory capacity! *Intelligence*, **14**, 389-433.
- LEGRENZI, P., GIROTTO, V., & JOHNSON-LAIRD, P. N. (1993). Focussing in reasoning and decision making. *Cognition*, **49**, 37-66.
- LEVINE, M. (1966). Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, **71**, 331-338.
- LIU, I.-M., LO, K.-C., & WU, J.-T. (1996). A probabilistic interpretation of "if-then." *Quarterly Journal of Experimental Psychology*, **49A**, 828-844.
- MANDEL, D. R., HILTON, D. J., & CATELLANI, P. (Eds.) (2005). *The psychology of counterfactual thinking*. London: Routledge.
- MYNATT, C. R., DOHERTY, M. E., & DRAGAN, W. (1993). Information relevance, working memory, and the consideration of alternatives. *Quarterly Journal of Experimental Psychology*, **46A**, 759-778.
- NEWSTEAD, S. E., HANDLEY, S. J., & BUCK, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Journal of Memory & Language*, **27**, 344-354.
- NEWSTEAD, S. E., HANDLEY, S. J., HARLEY, C., WRIGHT, H., & FARELLY, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*, **57A**, 33-60.
- NISBETT, R. E., PENG, K., CHOI, I., & NORENZAYAN, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, **108**, 291-310.

- OAKHILL, J., & JOHNSON-LAIRD, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology*, **37A**, 553-569.
- OAKHILL, J., JOHNSON-LAIRD, P. N., & GARNHAM, A. (1989). Believability and syllogistic reasoning. *Cognition*, **31**, 117-140.
- OAKSFORD, M., & CHATER, N. (1993). Reasoning theories and bounded rationality. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 31-60). London: Routledge.
- OAKSFORD, M., & CHATER, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning*, **1**, 121-152.
- OAKSFORD, M., & CHATER, N. (1998). *Rationality in an uncertain world*. Hove, U.K.: Psychology Press.
- OBERAUER, K., & WILHELM, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 680-693.
- OVER, D. E. (2004). Naive probability and its model theory. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason: Essays in honour of Paolo Legrenzi* (pp. 139-160). Hove, U.K.: Psychology Press.
- OVER, D. E., HADJICHRISTIDIS, C., EVANS, J. ST. B. T., HANDLEY, S. J., & SLOMAN, S. A. (in press). The probability of causal conditionals. *Cognitive Psychology*.
- POLETIEK, F. (1996). Paradoxes of falsification. *Quarterly Journal of Experimental Psychology*, **49A**, 447-462.
- POLETIEK, F. (2001). *Hypothesis-testing behaviour*. Hove, U.K.: Psychology Press.
- POPPER, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.
- RAMSEY, F. P. (1931). *The foundations of mathematics and other logical essays*. London: Routledge & Kegan Paul.
- REBER, A. S. (1993). *Implicit learning and tacit knowledge*. Oxford: Oxford University Press.
- RIPS, L. J. (1984). Reasoning as a central intellectual ability. In R. J. Sternberg (Ed.), *Advances in the study of human intelligence* (pp. 105-147). Hillsdale, NJ: Erlbaum.
- RIPS, L. J. (1989). The psychology of knights and knaves. *Cognition*, **31**, 85-116.
- RIPS, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- ROBERTS, M. J. (1998). Inspection times and the selection task: Are they relevant? *Quarterly Journal of Experimental Psychology*, **51A**, 781-810.
- ROBERTS, M. J., & NEWTON, E. J. (2002). Inspection times, the change task, and the rapid-response selection task. *Quarterly Journal of Experimental Psychology*, **54A**, 1031-1048.
- ROESE, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, **121**, 133-148.
- SCHROYENS, W., & SCHAEKEN, W. (2004). Guilt by association: On iffy propositions and the proper treatment of mental-models theory. *Current Psychology Letters*, **12**, Vol. 1. Available at <http://cpl.revues.org/document411.html>.
- SCHROYENS, W., SCHAEKEN, W., & D'YDEWALLE, G. (2001). The processing of negations in conditional reasoning: A meta-analytic study in mental models and/or mental logic theory. *Thinking & Reasoning*, **7**, 121-172.
- SCHROYENS, W., SCHAEKEN, W., FIAS, W., & D'YDEWALLE, G. (2000). Heuristic and analytic processes in propositional reasoning with negative conditionals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1713-1734.
- SCHROYENS, W., SCHAEKEN, W., & HANDLEY, S. J. (2003). In search of counterexamples: Deductive rationality in human reasoning. *Quarterly Journal of Experimental Psychology*, **56A**, 1129-1145.
- SCHWARZ, N., & VAUGHN, L. A. (2002). The availability heuristic revisited: Ease of recall and content of recall as distinct sources. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 103-119). Cambridge: Cambridge University Press.
- SIMON, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.
- SLOMAN, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, **119**, 3-22.
- SLOMAN, S. A. (2005). *Causal models*. Oxford: Oxford University Press.
- SLOMAN, S. A., OVER, D. E., & SLOVACK, L. (2003). Frequency illusions and other fallacies. *Organizational Behavior & Human Decision Processes*, **91**, 296-309.
- SPERBER, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations* (pp. 117-138). Oxford: Oxford University Press.
- SPERBER, D., CARA, F., & GIROTTO, V. (1995). Relevance theory explains the selection task. *Cognition*, **57**, 31-95.
- SPERBER, D., & WILSON, D. (1995). *Relevance* (2nd ed.). Oxford: Blackwell.
- STANOVICH, K. E. (1999). *Who is Rational? Studies of Individual Differences in Reasoning*. Mahway, NJ: Erlbaum.
- STANOVICH, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: Chicago University Press.
- STANOVICH, K. E., & WEST, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, **89**, 342-357.
- STANOVICH, K. E., & WEST, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking & Reasoning*, **4**, 193-230.
- STANOVICH, K. E., & WEST, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral & Brain Sciences*, **23**, 645-726.
- STEVENSON, R. J., & OVER, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology*, **48A**, 613-643.
- THOMPSON, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, **22**, 742-758.
- THOMPSON, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition*, **76**, 209-268.
- TVERSKY, A., & KAHNEMAN, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207-232.
- TVERSKY, A., & KAHNEMAN, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293-315.
- TWENEY, R. D., DOHERTY, M. E., WORNER, W. J., PLISKE, D. B., MYNATT, C. R., GROSS, K. A., & ARKKELIN, J. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, **32**, 109-124.
- VERSCHUEREN, N., SCHAEKEN, W., & D'YDEWALLE, G. (2005a). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, **11**, 239-278.
- VERSCHUEREN, N., SCHAEKEN, W., & D'YDEWALLE, G. (2005b). Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory & Cognition*, **33**, 107-119.
- WASON, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, **12**, 129-140.
- WASON, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology I* (pp. 106-137). Harmondsworth, U.K.: Penguin.
- WASON, P. C. (1968). On the failure to eliminate hypotheses: A second look. In P. C. Wason & P. N. Johnson-Laird (Eds.), *Thinking and reasoning* (pp. 165-174). Harmondsworth, U.K.: Penguin.
- WASON, P. C., & BROOKS, P. G. (1979). THOG: The anatomy of a problem. *Psychological Research*, **41**, 79-90.
- WASON, P. C., & EVANS, J. ST. B. T. (1975). Dual processes in reasoning? *Cognition*, **3**, 141-154.
- WETHERICK, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quarterly Journal of Experimental Psychology*, **14**, 246-249.
- WOODWORTH, R. S., & SELLS, S. B. (1935). An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology*, **18**, 451-460.