Human rationality and the psychology of reasoning: Where do we go from here?

Nick Chater*

Department of Psychology, University of Warwick, UK

Mike Oaksford

School of Psychology, Cardiff University, UK

British psychologists have been at the forefront of research into human reasoning for 40 years. This article describes some past research milestones within this tradition before outlining the major theoretical positions developed in the UK. Most British reasoning researchers have contributed to one or more of these positions. We identify a common theme that is emerging in all these approaches, that is, the problem of explaining how prior general knowledge affects reasoning. In our concluding comments we outline the challenges for future research posed by this problem.

The articles in this special issue illustrate the diversity and strength of psychological research in Britain. In this regard, the topic of human reasoning is particularly noteworthy. British researchers have had a disproportionately large role in the initiation and development of the field, as evidenced, for example, by the authorship of the major textbooks in this area (e.g. Evans, 1982, 1989; Evans, Newstead, & Byrne, 1993; Garnham & Oakhill, 1994; Manktelow, 1999). This article outlines how human reasoning has been studied, concentrating on the contribution of British research. Inevitably our views on the current state and future development of the area of human reasoning are highly personal. We do not expect that everyone working in the area will agree with everything we say here, or that they will agree on the selection of work on which we have concentrated. However, our goal was to review the main theories in the area and discuss work that allows us to look forward and speculate on an agenda for future work.

We first point out the paradoxical nature of reasoning research: it seems impossible to assess the quality of human reasoning without circular appeal to the way people reason. We then show how the psychology of reasoning seems to have got round this conceptual problem in the research programme initiated by Peter Wason in the late 1950s and early 1960s. The next three sections concentrate on the major theoretical approaches to human reasoning that have been developed by British researchers in the years since Wason's

* Requests for reprints should be addressed to Nick Chater, Department of Psychology, University of Warwick, Coventry CV4 7AL (e-mail: nick.chater@warwick.ac.uk) or to Mike Oaksford, School of Psychology, Cardiff University, PO Box 901, Cardiff CF10 3YG, Wales, (e-mail: oaksford@cardiff.ac.uk).

pioneering work. Most reasoning researchers in the UK have contributed to one or more of these theoretical approaches. Our goal was not to be exhaustive but to try to identify the main common themes that are emerging from what often seems like very disparate and unrelated approaches. Along the way we also hope to demonstrate the richness, variety and fecundity of the research being conducted in this area by British researchers.

Paradoxes and formal systems

From its inception, the psychology of reasoning has appeared close to paradox. Its goal is to assess empirically the nature and quality of human reasoning. Yet against what standards can such reasoning be assessed? Any putative standards will be human constructions (i.e. products of the very reasoning system, the human brain, that we are attempting to assess). This seems, at the least, dangerously circular, rather akin to checking the veracity of a story in one copy of a newspaper by looking in another copy (to take an example from Wittgenstein (1953) in a different context). A deeper circularity also lurks. If we cast the rationality of human reasoning into doubt, then we risk undermining the very reasoning that went into drawing this conclusion (e.g. in making theoretical predictions, interpreting data, and so on).

These alarming concerns seem to suggest that the psychology of reasoning cannot really assess the *quality* of human reasoning. Instead, the quality of such reasoning must simply be taken for granted, on pain of conceptual self-destruction. On this view, advocated by the Oxford philosopher Jonathan Cohen (1981), the psychology of reasoning is a purely descriptive enterprise: it concerns how people think, but cannot question how well they think. Human rationality is taken as axiomatic, and cannot be assessed empirically.

But, fortunately, there is a way to break out of this viewpoint. It turns out that sophisticated mathematical theories of good reasoning can be derived from extremely simple and apparently uncontroversial assumptions. These theories, though derived indirectly from people's inferential intuitions, stand as independent mathematically specified accounts of good reasoning. They can serve as objective standards against which actual, real-time, human reasoning can be measured. For example, all of propositional logic can be derived from the assumption that one should avoid contradictions (A. R. Anderson & Belnap, 1975). Moreover, the whole of probability theory can be derived from the assumption that one should avoid bets which one is *certain* to lose, whatever the outcome (this is the so-called Dutch book justification for probability; de Finetti, 1937; Ramsey, 1931; Skyrms, 1977). Similar justifications can be given for decision theory and game theory (e.g. Cox, 1961; Savage, 1954; von Neumann & Morgenstern, 1944). These formal theories can be viewed as defining *normative* standards for good reasoning. An empirical programme of psychological research can assess how well actual human reasoning fits these norms.

This is the starting point of the experimental psychology of reasoning. Typically, a reasoning task is defined for which some normative theory is presumed to specify the 'right' answer (we shall see later that this can sometimes be problematic). People then solve the task, and the nature and quality of their reasoning is assessed against the 'right answer', thus providing an assessment of the quality of human reasoning.

Rationality in doubt: Wason's research programme

The usual scattering of precursors aside, this programme of research was initiated systematically in the 1960s by Peter Wason, at University College London (UCL). Wason's experimental work was astonishingly innovative, fruitful and broad (see e.g. the essays in Newstead & Evans, 1994). We focus here on his most celebrated experimental task, the selection task (Wason, 1966, 1968), which has remained probably the most intensively studied task in the field.

In the selection task, people must assess whether some evidence is relevant to the truth or falsity of a conditional rule of the form *if p then q*, where by convention '*p*' stands for the antecedent clause of the conditional and '*q*' for the consequent clause. In the standard abstract version of the task, the rule concerns cards, which have a number on one side and a letter on the other. A typical rule is 'if there is a vowel on one side (*p*), then there is an even number on the other side (*q*)'. Four cards are placed before the participant, so that just one side is visible; the visible faces show an 'A' (*p* card), a 'K' (*not-p* card), a '2' (*q* card) and a '7' (*not-q* card) (see Fig. 1). Participants then select those cards they must turn over to determine whether the rule is true or false. Typical results are: *p* and *q* cards (46%); *p* card only (33%); *p*, *q* and *not-q* cards (7%); and *p* and *not-q* cards (4%) (Johnson-Laird & Wason, 1970a).

The task participants confront is analogous to a central problem of experimental science: the problem of which experiment to perform. The scientist has a hypothesis (or a set of hypotheses) which must be assessed (for the participant, the hypothesis is the conditional rule), and must choose which experiment (card) will be likely to provide data (i.e. what is on the reverse of the card) which bears on the truth of the hypothesis.

The selection task traditionally has been viewed as a deductive task. This is because psychologists of reasoning have tacitly accepted Popper's hypothetico-deductive philosophy of science as an appropriate normative standard, against which people's performance can be judged. Popper (1935/1959) assumes that evidence can falsify but not confirm scientific theories. Falsification occurs when predictions that follow deductively from the theory do not accord with observation. This leads to a recommendation for the choice of experiments: only to conduct experiments that have the potential to falsify the hypothesis under test.

Applying the hypothetico-deductive account to the selection task, the recommendation is that participants should only turn cards that are potentially logically incompatible with the conditional rule. When viewed in these terms, the selection task has a deductive component, in that the participant must deduce logically which cards would be incompatible with the conditional rule. According to the rendition of the conditional as material implication (which is standard in elementary logic; see Haack, 1978), the only

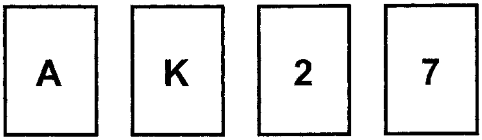

Figure 1. The four cards in the abstract version of Wason's selection task.

observation that is incompatible with the conditional rule *if p then q* is a card with *p* on one side and *not-q* on the other. Hence the participant should select only cards that could potentially yield such an instance. That is, they should turn the *p* card, since it might have a *not-q* on the back; and the *not-q* card, since it might have a *p* on the back.

These selections are rarely observed in the experimental results outlined above. Participants typically select cards that could *confirm* the rule (i.e. the *p* and *q* cards). However, according to falsificationism, the choice of the *q* card is irrational, and is an example of 'confirmation bias' (Evans & Lynch, 1973; Wason & Johnson-Laird, 1972). The rejection of confirmation as a rational strategy follows directly from the hypothetico-deductive perspective.

The dissonance between this normative standard and observed behaviour appears to cast human rationality into severe doubt (e.g. Cohen, 1981; Stein, 1996; Stich, 1985, 1990; Sutherland, 1992). Moreover, a range of other experimental tasks studied by Wason and his co-workers (see Newstead & Evans, 1994, for a review) appeared to suggest that human reasoning is consistently faulty.

Rationality restored? Content and mental models

Rationality can, however, remain in the picture, according to the mental models theory of reasoning, developed by Phil Johnson-Laird (1983), who began his career as a student of Wason's at UCL, later working at Sussex University, the MRC Applied Psychology Unit in Cambridge, and currently Princeton University. Johnson-Laird has the (perhaps unique) distinction of being elected both a Fellow of the Royal Society and of the British Academy. Working closely with Wason, Johnson-Laird began as a major contributor to the experimental reasoning literature. In the first major summary of the area of human reasoning, following the Gestalt problem solving literature (Wertheimer, 1959), Wason and Johnson-Laird (1972; see also Johnson-Laird & Wason, 1970a, 1970b) explained the patterns of performance on the selection task in terms of various levels of insight. People were capable of reasoning logically but required insight into the fact that logic applied to the task. This pattern of explanation also seemed to account for some new and surprising results that using certain contentful materials in the selection task appeared to cause people to switch to apparently logically correct performance (Johnson-Laird, Legrenzi, & Legrenzi, 1972; Wason & Shapiro, 1971). So, for example, using rules such as 'If I travel to Manchester, I take the train' and cards representing train journeys with destination and mode of transport on either side seemed to facilitate selection of the logical *p* (Manchester) and *not-q* (car) cards. According to insight models contentful materials improved insight into the relevance of logic to the task. Thus, it appeared that perhaps people are rational after all—but this rationality is somehow suppressed when reasoning about unnatural, abstract materials. However, the finding that content affects reasoning appears fundamentally to undercut 'formal' views of reasoning, because the logical *form* appears to be the same, independent of the change of content.

Mental models: the theory

During the late 1970s Johnson-Laird began to develop a radically new theoretical perspective on reasoning, culminating in his celebrated book *Mental models*

(Johnson-Laird, 1983). In mental models theory, rationality does have a central place—specifically, standard deductive logic gives the *competence* theory for reasoning, specifying what inferences are valid and which are not. Moreover, the reasoning system is viewed as adapted to drawing valid inferences, at least in principle. What is distinctive about the mental models approach is that reasoning is assumed to involve not the direct application of logical rules in the mind, as had been assumed by Piaget (e.g. Inhelder & Piaget, 1958) and developed in the US (e.g. Braine, 1978; Rips, 1983, 1994), but by creating 'models' of the circumstances described in the premises. Reasoning involves constructing such models, one at a time, reading off conclusions which appear to hold in a model, and then searching for 'counter-example' models and checking whether the conclusion also follows in these models. If a counter-example model is not found, then the conclusion is assumed to be valid (this procedure follows the 'negation as failure' pattern used in logic programming in computer science; Clark, 1978). Mental model theory assumes that errors arise in practice because this process goes awry, most notably when people fail to construct relevant counter-example models.

There are constraints on the states of affairs that mental models can represent which are not captured by traditional 'propositional' ways of representing information. This means that some sets of statements cannot be represented by a single mental model, but rather by a set of mental models. According to mental models theory, the number of models that must be entertained in order to make a valid inference is a key determinant of reasoning difficulty. To see how the theory works, we consider the difference between two apparently similar syllogisms. First, we consider the syllogism

$$\begin{array}{l} \text{Some As are Bs} \\ \text{All Bs are Cs} \\ \hline \therefore \text{ Some As are Cs} \end{array}$$

which can be represented by the mental model:

$$\begin{array}{ccc} \mathrm{A} & [\mathrm{B}] & \mathrm{C} \\ & [\mathrm{B}] & \mathrm{C} \\ & \ldots & \end{array}$$

Here, rows correspond to objects: an object corresponding to the second line of the mental model has properties A, B and C. Square brackets indicate that the item is represented exhaustively in the mental model. Hence in the above mental model there are no additional Bs that are not represented in the model, but there might be additional As or Cs. The fact that there may be additional items not explicitly represented in the model is indicated by the '...' symbol below the model. (This notation is described informally in more detail in Johnson-Laird & Byrne, 1991.) A single model suffices to represent the premises above, so that the reasoner can draw a conclusion simply by reading off that 'Some As are Cs'. Because only a single model need be considered, mental models theory predicts that this is an easy syllogism, which indeed is confirmed empirically.

By contrast, consider the syllogism

Some Bs are As

No Bs are Cs

∴ Some As are not Cs

which is compatible with three different models:

```
A [B]           A [B]           A [B]
A [B]           A [B]           A [B]
      [C]  A          [C]  A          [C]
      [C]             [C]  A          [C]
  ...             ...             ...
```

The crucial difference between the first and the second model is that in the second model there are As that are also Cs, whereas this is not the case in the first model. If a person constructs only the first model, then he or she is likely to conclude that *No As Are Cs*—and indeed, this is a frequently produced conclusion (e.g. Johnson-Laird & Byrne, 1991). To realize that this conclusion does not follow requires that the person also consider the second model. The second model may lead to the erroneous conclusion that *Some Cs are not As*. However, this is ruled out by the third model, where *All the Cs are As*. All these models are consistent with the premises. The correct conclusion, that *Some As are not Cs*, holds in all models. Hence, reaching this conclusion should be substantially more difficult than in the one model syllogism above, and this is observed experimentally. Quite generally, the number of models associated with representing a syllogism is a good predictor of syllogism difficulty. This is an advantage of mental models theory over the competing theory, mental logic, that has been particularly influential in the USA, according to which reasoning involves the application of logical rules (Braine, 1978; Rips, 1983, 1994). The two syllogisms above are associated with very similar logical proofs and hence mental logic does not predict the difference in difficulty that mental models theory captures.

Explanatory breadth

The mental models theory has been very influential, with many researchers in most European countries and in the USA applying and extending the theory to many different domains of reasoning.[1] Johnson-Laird's principle collaborator of the last 10–15 years has been Ruth Byrne who, after working with Johnson-Laird at Cambridge, is now at Trinity College, Dublin. She has not only collaborated with Johnson-Laird in developing mental models theory, she has also pioneered the development of the mental models approach to further domains of reasoning. Across these domains, the number of mental models required to solve a reasoning problem is used to provide a measure of the difficulty

[1] The remarkable influence of mental models theory can perhaps be better appreciated by looking at the Mental Models website maintained by Ruth Byrne at Trinity College, Dublin (http://www2.tcd.ie/Psychology/Ruth_Byrne/mental_models/index.html).

of reasoning problems, and reasoning errors are predicted on the assumption that people do not correctly always entertain all models. Aside from syllogistic reasoning (Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1989; Johnson-Laird & Steedman, 1978), mental models theory has been applied to reasoning with multiply quantified statements (Johnson-Laird, Byrne, & Tabossi, 1990), meta-logical reasoning about truth and falsity (Johnson-Laird & Byrne, 1990), model reasoning (Bell & Johnson-Laird, 1998; Johnson-Laird & Byrne, 1992), counterfactual reasoning (Byrne, Segura, Culhane, Tasso, & Berrocal, 2000; Byrne & Tasso, 1999), spatial reasoning (Byrne & Johnson-Laird, 1989; Mani & Johnson-Laird, 1982), temporal reasoning (Schaeken, Johnson-Laird, & d'Ydewalle, 1996), propositional reasoning (Johnson-Laird, Byrne & Schaeken, 1992, 1995), conditional reasoning (Byrne, 1989; Johnson-Laird & Byrne, 1991, 1992), the selection task (Johnson-Laird & Byrne, 1991), and to a limited class of reasoning about probability (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999; Johnson-Laird & Savary, 1996). The extension to probabilistic reasoning is particularly important because a variety of probabilistic effects have been observed in, for example, conditional inference (Stevenson & Over, 1995; but see Byrne, Espino, & Santamaria, 1999).

Aside from its explanatory breadth in the area of human reasoning, a further attractive feature of the mental models account of reasoning is that it can be applied in other areas of cognition. Most notably it has been applied to theories of language understanding—constructing mental models is assumed to be an intrinsic part of normal language processing, rather than part of a separate 'reasoning system.' The notion that we reason by constructing concrete internal representations of situations making the premises true also has considerable introspective and intuitive plausibility. This plausibility is strengthened by recent work showing an equivalence between the mental models view of syllogistic reasoning and a mode of 'graphical' reasoning using Euler circles, developed by Keith Stenning and colleagues at the University of Edinburgh (Stenning & Oaksford, 1993; Stenning & Oberlander, 1995; Stenning & Yule, 1997). Such graphical reasoning might be part of a general visual or imagistic reasoning capacity.

Effects of content on reasoning performance might appear problematic for the mental models view, to the degree that it is based on logic, because logical inference is independent of the content of the materials being reasoned about. However, research in the USA in the mid-1980s indicated that content effects in Wason's selection task were specific to particular kinds of content (Cheng & Holyoak, 1985). The contents that seemed to produce logical performance involved 'deontic regulations' concerning what preconditions were obliged to be satisfied to perform some action (e.g. 'if you want to enter the country you must be inoculated against cholera'). As Manktelow and Over (1987) (at that time both at Sunderland University; Ken Manktelow is now at the University of Wolverhampton) pointed out that this move to deontic regulations actually changes the task's logical structure. Deontic regulations cannot be falsified; for example, it makes no sense to argue that someone attempting to enter the country without an inoculation falsifies the regulation that he or she should be inoculated. Such an individual *violates* the law, but does not *falsify* it.[2] Consequently, performance is not 'facilitated' from

[2] There is potentially a question of truth or falsity regarding which norms are in force in a particular society or culture (e.g. what the law is in Britain concerning drinking ages). But this is a claim *about* the norm; there is no question of the norm *itself* being true or false.

an initially irrational baseline by the use of these materials, rather the whole problem confronting participants has changed from a problem in inductive inference to a problem of how to enforce regulations. Johnson-Laird and Byrne (1992) argues that this change in the problem structure can be explained readily by mental models theory and consequently so-called 'content effects' can be explained within this framework (although this is disputed by Manktelow & Over, 1992).

Recent developments

Perhaps the most distinctive feature of mental models theory is the claim that people search for counter-examples. However, recent work in this tradition seems to show that although people may be able to construct such alternative interpretations, they tend not to do so spontaneously (Bucciarelli & Johnson-Laird, 1999; Evans, Handley, Harper, & Johnson-Laird, 1999; Newstead, Handley, & Buck, 1999). That is, in the main the experimental results are consistent with people only constructing a single model from which a conclusion is read off. This possibility has always been important in mental models theory in explaining apparent errors in reasoning: if more than one model is required to reach the valid conclusion, people may balk and simply state a conclusion that is true in their initial model. However, this strategy places a heavy explanatory burden on the processes by which initial models are constructed. The way this is achieved has not changed much between the original formulations of mental models theory (e.g. Johnson-Laird & Steedman, 1978) and more recent computer programs for syllogistic reasoning (Johnson-Laird, 1992). The surface form of each premise is parsed into its mental model representation and then the two models are combined. This process of interpretation makes no reference to prior beliefs in long-term memory.

However, in their work on belief bias effects, where people tend to endorse erroneous but believable conclusions, Klauer, Musch, and Naumer (in press) have argued that 'people consider only one mental model of the premises and that belief biases the process of model construction rather than influencing a search for alternative models' (see also, Newstead & Evans, 1993; Newstead, Pollard, & Evans, 1993). This position is also consistent with Evans, Handley, Harper, and Johnson-Laird's (1999) conclusions about how people draw syllogistic arguments (see also Evans & Over, 1996a). That is, prior beliefs can be expected to have a strong influence on the processes of comprehension and hence of initial model construction. As we observed above, according to recent research it is these processes that appear to bear the main explanatory burden in mental models theory. However, 'mental models theorists have . . . focused their research efforts on the "mental model theory per se," and have generally not specified the mechanisms operating in the comprehension stage' (Schroyens, Schaeken, Fias, & d'Ydewalle, in press). Consequently, the future for mental models would appear to be in the direction of outlining how prior beliefs about the world influence how initial models are constructed.

It could be argued that although this must be the case for contentful materials used in belief bias experiments, prior beliefs will play no role in interpreting the abstract materials used in the majority of reasoning experiments. There are two points to make. First, as with any theory of reasoning, mental models theory is not intended to be a theory of abstract laboratory tasks but of everyday human inference about the real world (Chater & Oaksford, 1993; Oaksford & Chater, 1993, 1995a). Thus any theory of reasoning must cope with content and prior knowledge. Secondly, even when abstract materials are used people use their prior knowledge to interpret the claims that are made. For example, Schroyens *et al.*'s (in press) recent work on the mental models approach to conditional inference has shown that 'even in the most abstract contexts probabilistic, knowledge-based processes are called upon'. They showed that the probability of counter-examples using letters and numbers was influenced by prior knowledge of set sizes (see also Oaksford & Chater, 1994; Oaksford & Stenning, 1992).

In summary, it is becoming increasingly clear in research carried out in the mental models framework that the influence of prior knowledge on initial model construction is a key issue. We argue that a similar theme emerges from the other main theoretical positions developed in the UK.

Biases and two types of rationality

Jonathan Evans, another former student of Wason's, has built-up one of the world's leading reasoning research groups at the University of Plymouth. Probably the most prolific experimenter in modern reasoning research, Evans has uncovered many of the core findings that prospective theories of reasoning must explain. Over many years he has also been developing an approach to explaining patterns of human reasoning that sometimes builds on the mental models approach. In the selection task, for example, Evans has pointed out that participants may not be reasoning at all. In an ingenious experiment he introduced what has become known as the Evans' Negations Paradigm, where the task rules have negations inserted systematically into the antecedent and consequent producing three more rule forms: *if p then not-q*; *if not-p then q*; and *if not-p then not-q* (Evans, 1972; Evans & Lynch, 1973). Evans argued that people could just be matching the items named in the rule rather than reasoning. If this were the case then participants should continue to select the *p* and the *q* cards even for the rules containing negations. He argued therefore that people were prone to a matching bias in this task and were not reasoning at all (Evans & Lynch, 1973; see also Evans, 1998; Evans, Legrenzi, & Girotto, 1999).

Evans aims to explain these phenomena by postulating various relevance heuristics which the reasoning system follows (e.g. Evans, 1983, 1984, 1989, 1991). One such heuristic comes from an account of language processing (Wason, 1965) which suggests that the topic of a negated constituent is still that constituent, so the topic of the sentence 'The train is not late' is still the lateness of trains (this is called the *not*-heuristic). That is, this is still the relevant information to attend to. Despite these motivations, these heuristics are tied quite closely to the empirical data, and hence the explanation is quite direct, not arising from an overarching theory. Evans and colleagues have, however, also employed Johnson-Laird's mental models theory as an explanatory framework, especially in the area of conditional inference (e.g. Evans, 1993; Evans, Clibbens, & Rood, 1995; Evans & Handley, 1999). They have also stressed a 'two process' view of human deductive reasoning—one analytic process is based on logic and may be implemented via mental models; the other, more dominant process concerns the heuristics for linguistic interpretation we have sketched above. The overarching goal of these heuristics is to direct attention to the most relevant information, which connects Evans work to recent work on the European continent by Dan Sperber and colleagues (Sperber, Cara, & Girotto,

1995) and to probabilistic approaches to reasoning, described below (Oaksford & Chater, 1994, 1995b).

In general Evans' account suggests that normative theories of good reasoning, such as logic and probability theory, may not play a major role in psychological accounts, whereas Johnson-Laird assumes that logic provides at least a competence theory for inference. Indeed, Evans and his collaborators recently have distinguished two notions of rationality (Evans & Over, 1996a, 1997; Evans, Over, & Manktelow, 1994).

> Rationality$_1$: Thinking, speaking, reasoning, making a decision, or acting in a way that is generally reliable and efficient for achieving one's goal.
>
> Rationality$_2$: Thinking, speaking, reasoning, making a decision, or acting when one has a reason for what one does sanctioned by a normative theory {Evans & Over, 1997, p. 2}.

They argue that 'people are largely rational in the sense of achieving their goals (rationality$_1$) but have only a limited ability to reason or act for good reasons sanctioned by a normative theory (rationality$_2$)' (Evans & Over, 1997, p. 1). If this is right, then achieving one's goals can be achieved without in any sense following a formal normative theory. This viewpoint challenges fundamentally the core assumption in the psychology of reasoning that we mentioned above: that human reasoning performance should be compared against the dictates of theories such as logic and probability. But it leaves a crucial question unanswered: Why is human reasoning so successful in everyday life? After all, human intelligence vastly exceeds that of any artificial computer system, and generally deals effectively with an immensely complex and partially understood physical and social environment. If human reasoning is somehow tied to normative standards, then the justification of those standards as providing methods of good reasoning carries over to explain why human reasoning succeeds. But if this tie is broken, then the remarkable effectiveness of human reasoning remains unexplained.

Evans' work provides a bridge between approaches that concentrate on analytic reasoning and those that concentrate mainly on how people achieve their goals (see below). Evans assumes that the analytic component is based on mental models, and much of his work is in this framework. However, Evans together with David Over have also concentrated their efforts on understanding human reasoning behaviour as resulting from processes adapted to achieving peoples' goals in the environment (Evans & Over, 1996a), in other words from a rationality$_1$ perspective. This has involved a two-pronged approach. First, as we saw above, certain heuristics are suggested that are motivated by the pragmatic goals of successful communication, such as the *not*-heuristic that focuses attention on named items regardless of the negation because they are still the topic of the discourse. Such heuristics are influenced by prior knowledge. For example, assumed knowledge of the purpose of the utterance 'the train is not late' can completely alter its topic. Said ironically to a fellow traveller on the platform, the topic is still the lateness of trains, but said urgently to the ticket seller, the topic is the train being on time. The interpretation of communicative speech depends crucially on prior knowledge.

This is an area that a long-time collaborator of Evans at Plymouth, Stephen Newstead, has been researching for many years. Newstead has suggested that errors may occur in reasoning because people take account of pragmatic communicative factors (Newstead, 1989, 1995) and prior knowledge (Newstead & Evans, 1993; Newstead *et al.*, 1993). However, the evidence for pragmatic influences seems equivocal, some data supporting the view (Newstead, 1989) while other data seem to show less influence (Newstead, 1995). Newstead's work on the belief bias effect (see above) also argues for the view that people only construct one model (Newstead & Evans, 1993; Newstead *et al.*, 1993). More recently Newstead *et al.* (1999) have also shown that there is little evidence that people search for alternative models in syllogistic reasoning (but see Bucciarelli & Johnson-Laird, 1999). This finding again suggests that the main explanatory burden lies with the processes that construct initial models and not with the search for counter-examples.

The second prong of Evans and Over's approach is an appeal to Bayesian decision theory to characterize the way that prior knowledge can influence reasoning. Although such an approach was described in Evans and Over (1996b) they have not pursued it in detail for particular tasks (although see Green & Over, 1997, 2000; Green, Over, & Pyne, 1997; Over & Green, in press). However, we have also been developing this approach systematically to explain behaviour on the main experimental tasks investigated in reasoning research. Consequently, we put off discussion of this approach until the next section. In sum, the main thrust of Evans' dual process view is that analytic processes are extremely limited—people only tend to construct single models of the premises—and that heuristic relevance mechanisms constrain the contents of initial models. More generally, Evans and Over (1996a) suggest that the effects of prior world knowledge may also be captured by adopting a probabilistic approach. It is this approach to which we now turn.

A probabilistic approach

A more recent theoretical proposal concerns our own work on a probabilistic approach to the inferences that people should make in reasoning tasks (Chater & Oaksford, 1999a, 1999b, 1999c, 2000; Oaksford & Chater, 1994, 1995a, 1995b, 1996, 1998a, 1998b, 1998c; Oaksford, Chater, & Grainger, 1999; Oaksford, Chater, Grainger, & Larkin, 1997; Oaksford, Chater, & Larkin, 2000). We can contrast our approach to the other two approaches we have reviewed by concentrating on how each accounts for human rationality. According to the mental models view, we are rational in principle but err in practice, that is, we have sound procedures for deductive reasoning but the algorithms that we use can fail to produce the right answers because of cognitive limitations such as working memory capacity. Such an approach seems difficult to reconcile with two facts. First, these faulty algorithms can lead to error rates as high as 96% (in Wason's selection task) compared to the standard provided by formal logic. Secondly, our everyday rationality in guiding our thoughts and actions seems in general to be highly successful. How is this success to be understood if the reasoning system people use is prone to so much error? The distinction between rationality$_1$ and rationality$_2$ seems to resolve this problem. Our everyday rationality (rationality$_1$) does not depend on formal systems like logic and it is only our formal rationality (rationality$_2$) that is highly constrained and error prone. This line of reasoning follows the old philosophical adage: if you reach a contradiction, draw a distinction. However, we then confront the problem of explaining the success of everyday inference. The problem here is that there are no obvious alternative explanations, aside from arguing that everyday rationality is also somehow based on normative formal reasoning principles, for which good justifications can be given. But this seems to bring us full circle.

We attempt to resolve this problem by arguing that people's everyday reasoning can be understood from the perspective of probability theory and that people make errors in so-called 'deductive tasks' because they generalize their everyday strategies to these laboratory tasks. This approach has been much influenced by J. R. Anderson's (1990, 1991) account of rational analysis. Any laboratory task will recruit some set of cognitive mechanisms that determine the participant's behaviour. But it is not obvious what problem these mechanisms are adapted to solving. This adaptive problem is not likely to be related directly to the problem given to the participant by the experimenter, precisely because adaptation is to the everyday world, not to laboratory tasks. In particular, this means that participants may fail with respect to the tasks that the experimenter thinks he or she has set. But this may be because this task is unnatural with respect to the participant's normal environment. Consequently people may assimilate the task that they are given to a more natural task, recruiting adaptively appropriate mechanisms which solve this, more natural, task successfully.

The psychology of deductive reasoning involves giving people problems that the experimenters conceive of as requiring logical inference. But people respond consistently in a non-logical way, thus calling human rationality into question (Stein, 1996; Stich, 1985, 1990). On our view, everyday rationality is founded on uncertain rather than certain reasoning (Oaksford & Chater, 1991, 1998b) and so probability provides a better starting point for an account of human reasoning than logic. It also resolves the problem of explaining the success of everyday reasoning: it is successful to the extent that it approximates a probabilistic theory of the task. Secondly, we suggest that a probabilistic analysis of classic 'deductive' reasoning tasks provides an excellent empirical fit with observed performance. The upshot is that much of the experimental research in the 'psychology of deductive reasoning' does not engage people in deductive reasoning at all but rather engages strategies suitable for probabilistic reasoning. According to this viewpoint, the field of research appears crucially to be misnamed!

We illustrate our probabilistic approach in the three main tasks that have been the focus of research into human reasoning: conditional inference, Wason's selection task, and syllogistic inference.

Conditional inference

Conditional inference is perhaps the simplest inference form investigated in the psychology of reasoning. It involves presenting participants with a conditional premise, *if p then q*, and then one of four categorical premises, *p*, *not-p*, *q*, or *not-q*. Logically, given the categorical premise *p*, participants should draw the conclusion *q*; and given the categorical premise *not-q* they should draw the conclusion *not-p*. These are the logically valid inferences of modus ponens (MP) and modus tollens (MT) respectively. Moreover, given the categorical premise *not-p*, participants should *not* draw the conclusion *not-q*; and given the categorical premise *q* they should *not* draw the conclusion *p*. These are the logical fallacies of denying the antecedent (DA) and affirming the consequent (AC) respectively. So logically participants should endorse MP and MT in equal proportion and they should refuse to endorse DA or AC. However, they endorse MP significantly more than MT and they endorse DA and AC at levels significantly above zero.

Following some British researchers in this area (Stevenson & Over, 1995) and many others world wide (J. R. Anderson, 1995; Chan & Chua, 1994; George, 1997; Liu, Lo, & Wu, 1996), Oaksford *et al.* (2000) proposed a model of conditional reasoning based on conditional probability. The greater the conditional probability of an inference, the more it should be endorsed. On their account the meaning of a conditional statement can be defined using a 2 ×2 contingency table, as in Table 1 (see Oaksford & Chater, 1998c).

Table 1. The contingency table for a conditional rule, *if p then q*, where there is a dependency between the p and q that may admit exceptions (ε). $a = P(p)$, $b = P(q)$, and $\varepsilon = P(\textit{not-}q|p)$

	q	*not-q*
p	$a(1-\varepsilon)$	$a\varepsilon$
not-p	$b - a(1-\varepsilon)$	$(1-b) - a\varepsilon$

Table 1 represents a conditional rule, *if p then q*, where there is a dependency between p and q that may admit exceptions (ε) and where a is the probability of the antecedent, $P(p)$; b is the probability of the consequent, $P(q)$; and ε is the probability of exceptions (i.e. the probability that q does not occur even though p has), $P(\textit{not-}q|p)$. It is straightforward to then derive conditional probabilities for each inference. For example, the conditional probability associated with MP (i.e. $P(q|p) = 1 - \varepsilon$) depends only on the probability of exceptions. If there are few exceptions the probability of drawing the MP inference will be high. However, the conditional probability associated with MT:

$$P(\textit{not-}p|\textit{not-}q) = \frac{1 - b - a\varepsilon}{1 - b}$$

depends on the probability of the antecedent $P(p)$, and the probability of the consequent $P(q)$, as well the probability of exceptions. As long as there are exceptions ($\varepsilon > 0$) and the probability of the antecedent is greater than the probability of the consequent not occurring ($P(p) > 1 - P(q)$), then the probability of MT is less than MP ($P(\textit{not-}p|\textit{not-}q) < P(q|p)$). For example, if $P(p) = .5$, $P(q) = .8$ and $\varepsilon = .1$, then $P(q|p) = .9$ and $P(\textit{not-}p|\textit{not-}q) = .75$. This behaviour of the model accounts for the preference for MP over MT in the empirical data. In the model, conditional probabilities associated with DA and AC also depend on these parameters which means that they can be non-zero. Consequently the model also predicts that the fallacies should be endorsed to some degree.

Oaksford *et al.* (2000) argue that this simple model can also account for other effects in conditional inference. For example, using Evans' Negations Paradigm (see above) in the conditional inference task leads to a bias towards negated conclusions. Oaksford and Stenning (1992; see also Oaksford & Chater, 1994) proposed that negations define higher probability categories than their affirmative counterparts: for example, the probability that an animal is not a frog is much higher than the probability that it is. Oaksford *et al.* (2000) show that according to their model the conditional probability of an inference increases with the probability of the conclusion. Consequently, the observed bias towards negated conclusions may actually be a rational preference for high probability conclusions. If this is correct then when given rules containing high and low probability categories, people should show a preference to draw conclusions that have a high probability analogous to negative conclusion bias. Oaksford *et al.* (2000) confirmed this prediction in a series of three experiments.

Wason's selection task

The probabilistic approach was applied originally to Wason's selection task, which we introduced above (Oaksford & Chater, 1994, 1995b, 1996, 1998a, 1998b; Oaksford *et al.*, 1999; Oaksford *et al.*, 1997). According to Oaksford and Chater's (1994) optimal data selection model, people select evidence (i.e. turn cards) to determine whether *q* depends on *p*, as in Table 1, or whether *p* and *q* are statistically independent (i.e. the cell values in Table 1 are simply the products of the marginal probabilities). What participants are looking for in the selection task is evidence that gives the greatest probability of discriminating between these two possibilities. Initially, participants are assumed to be maximally uncertain about which possibility is true, i.e. a prior probability of .5 is assigned to both the possibility of a dependency (the dependence hypothesis, H_D) and to the possibility of independence (the independence hypothesis, H_I). Participants' goal is to select evidence (turn cards) that would be expected to produce the greatest reduction in this uncertainty. This involves calculating the posterior probabilities of the hypotheses, H_D or H_I, being true given some evidence. These probabilities are calculated using Bayes' theorem which requires information about prior probabilities ($P(H_D) = P(H_I) = .5$) and the likelihoods of evidence given a hypothesis, for example the probability of finding an A when turning the 2 card assuming H_D ($P(A|2, H_D)$). These likelihoods can be calculated directly from the contingency tables for each hypothesis: for H_D, Table 1; and for H_I, the independence model. With these values it is possible to calculate the reduction in uncertainty that can be expected by turning any of the four cards in the selection task. Oaksford and Chater (1994) observed that assuming that the marginal probabilities $P(p)$ and $P(q)$ were small (their 'rarity assumption'), the *p* and the *q* cards would be expected to provide the greatest reduction in uncertainty about which hypothesis was true. Consequently, the selection of cards that has been argued to demonstrate human irrationality may actually reflect a highly rational data selection strategy. Indeed this strategy may be optimal in an environment where most properties are rare (e.g. most things are not black, not ravens and not apples (but see Klauer, 1999; Chater & Oaksford, 1999b, for a reply).

Oaksford and Chater (1994) argued that this model can account for most of the evidence on the selection task, and defended the model against a variety of objections (Oaksford & Chater, 1996). For example, Evans and Over (1996b) criticized the notion of information used in the optimal data selection model and proposed their own probabilistic model. This model made some predictions that diverged from Oaksford and Chater's model and these have been experimentally tested by Oaksford *et al.* (1999). Although the results seem to support the optimal data selection model, there is still much room for further experimental work in this area. Manktelow and Over have been exploring probabilistic effects in deontic selection tasks (Manktelow, Sutherland, & Over, 1995). Moreover, David Green, who is at University College London, and David Over have also been exploring the probabilistic approach to the standard selection task (Green, Over, & Pyne, 1997; see also, Oaksford, 1998; Green & Over, 1998; Over & Jessop, 1998). They have also extended this approach to what they refer to as 'causal selection tasks' (Green & Over, 1997, 2000; Over & Green, in press). This is important because their work develops the link between research on causal estimation (e.g. J. R. Anderson & Sheu, 1995; Cheng, 1997) and research on the selection task originally suggested by Oaksford and Chater (1994).

Syllogistic reasoning

Chater and Oaksford (1999c) have further extended the probabilistic approach to the more complex inferences involved in syllogistic reasoning that we discussed in looking at mental models. In their probability heuristics model (PHM) they extend their probabilistic interpretation of conditionals to quantified claims, such as All, Some, None, and Some . . not. In Table 1, if there are no exceptions, then the probability of the consequent given the antecedent ($P[q|p]$) is 1. The conditional and the universal quantifier 'All' have the same underlying logical form: $\forall x(P(x) \Rightarrow Q(x))$. Consequently Chater and Oaksford interpreted universal claims such as All Ps are Qs, as asserting that the probability of the predicate term (Q) given the subject term (P) is 1 (i.e. $P(Q|P) = 1$). Probabilistic meanings for the other quantifiers are then easily defined (None, $P(Q|P) = 0$; Some $P(Q|P) > 0$; Some . . not, $P(Q|P) < 1$). Given these probabilistic interpretations it is possible to prove what conclusions follow probabilistically for all 64 syllogisms (i.e. which syllogisms are *p*-valid). Moreover, given these interpretations and again making the rarity assumption (see above on the selection task), the quantifiers can be ordered in terms of how informative they are (All > Some > None > Some . . not). It turns out that a simple set of heuristics defined over the informativeness of the premises can successfully predict the *p*-valid conclusion if there is one. The most important of these heuristics is the *min*-heuristic, which states that the conclusion will have the form of the least informative premise. So, for example, a *p*-valid syllogism such as *All B are A*, *Some B are not C*, yields the conclusion *Some A are not C*. Note that the conclusion has the same form as the least informative premise. This simple heuristic captures the form of the conclusion for most *p*-valid syllogisms. Moreover, if overgeneralized to the invalid syllogisms, the conclusions it suggests match the empirical data very well. Other heuristics determine the confidence that people have in their conclusions and the order of terms in the conclusion.[3]

Perhaps the most important feature of PHM is that it can generalize to syllogisms containing quantifiers such as Most and Few that have no logical interpretation. In terms of Table 1, the suggestion is that these terms are used instead of All when there are some (Most) or many (Few) exceptions. So the meaning of Most is $1 - \Delta < P(Q|P) < 1$, and the meaning of Few is $0 < P(Q|P) < \Delta$, where Δ is small. These interpretations lead to the following order of informativeness: All > Most > Few > Some > None > Some . . not. Consequently, PHM uniquely makes predictions for the 144 syllogisms that are produced when Most and Few are combined with the standard logical quantifiers. Chater and Oaksford (1999c) show that (1) their heuristics pick out the *p*-valid conclusions for these new syllogisms; and (2) they report experiments confirming the predictions of PHM when Most and Few are used in syllogistic arguments.

[3] In our example, the *min*-heuristic only dictates that the conclusion should be *Some . . not*, but this is consistent with either *Some A are not C* or *Some C are not A*; however, only *Some A are not C* is *p*-valid.

There has already been some work on syllogistic reasoning consistent with PHM. Newstead *et al.* (1999) found that the conclusions participants drew in their experiments were mainly as predicted by the *min*-heuristic, although they found little evidence of the search for counter-examples predicted by mental models theory for multiple model syllogisms. Evans, Handley, Harper & Johnson-Laird (1999) also found evidence consistent with PHM. Indeed, they found that an important novel distinction they discovered between strong and weak possible conclusions could be captured as well by the *min*-heuristic as by mental models theory. A conclusion is necessarily true if it is true in all models of the premises, a conclusion is possibly true if it is true in at least one model of the premises, and a conclusion is impossible if it is not true in any model of the premises. Evans, Handley, Harper & Johnson-Laird (1999) found that some possible conclusions were endorsed by as many participants as necessary conclusions and that some were endorsed by as few participants as impossible conclusions. According to mental models theory this happens because strong possible conclusions are those that are true in the initial model constructed, but not in subsequent models. Moreover, weak possible conclusions are those that are only true in non-initial models.[4] Possible strong conclusions all conform to the *min*-heuristic in that they either match the *min*-premise or are less informative than the *min*-premise. Possible weak conclusions all violate the *min*-heuristic (bar one), in that they have conclusions that are more informative than the *min*-premise. In sum, PHM would appear to be gaining some empirical support.

Summary

Our probabilistic approach contrasts with Evans and Over (1996a, 1996b) in that we see probability theory as a wholesale replacement for logic as a computational level theory of what inferences people should draw. Consequently, other than a learned facility for logical reasoning, we do not regard logical inference as a part of the innate architecture of cognition. Evans and Over, on the other hand, still seem to view some, however limited, facility for logical thought as part of our innate cognitive endowment. However, this difference is superficial compared to the major points on which we agree. The one problem for all probabilistic approaches is that they are largely formulated at the computational level (i.e. they concern *what* gets computed, not *how*). However, if such an approach is to be viable at the algorithmic level then there is a tacit assumption that the mind/brain is capable of working out how probable various events are.[5] This means that the probabilistic approach faces similar problems to mental models and Evans' relevance approach. The key to human reasoning appears to lie in how world knowledge provides only the most plausible model of some premises, or accesses only the most relevant information, or permits an assessment of the probabilities of events. This problem provides reasoning research with a challenge that is likely to keep researchers in this area busy for some years to come for reasons that we outline in the next and final section of the article.

[4] This again shows that the main burden of explanation lies with the processes that construct initial models.

[5] Of course, such information need not be represented as a number over which the probability calculus operates (e.g. it could be the activation level of a node of a neural network).

Where do we go from here?

Despite the intensive research effort outlined above, human reasoning remains largely mysterious. While there is increased understanding of laboratory performance as we have discussed above, deep puzzles over the nature of everyday human reasoning processes remain. We suggest that three key issues may usefully frame the agenda for future research: (1) establishing the relation between reasoning and other cognitive processes; (2) developing formal theories which capture the full richness of everyday reasoning; and (3) explaining how such theories can be implemented in real-time in the brain.

Reasoning and cognition

From an abstract perspective, almost every aspect of cognition can be viewed as involving inference. Perception involves inferring the structure of the environment from perceptual input; motor control involves inferring appropriate motor commands from proprioceptive and perceptual input, together with demands of the motor task to be performed; learning from experience, in any domain, involves inferring general principles from specific examples; and understanding a text or utterance typically requires inferences relating the linguistic input to an almost unlimited amount of general background knowledge. Is there a separate cognitive system for *reasoning*, or are the processes studied by reasoning researchers simply continuous with the whole of cognition? A key sub-question concerns the modularity of the cognitive system. If the cognitive system is non-modular, then reasoning would seem, of necessity, to be difficult to differentiate from other aspects of cognition. If the cognitive system is highly modular, then different principles may apply in different cognitive domains. Nonetheless, it might still turn out that even if modules are informationally sealed off from each other (e.g. Fodor, 1983; Pylyshyn, 1984) the inferential principles that they use might be the same; the same underlying principles and mechanisms might simply be reused in different domains. Even if the mind is modular, it seems unlikely that there could be a module for *reasoning* in anything like the sense studied in psychology. This is because everyday reasoning (in contrast to some artificial laboratory tasks) requires engaging arbitrary world knowledge. Consequently, understanding reasoning would appear to be part of the broader project of understanding central cognitive processes and the knowledge they embody in full generality.

This is an alarming prospect for reasoning researchers because current formal research is unable to provide adequate tools for capturing even limited amounts of general knowledge, let alone reasoning with it effectively and in real-time, as we discuss below. Reasoning researchers often attempt to seal off their theoretical accounts from the deep waters of general knowledge by assuming that these problems are solved by other processes (e.g. processes constraining how mental models are 'fleshed out' (Johnson-Laird & Byrne, 1991) or when particular premises can be used in inference (Politzer & Braine, 1991), what information is relevant (Evans, 1989; Sperber *et al.*, 1995) or how certain probabilities are determined (Oaksford & Chater, 1994)). Whether or not this strategy is methodologically appropriate in the short term, substantial progress in understanding everyday reasoning will require theories which address, rather than duck, these crucial issues (i.e. that explicate, rather than presuppose, our judgments concerning what is plausible, probable or relevant). Moreover, as we have seen, the recent empirical work seems strongly to suggest that progress in understanding human reasoning even in the laboratory requires the issue of general knowledge to be addressed.

Formal theories of everyday reasoning

Explaining the cognitive processes involved in everyday reasoning requires developing a formal theory that can capture everyday inferences. Unfortunately, however, this is far from straightforward, because everyday inferences are *global*: whether a conclusion follows typically depends not just on a few circumscribed 'premises', but on arbitrarily large amounts of general world knowledge (see e.g. Fodor, 1983; Oaksford & Chater, 1991, 1998b). From a statement such as 'While John was away, Peter changed all the locks in the house', we can provisionally infer, for example, that Peter did not want John to be able to enter the house, that John possesses a key, that Peter and John have had a disagreement, and so on. But such inferences draw on background information, such as that the old key will not open the new lock, that locks secure doors, that houses can usually only be entered through doors, and a host more information about the function of houses, the nature of human relationships, and the law concerning breaking and entering. Moreover, deploying each piece of information requires an inference which is just as complex as the original one. Thus, even to infer that John's key will not open the new lock requires background information concerning the way in which locks and keys are paired together, the convention that when locks are replaced they will not fit the old key, that John's key will not itself be changed when the locks are changed, that the match between lock and key is stable over time, and so on. This is what we call the 'fractal' character of everyday reasoning—just as, in geometry, each part of a fractal is as complex as the whole, each part of an everyday inference is as complex as the whole piece of reasoning.

How can such inferences be captured formally? Deductive logic is inappropriate, because everyday arguments are not deductively valid, but can be overturned when more information is learned.[6] The essential problem is that these methods fail to capture the global character of everyday inference successfully (Oaksford & Chater, 1991, 1992, 1993, 1998b). In artificial intelligence, this has led to a switch to using probability theory, the calculus of *un*certain reasoning, to capture patterns of everyday inference (e.g. Pearl, 1988). This is an important advance, but only a beginning. Probabilistic inference can only be used effectively if it is possible to separate knowledge into discrete chunks—with a relatively sparse network of probabilistic dependencies between the chunks. Unfortunately, this just does not seem to be possible for everyday knowledge. The large variety of labels for the current impasse (the 'frame' problem (McCarthy & Hayes, 1969; Pylyshyn, 1987), the 'world knowledge' problem or the problem of knowledge representation (Ginsberg, 1987), the problem of non-monotonic reasoning (Paris, 1994), the criterion of completeness* (Oaksford & Chater, 1991, 1998b)) is testimony to its fundamental importance and profound difficulty. The problem of providing a formal calculus of everyday inference presents a huge intellectual challenge,

[6] Technically, these inferences are *non-monotonic*: and extensions to logic to capture everyday inference, though numerous, have been uniformly unsuccessful—they just do not capture inferences that people make routinely, or they fall into paradox (Ginsberg, 1987; Hanks & McDermott, 1985; Oaksford & Chater, 1991).

not just in psychology, but in the study of logic, probability theory, artificial intelligence and philosophy.

Everyday reasoning and real-time neural computation

Suppose that a calculus which captured everyday knowledge and inference could be developed. If this calculus underlies thought, then it must be implemented (probably to an approximation) in real-time in the human brain. Current calculi for reasoning, including standard and non-standard logics, probability theory, decision theory and game theory, are computationally intractable (Garey & Johnson, 1979; Paris, 1994). That is, as the amount of information that they have to deal with increases, the amount of computational resources (in memory and time) required to derive conclusions explodes very rapidly (or, in some cases, inferences are not computable at all, even given limitless time and memory). Typically, attempts to extend standard calculi to mimic everyday reasoning more effectively make problems of tractability *worse* (e.g. this is true of 'non-monotonic logics' developed in artificial intelligence). Somehow, a formal calculus of everyday reasoning must be developed which, instead, eases problems of tractability.

This piles difficulty upon difficulty for the problem of explaining human reasoning computationally. Nonetheless, there are interesting directions to explore. For example, modern 'graphical' approaches to probabilistic inference in artificial intelligence and statistics (e.g. Pearl, 1988) are very directly related to connectionist computation; and more generally, connectionist networks can be viewed as probabilistic inference machines (Chater, 1995; Mackay, 1992; McClelland, 1998). To the extent that the parallel, distributed style of computation in connectionist networks can be related to the parallel, distributed computation in the brain, this suggests that the brain may be understood, in some sense, as directly implementing rational calculations. Nonetheless, there is presently little conception either of how such probabilistic models can capture the 'global' quality of everyday reasoning, or how these probabilistic calculations can be carried out in real-time to support fluent and rapid inference, drawing on large amounts of general knowledge, in a brain consisting of notoriously slow and noisy neural components (Feldman & Ballard, 1982).

Where do we stand?

British research has been at the forefront of the psychology of reasoning, both in uncovering empirical phenomena and in developing theoretical proposals. This research appeared to bring human rationality into question—a conclusion which is both conceptually puzzling and apparently at variance with the manifest practical effectiveness of human intelligence. We have suggested that the probabilistic approach to reasoning provides a way of reconciling experimental data with human rationality, by allowing that the rational theory of the task is not specified *a priori* but is part of an empirical scientific explanation. But the psychology of reasoning faces vast challenges, in developing theoretical accounts that are rich enough to capture the 'global' character of human everyday reasoning, and that can be implemented in real-time in the human brain. This is probably a challenge not merely for the next century, but for the next millennium.

References

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity* (Vol. 1). Princeton, NJ: Princeton University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14,* 471–517.

Anderson, J. R. (1995). *Cognitive psychology and its implications.* New York: Freeman & Co.

Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Memory and Cognition, 23,* 510–524.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22,* 25–51.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85,* 1–21.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23,* 247–303.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 1–21.

Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language, 40,* 347–373.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564–575.

Byrne, R. M. J., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000). The temporality effect in counterfactual thinking about what might have been. *Memory and Cognition, 28,* 264–281.

Byrne, R. M. J., & Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory and Cognition, 27,* 726–740.

Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition, 53,* 217–238.

Chater, N. (1995). Neural networks: The new statistical models of mind. In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (pp. 207–227). London: UCL Press.

Chater, N., & Oaksford, M. (1993). Logicism, mental models and everyday reasoning: Reply to Garnham. *Mind & Language, 8,* 72–89.

Chater, N., & Oaksford, M. (1999a). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences, 3,* 57–65.

Chater, N., & Oaksford, M. (1999b). Information given vs. decision-theoretic approaches to data selection: Response to Klauer. *Psychological Review, 106,* 223–227.

Chater, N., & Oaksford, M. (1999c). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38,* 191–258.

Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behaviour. *Synthese, 122,* 93–131.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405.

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

Clark, K. L. (1978). Negation as failure. In *Logic and databases* (pp. 293–322). New York: Plenum Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317–370.

Cox, R. T. (1961). *The algebra of probable inference.* Baltimore: The Johns Hopkins University Press.

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives {Foresight: Its logical laws, its subjective sources}. *Annales de l'Institute Henri Poincaré, 7,* 1–68. (Trans. in H. E. Kyburg & H. E. Smokler (Eds.) (1964). *Studies in subjective probability.* Chichester: Wiley.)

Evans, J. St.B. T. (1972). Interpretation and 'matching bias' in a reasoning task. *Quarterly Journal of Experimental Psychology, 24,* 193–199.

Evans, J. St.B. T. (1982). *The psychology of deductive reasoning.* London: Routledge & Kegan Paul.

Evans, J. St.B. T. (1983). Selective processes in reasoning. In Evans, J. St.B. T. (Ed.), *Thinking and reasoning: Psychological approaches* (pp. 135–163). London: Routledge & Kegan Paul.

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. (1991). Theories of human reasoning: The fragmented state of the art. *Theory & Psychology, 1,* 83–105.

Evans, J. St.B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition, 48,* 1–20.

Evans, J. St.B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking and Reasoning, 4,* 45–82.

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., & Handley, S. J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology, 52,* 739–770.

Evans, J. St.B. T., Handley, S. J., Harper, C. N. J., & 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, and Cognition, 25,* 1495–1513.

Evans, J. St.B. T., Legrenzi, P., & Girotto, V. (1999). The influence of linguistic form on reasoning: The case of matching bias. *Quarterly Journal of Experimental Psychology, 52A,* 185–216.

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.* Hove: Erlbaum.

Evans, J. St.B. T., & Over, D. (1996a). *Rationality and reasoning.* Hove: Psychology Press.

Evans, J. St.B. T., & Over, D. (1996b). Rationality in the selection task: Epistemic utility vs. uncertainty reduction. *Psychological Review, 103,* 356–363.

Evans, J. St.B. T., & Over, D. (1997). Rationality in reasoning: The problem of deductive competence. *Cahiers de Psychologie Cognitive, 16,* 1–35.

Evans, J. St.B. T., Over, D., & Manktelow, K. I. (1994). Reasoning, decision making and rationality. *Cognition, 49,* 165–187.

Feldman, J., & Ballard, D. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205–254.

Fodor, J. A. (1983). *Modularity of mind.* Cambridge, MA: MIT Press.

Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness.* San Fransisco: W. H. Freeman.

Garnham, A. & Oakhill, J. (1994). *Thinking and reasoning.* Oxford: Blackwell.

George, C. (1997). Reasoning from uncertain premises. *Thinking and reasoning, 3,* 161–190.

Ginsberg, M. L. (Ed.) (1987). *Readings in nonmonotonic reasoning.* Los Altos, CA: Morgan Kaufman.

Green, D. W., & Over, D. E. (1997). Causal inference, contingency tables and the selection task. *Current Psychology of Cognition, 16,* 459–487.

Green, D. W., & Over, D. E. (1998). Reaching a decision: A reply to Oaksford. *Thinking and Reasoning, 4,* 231–248.

Green, D. W., & Over, D. E. (2000). Decision theoretic effects in the selection task. *Current Psychology of Cognition, 19,* 51–68.

Green, D. W., Over, D. E., & Pyne, R. A. (1997). Probability and choice in the selection task. *Thinking and Reasoning, 3,* 209–235.

Haack, S. (1978). *Philosophy of logics.* Cambridge: Cambridge University Press.

Hanks, S., & McDermott, D. (1985). Default reasoning, nonmonotonic logics, and the frame problem. *Proceedings of the American Association for Artificial Intelligence* (pp. 328–333). Philadelphia, PA.

Inhelder, B., & Piaget, J. (1958). *The growth of logical reasoning.* New York: Basic Books.

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

Johnson-Laird, P. N. (1992). Syllogs {computer program}. http://www.cogsci.princeton.edu/~phil/syl.txt.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16,* 1–62.

Johnson-Laird, P. N., & Byrne, R. M. J. (1989). Only reasoning. *Memory and Language, 28,* 313–330.

Johnson-Laird, P. N., & Byrne, R. M. J. (1990). Meta-logical reasoning: Knights, knaves and Rips. *Cognition, 36,* 173–182.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1992). Modal reasoning, models, and Manktelow and Over. *Cognition, 43,* 173–182.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99,* 418–439.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1995). Why models rather than rules give a better account of propositional reasoning: A reply to Bonatti and to O'Brien, Braine, and Yang. *Psychological Review, 101,* 734–739.

Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1990). Reasoning by model: The case of multiple quantification. *Psychological Review, 96,* 658–673.

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.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395–400.

Johnson-Laird, P. N., & Savary, F. (1996). Illusory inferences about probabilities. *Acta Psychologica, 93,* 69–90.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology, 10,* 64–99.

Johnson-Laird, P. N., & Wason, P. C. (1970a). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology, 1,* 134–148.

Johnson-Laird, P. N., & Wason, P. C. (1970b). Insight into a logical relation. *Quarterly Journal of Experimental Psychology, 22,* 49–61.

Klauer, K. C. (1999). On the normative justification for information gain in Wason's selection task. *Psychological Review, 106,* 215–222.

Klauer, K. C., Musch, J., & Naumer, B. (in press). On belief bias in syllogistic reasoning. *Psychological Review.*

Liu, I., Lo, K., & Wu, J. (1996). A probabilistic interpretation of 'If-then'. *Quarterly Journal of Experimental Psychology, 49A,* 828–844.

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation, 4,* 590–604.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10,* 181–187.

Manktelow, K. I. (1999). *Reasoning and thinking.* Hove: Psychology Press.

Manktelow, K. I., & Over, D. E. (1987). Reasoning and rationality. *Mind and Language, 2,* 199–219.

Manktelow, K. I., & Over, D. E. (1992). Utility and deontic reasoning: Some comments on Johnson-Laird and Byrne. *Cognition, 43,* 183–188.

Manktelow, K. I., Sutherland, E. J., & Over, D. E. (1995). Probabilistic factors in deontic reasoning. *Thinking and Reasoning, 1,* 201–220.

McCarthy, J. M., & Hayes, P. (1969). Some philosophical problems from the standpoint of Artificial Intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence* (Vol. 4, pp. 463–502). Edinburgh: Edinburgh University Press.

McClelland, J. L. (1998). Connectionist models and Bayesian inference. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 21–53). Oxford: Oxford University Press.

Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language, 28,* 78–91.

Newstead, S. E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language, 34,* 644–664.

Newstead, S. E., & Evans, J. St.B. T. (1993). Mental models as an explanation of belief bias effects in syllogistic reasoning. *Cognition, 46,* 93–97.

Newstead, S. E., & Evans, J. St.B. T. (Eds.) (1994). *Perspectives in thinking and reasoning: Essays in honour of Peter Wason.* Hove: Erlbaum.

Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory & Cognition, 27,* 344–354.

Newstead, S. E., Pollard, P., & Evans, J. St.B. T. (1993). The source of belief bias effects in syllogistic reasoning. *Cognition, 45,* 257–284.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language, 6,* 1–38.

Oaksford, M., & Chater, N. (1992). Bounded rationality in taking risks and drawing inferences. *Theory and Psychology, 2,* 225–230.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 131–60). London: Routledge.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631.

Oaksford, M., & Chater, N. (1995a). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning, 1,* 121–152.

Oaksford, M., & Chater, N. (1995b). Information gain explains relevance which explains the selection task. *Cognition, 57,* 97–108.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381–391.

Oaksford, M., & Chater, N. (Eds.) (1998a). *Rational models of cognition.* Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (1998b). *Rationality in an uncertain world.* Hove: Psychology Press.

Oaksford, M., & Chater, N. (1998c). A revised rational analysis of the selection task: Exceptions and sequential sampling. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 372–398). Oxford: Oxford University Press.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5,* 193–243.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 441–458.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 883–899.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 835–854.

Over, D. E., & Green, D. W. (2000). Contingency, causation, and adaptive heuristics. *Psychological Review.*

Over, D. E., & Jessop, A. (1998). Rational analysis of causal conditionals and the selection task. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 399–414). Oxford: Oxford University Press.

Paris, J. (1994). *The uncertain reasoner's companion.* Cambridge: Cambridge University Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufman.

Politzer, G., & Braine, M. D. S. (1991). Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition, 38,* 103–108.

Popper, K. R. (1935/1959). *The logic of scientific discovery.* London: Hutchinson.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Montgomery, VT: Bradford.

Pylyshyn, Z. W. (Ed.) (1987). *The robot's dilemma: The frame problem in artificial intelligence.* Norwood, NJ: Ablex.

Ramsey, F. P. (1931). *The foundations of mathematics and other logical essays.* London: Routledge and Kegan Paul.

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

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition, 60,* 205–234.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle, G. (in press). Heuristic and analytic processes in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Skyrms, B. (1977). *Choice and chance.* Belmont, CA: Wadsworth.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57,* 31–95.

Stein, E. (1996). *Without good reason.* Oxford: Oxford University Press.

Stenning, K., & Oaksford, M. (1993). Rational reasoning and human implementations of logics. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 136–177). London: Routledge.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science, 19,* 97–140.

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology, 34,* 109–159.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology, 48A,* 613–643.

Stich, S. (1985). Could man be an irrational animal? *Synthese, 64,* 115–135.

Stich, S. (1990). *The fragmentation of reason.* Cambridge, MA: MIT Press.

Sutherland, N. S. (1992). *Irrationality: The enemy within.* London: Constable.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior, 4,* 7–11.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology,* (pp. 135–151). Harmondsworth: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20,* 273–281.

Wason, P. C., & Johnson-Laird, P. N. (1972). *The psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23,* 63–71.

Wertheimer, M. (1959). *Productive thinking.* New York: Harper & Row.

Wittgenstein, L. (1953). *Philosophical investigations.* Oxford: Basil Blackwell.