BEHAVIORAL AND BRAIN SCIENCES (2009) **32**, 69–120
Printed in the United States of America
doi:10.1017/S0140525X09000284

Précis of *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*

Mike Oaksford
School of Psychology, Birkbeck College London, London, WC1E 7HX, United Kingdom
m.oaksford@bbk.ac.uk
www.bbk.ac.uk/psyc/staff/academic/moaksford

Nick Chater
Division of Psychology and Language Sciences, and ESRC Centre for Economic Learning and Social Evolution, University College London, London, WC1E 6BT, United Kingdom
n.chater@ucl.ac.uk
www.psychol.ucl.ac.uk/people/profiles/chater_nick.htm

Abstract: According to Aristotle, humans are the rational animal. The borderline between rationality and irrationality is fundamental to many aspects of human life including the law, mental health, and language interpretation. But what is it to be rational? One answer, deeply embedded in the Western intellectual tradition since ancient Greece, is that rationality concerns reasoning according to the rules of logic – the formal theory that specifies the inferential connections that hold with certainty between propositions. Piaget viewed logical reasoning as defining the end-point of cognitive development; and contemporary psychology of reasoning has focussed on comparing human reasoning against logical standards.

Bayesian Rationality argues that rationality is defined instead by the ability to reason about *un*certainty. Although people are typically poor at numerical reasoning about probability, human thought is sensitive to subtle patterns of qualitative Bayesian, probabilistic reasoning. In Chapters 1–4 of *Bayesian Rationality* (Oaksford & Chater 2007), the case is made that cognition in general, and human everyday reasoning in particular, is best viewed as solving probabilistic, rather than logical, inference problems. In Chapters 5–7 the psychology of "deductive" reasoning is tackled head-on: It is argued that purportedly "logical" reasoning problems, revealing apparently irrational behaviour, are better understood from a probabilistic point of view. Data from conditional reasoning, Wason's selection task, and syllogistic inference are captured by recasting these problems probabilistically. The probabilistic approach makes a variety of novel predictions which have been experimentally confirmed. The book considers the implications of this work, and the wider "probabilistic turn" in cognitive science and artificial intelligence, for understanding human rationality.

Keywords: Bayes' theorem, conditional inference, logic, non-monotonic reasoning, probability, rational analysis, rationality, reasoning, selection task, syllogisms

Bayesian Rationality (Oaksford & Chater 2007, hereafter *BR*) aims to re-evaluate forty years of empirical research in the psychology of human reasoning, and cast human rationality in a new and more positive light. Rather than viewing people as flawed logicians, we focus instead on the spectacular success of human reasoning under uncertainty. From this perspective, everyday thought involves astonishingly rich and subtle probabilistic reasoning – but probabilistic reasoning which is primarily qualitative, rather than numerical. This viewpoint leads to a radical re-evaluation of the empirical data in the psychology of reasoning. Previously baffling logical "errors" in reasoning about even the simplest statements can be understood as arising naturally from patterns of qualitative probabilistic reasoning.

Why "Bayesian" rationality, rather than mere "probabilistic" rationality? The answer is that our approach draws crucially on a particular interpretation of probability, not merely on the mathematics of probability itself. Probability is often taught as capturing "objective" facts about something, for example, gambling devices such as dice or cards. It is sometimes presumed to be a fact, for example, that the probability of a fair coin producing three consecutive heads is 1/8. However, in the context of cognitive science, probability refers not to objective facts about gambling devices or anything else, but rather, it describes a reasoner's *degrees of belief*. Probability theory is then a calculus not for solving mathematical problems about objects in the world, but a calculus for rationally updating beliefs. This perspective is the *subjective*, or *Bayesian* view of probability. We thus argue that human rationality, and the coherence of human thought, is defined not by logic, but by probability.

The Bayesian perspective on human reasoning has radical implications. It suggests that the meaning of even the most elementary natural language sentences may have been fundamentally mischaracterized: many such statements may make probabilistic rather than logical

© 2009 Cambridge University Press 0140-525X/09 $40.00

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

claims. And the most elementary aspects of human reasoning may have been misunderstood – what appeared to be logically certain inferences may often instead be better understood as plausible, probabilistic reasoning. Shifting from a logical to a Bayesian perspective entirely changes our predictions concerning the *patterns* of reasoning that we should expect people to exhibit. And experimental work in the psychology of reasoning provides the data against which these predictions can be compared.

This Précis outlines the argument of *BR* chapter by chapter; the section numbering corresponds to the chapter numbering of the book, with occasional modifications to assist the flow of what is now a somewhat compressed argument. The first section of the book, Chapters 1–4, outlines the theoretical background of our shift from logical to Bayesian rationality as an account of everyday human reasoning, drawing on relevant areas of psychology, philosophy, and artificial intelligence. The second section of the book, Chapters 5–7, relates this approach to the key empirical data in the psychology of reasoning: conditional reasoning, Wason's selection task, and syllogistic reasoning. We argue that the patterns of results observed in the empirical data consistently favour a Bayesian analysis, even for purportedly paradigmatically "logical" reasoning problems. Chapter 8 reflects on the implications of this approach.

1. Logic and the Western conception of mind

Since the Greeks, the analysis of mind has been deeply entwined with logic. Indeed, the study of logical argument and the study of mind have often been viewed as overlapping substantially. One swift route to such a deep connection is to argue that minds are distinctively rational; and that rationality is partly, or perhaps even wholly, characterized by logic. That is, logical relations are viewed primarily as unbreakable, inferential relations between thoughts; and a coherent, intelligible agent must respect such relations. In particular, logic aims to specify inferential relations that hold *with absolute certainty*: logical inference is *truth preserving*, that is, if the premises are true, the conclusion must also be true.

But which inferences *are* absolutely certain? Which can be relied upon to preserve truth reliably? We may feel confident, from our knowledge of science, that, for example, all women are mortal. We might generalize from the mortality of all other living things; or note that even the most long-lived creature will succumb in the heat-death of the universe of the far future. But such considerations, however convincing, do not give the certainty of logic – they depend on contingent facts, and such facts are not themselves certain. Aristotle answered these questions by providing the first logical system: the theory of the syllogism. Syllogisms involve two premises, such as, *All women are people; All people are mortal*. Aristotle argued that these premises imply with absolute certainty that *All women are mortal*.

Logical certainty is more than mere overwhelming confidence or conviction. A logical argument depends purely on its *structure*: thus, Aristotle noted, our logical argument put forth here is of the form *All A are B; All B are C;* therefore, *All A are C*. And this argument is valid whatever *A*, *B*, or *C* stand for; hence there is no appeal to contingent facts of any kind. Aristotle's spectacular discovery was, therefore, that patterns of reliable reasoning could be obtained merely by identifying the *structure* of that reasoning. Logic, then, aims to provide a theory that determines which argument structures are truth-preserving, and which are not. In a very real sense, in a logical inference, if you believe the premises, you already believe the conclusion – the meaning of the conclusion is, somehow, contained in the meaning of the premises. To deny the constraints of logic would thus be *incoherent*, rather than merely mistaken. Thus, logic can be viewed as providing crucial constraints on the thoughts that any rational agent can entertain (Davidson 1984; Quine 1953).

Aristotle's theory of the logical structure of the syllogism proceeded by enumeration: Aristotle identified 64 forms of the syllogism, along with a systematic, though intuitive, approach to deciding which of these syllogisms had a valid conclusion, and if so, what the nature of this conclusion is. For more than two thousand years, Aristotle's theory of the syllogism almost exhausted logical theory—and indeed, Kant considered all logical questions to have been decisively resolved by Aristotle's account, stating: "It is remarkable also, that to the present day, it has not been able to make one step in advance, so that, to all appearance, it [i.e., logic] may be considered as completed and perfect" (Kant 1787/1961, p. 501).

As we have suggested, although Aristotle's logic is defined over patterns of verbally stated arguments (converted from everyday language into the appropriate formal structure), it is nonetheless tempting to view the primary subject matter of logic as thought itself. If the mind is viewed as constituted by rational thought, and logic captures patterns of rational thought, it seems natural to view logic as a central part of psychology. Such was Boole's perspective, in going beyond Aristotle's enumeration of patterns of logical argument. Boole aimed to describe the "The Laws of Thought" (Boole

MIKE OAKSFORD is Professor of Psychology and Head of the School of Psychology at Birkbeck College, London. He is the author of over one hundred scientific publications in psychology, philosophy, and cognitive science, and has written or edited six books. He currently serves on the Editorial Boards of the journals *Psychological Review*, *Memory and Cognition*, and *Thinking and Reasoning*. He has also served as Associate Editor for the *Quarterly Journal of Experimental Psychology* and on the Editorial Board of the *Journal of Experimental Psychology: Learning, Memory, & Cognition*. His research explores the nature of reasoning, argumentation, and the influence of emotion on cognition.

NICK CHATER is Professor of Cognitive and Decision Sciences at University College London. He is the author of over one hundred and fifty scientific publications in psychology, philosophy, linguistics, and cognitive science, and has written or edited eight books. He currently serves as Associate Editor for *Psychological Review*. He has also served as Associate Editor for *Cognitive Science* and on the Editorial Boards of *Psychological Review* and *Trends in Cognitive Sciences*. His research explores formal models of inference, choice, and language.

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

1854/1958); and, in doing so, provided, for the first time, explicit mathematical rules for logical reasoning. This allowed him to develop a *calculus* for logical reasoning, albeit limited in scope. Boole also opened up the possibility that logical reasoning could be carried out mechanistically, purely by the manipulation of logical symbols. This insight provided a partial foundation for modern computation and, by extension, cognitive science.

The view that rational thought is governed by logic, which we term the *logicist* conception of the mind (Oaksford & Chater 1991), was adopted wholeheartedly by early cognitive theorists such as Piaget (e.g., Inhelder & Piaget 1955). Piaget viewed the pinnacle of cognitive development as attaining the "formal operational" stage, at which point the mind is capable of reasoning according to a particular formal system of logic: propositional logic. He viewed the process of cognitive development as a series of stages of enrichment of the logical apparatus of the child, enabling increasingly abstract reasoning, which is less tied to the specific sensory-motor environment. Similarly, the early foundations of cognitive science and artificial intelligence involved attempting to realize logical systems practically, by building computer programs that can explicitly derive logical proofs. Tasks such as mathematical reasoning and problem solving were then viewed as exercises in logic, as in Newell and Simon's *Logic Theorist* and *General Problem Solver* (see Newell & Simon 1972; Newell et al. 1958). Moreover, Chomsky's (1957; 1965) revolutionary work in linguistics showed how the syntactic structure of language could be organized in a deductive logical system, from which all and only the grammatical sentences of the language could be generated. And in the psychology of adult reasoning, this logical conception of mind was again used as the foundation for explaining human thought.

Simultaneous with the construction of the logicist program in cognition, there were some discordant and puzzling observations. Specifically, researchers such as Wason, who attempted to verify the Piagetian view of the adult mind as a perfect logic engine, found that people appeared surprisingly and systematically illogical in some experiments. Given the dissonance between these results and the emerging logicist paradigm in cognitive science, these results were largely set aside by mainstream cognitive theorists, perhaps to be returned to once the logicist approach had reached a more developed state. But the general form that an account of apparent irrationality might take was that all illogical performance resulted from misunderstandings and from the faulty way in which the mind might sometimes apply logical rules. For example, Henle stated: "I have never found errors which could unambiguously be attributed to faulty reasoning" (Henle 1978, p. xviii). But the central notion that thought is based on logic was to be retained.

This fundamental commitment to logic as a foundation for thought is embodied in contemporary reasoning theory in two of the main theoretical accounts of human reasoning. The *mental logic* view (Braine 1978; Rips 1983; 1994) assumes that human reasoning involves logical calculation over symbolic representations, using systems of proof which are very similar to those developed by Hilbert in mathematics, and used in computer programs for theorem-proving in artificial intelligence and computer science. By contrast, the *mental models* view (Johnson-Laird 1983; Johnson-Laird & Byrne 1991) takes its starting point as the denial of the assumption that reasoning involves formal operations over logical formulae, and instead assumes that people reason over concrete representations of situations or "models" in which the formulae are true. This provides a different method of proof (see Oaksford & Chater 1991; 1998a, for discussion), but one that can achieve logical performance by an indirect route.

Although mental logic and mental models both give logic a central role in human reasoning, they explain apparent irrationalities in different ways. For example, mental logics may explain errors in terms of the accessibility of different rules, whereas mental models explain errors in terms of limitations in how mental models are constructed and checked, and how many models must be considered.

These logicist reactions to data appearing to show human irrationality seem entirely reasonable. Every new theory in science could be immediately refuted if the mere existence of data apparently inconsistent with the theory were assumed to falsify it decisively (Kuhn 1962; Lakatos 1970). The crucial question is: Can a more plausible explanation of these puzzling aspects of human reasoning be provided? We argue that the Bayesian approach provides precisely such an alternative.

2. Rationality and rational analysis

BR aims to promote a Bayesian, rather than a logical, perspective on human reasoning. But to make sense of any debate between the logical and Bayesian standpoints, we need first to clarify how we interpret the relationship between a normative mathematical theory of reasoning (whether logic or probability), and empirical findings about human reasoning. In particular, how do we deal with any systematic clashes between the theory's dictates concerning how people *ought* to reason, and empirical observations of how they actually *do* reason?

Various viewpoints have been explored. One option is to take observed human intuitions as basic, and hence as the arbiter of what counts as a good formal theory of reasoning (e.g., Cohen 1981). Another is to take the mathematical theory as basic, and view it as providing a standpoint from which to evaluate the quality of observed reasoning performance (e.g., Rips 1994). Still a further possibility is that clashes between the formal theory and actual reasoning may arise because human thought itself is divided between two systems of reasoning (e.g., Evans & Over 1996a).

Here, we take a different line: We view normative theory as a component of the project of providing a "rational analysis" which aims to capture empirical data concerning thought and behavior. Rational analysis (e.g., Anderson 1990; 1991a; Oaksford & Chater 1998b) has six steps:

1. Specify precisely the goals of the cognitive system.
2. Develop a formal model of the environment to which the system is adapted.
3. Make minimal assumptions about computational limitations.
4. Derive the optimal behaviour function given steps 1–3.

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

(This requires formal analysis using rational norms, such as probability theory, logic, or decision theory.)

5. Examine the empirical evidence to see whether the predictions of the behaviour function are confirmed.

6. Repeat, iteratively refining the theory.

So the idea of rational analysis is to understand the problem that the cognitive system faces, and the environmental and processing constraints under which it operates. Behavioral predictions are derived from the assumption that the cognitive system is solving this problem, optimally (or, more plausibly, approximately), under these constraints. The core objective of rational analysis, then, is to understand the structure of the problem *from the point of view of the cognitive system*, that is, to understand what problem the brain is attempting to solve.

In the psychology of reasoning, this point is particularly crucial. We shall see that even when the experimenter intends to confront a participant with a logical reasoning puzzle, the participant may interpret the problem in probabilistic terms. If so, the patterns of reasoning observed may be well described in a Bayesian framework, but will appear to be capriciously errorful from a logical point of view. In Chapters 5–7 of *BR*, and summarized further on here, we argue that the core data in the psychology of reasoning, which has focussed on putatively "logical" reasoning tasks, can be dramatically clarified by adopting a Bayesian rational analysis.

It might appear that Step 2, concerning the environment, could not be relevant to rational analysis of the reasoning, as opposed to, say, perception. Mathematical theories of reasoning are supposed to apply across topics, and hence should surely be independent of environmental structure. We shall see further on that the reverse is the case. Very general features of the environment, such as the fact that almost all natural language categories occur with a low probability and that arbitrarily chosen probabilistic constraints are often independent or nearly independent, turn out to have substantial implications for reasoning. Indeed, the project of providing a rational analysis of human reasoning gains its empirical purchase precisely by explaining how a "topic neutral" mathematical theory applies to a specific goal, given a particular set of environmental and computational constraints.

Two caveats are worth entering concerning Bayesian rational analysis. The first is that rational analysis is not intended to be a theory of psychological *processes*. That is, it does not specify the representations or algorithms that are used to carry out this solution. Indeed, as Anderson (1990; 1991a) points out, these representations and algorithms might take many different forms – but certain general aspects of their behavior will follow irrespective of such specifics; they will arise purely because the cognitive system is well-adapted to solving this particular problem. Hence, the correct analysis of the rational structure of the cognitive problem at hand can have considerable explanatory power.

The second caveat is that the aim of understanding the structure of human reasoning, whether from a logical or a Bayesian perspective, should be carefully distinguished from the goal of measuring people's performance on logical or probabilistic problems (Evans et al. 1993; Kahneman et al. 1982). Indeed, both logic and probability provide a fresh and bracing challenge to each generation of students; performance on logical and probability problems results from explicit instruction and study, rather than emerging from capacities that are immanent within the human mind. But this observation need not impact our evaluation of logic or probability as explanations for patterns of everyday thought. Even if the mind is a probabilistic or logical calculating engine, it may not be possible to engage that engine with verbally, symbolically, or numerically stated probabilistic or logical puzzles, which it is presumably not adapted to handle. This point is no deeper than the observation that, although the early visual processes in the retina may compute elaborate convolutions and decorrelations of the image, this does not mean that people can thereby readily apply this machinery to solve mathematics problems concerning convolution or decorrelation. Thus, empirical evidence from the psychology of reasoning is not used, in *BR*, to evaluate people's logical or probabilistic reasoning competence. Rather, this evidence is used to explore the patterns of reasoning that people find natural; and to relate such patterns to how people reason outside the experimental laboratory.

From the standpoint of rational analysis, the question of whether logic or probability is the appropriate framework for understanding reasoning is an empirical question: Which rational analysis of human reasoning best captures the data? In Chapters 5–7 of *BR*, we argue, case-by-case, that a Bayesian rational analysis provides a better account of core reasoning data than its logicist rivals. First, though, we consider why. In Chapter 3, we argue that real-world, informal, everyday, reasoning is almost never deductive, that is, such reasoning is almost always logically *in*valid. In Chapter 4, we consider what has driven the broader "probabilistic turn" in cognitive science and related fields, of which the Bayesian analysis of human reasoning is a part.

3. Reasoning in the real world: How much deduction is there?

Logic provides a calculus for certain reasoning – for finding conclusions which follow, of necessity, from the premises given. But in everyday life, people are routinely forced to work with scraps of knowledge, each of which may be only partially believed. Everyday reasoning seems to be more a matter of tentative conjecture, rather than of water-tight argument.

Notice, in particular, that a successful logical argument cannot be overturned by any additional information that might be added to the premises. Thus, if we know that *All people are mortal*, and *All women are people*, then we can infer, with complete certainty, that *All women are mortal*. Of course, on learning new information we may come to doubt the premises – but we cannot come to doubt that the conclusion follows from the premises. This property of classical logic is known as *monotonicity*, meaning that adding premises can never overturn existing conclusions.

In reasoning about the everyday world, by contrast, *non*-monotonicity is the norm: almost any conclusion can be overturned, if additional information is acquired. Thus, consider the everyday inference from *It's raining* and *I am about to go outside* to *I will get wet*. This inference

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

is uncertain – indefinitely many additional premises (*the rain is about to stop*; *I will take an umbrella*; *there is a covered walkway*) can overturn the conclusion, even if the premises are correct. The nonmonotonicity of everyday inference is problematic for the application of logical methods to modelling thought. Nonmonotonic inferences are not logically valid and hence fall outside the scope of standard logical methods.

The nonmonotonicity of everyday reasoning often strikes in subtle and unexpected ways. Most notorious is the "frame problem" (McCarthy & Hayes 1969), which arose in early applications of logical methods in artificial intelligence. Suppose an agent, with knowledge base *K*, makes an action *A* (e.g., it turns a cup upside down). Which other information in *K* needs to be updated to take account of this action? Intuitively, almost all other knowledge should be unchanged (e.g., that the street is empty, or that the burglar alarm is off). But, from a logical point of view, the "interia" of such everyday knowledge does not follow, because it is logically possible that *A* may have all manner of consequences. For example, given the additional information that the cup is valuable and placed in an alarmed glass case, then turning it over *may* trigger the burglar alarm and may fill the street with curious bystanders. The difficulties generated by the frame problem have had a paralyzing effect on logical approaches to planning, action, and knowledge representation in artificial intelligence.

Analogous problems arise more generally (Fodor 1983; Pylyshyn 1987). Given a database with knowledge *K*, adding a new fact *F* (not necessarily concerning an action) can typically overthrow many of the previous consequences of *K*, in highly idiosyncratic ways. It proves to be impossible to delimit the inferential consequences of a new fact in advance. Learning a new fact about football can, for example, readily modify my beliefs about philosophy. For example, suppose one has been told footballing facts and philosophical facts by the same person, of uncertain trustworthiness. Then learning that a footballing fact is incorrect may cause one to doubt a putative philosophical fact. Thus, nonmonotonicty may apply to arbitrarily remote pieces of knowledge. And note, of course, that an inference that can be overturned by additional premises cannot be logically valid – because standard logic is monotonic by definition.

Inferences which are nonmonotonic, and hence cannot be captured by conventional logic, are described in different literatures using a variety of terms: *non-demonstrative* inference, *informal* argument, and *common-sense* reasoning. For the purposes of our arguments, these terms are interchangeable. But their import, across psychology, artificial intelligence, and philosophy, is the same: nonmonotonic arguments are outside the scope of deductive logic.

This conclusion has alarming implications for the hypothesis that thought is primarily based on logical inference. This is because the scope of monotonic inference is vanishingly small – indeed, it scarcely applies anywhere outside mathematics. As we shall see in Chapters 5–7, this point applies even to verbally stated inferences that are typically viewed as instances of deduction. For example, consider the argument from *if you put 50p in the coke machine, you will get a coke* and *I've put 50p in the coke machine*, to *I'll get a coke*. This argument appears to be an instance of a canonical monotonic logical inference: *modus ponens*.

Yet in the context of commonsense reasoning, this argument does not appear to be monotonic at all. There are innumerable possible additional factors that may block this inference (power failure, the machine is empty, the coin or the can become stuck, etc.). Thus, you can put the money in, and no can of coke may emerge. Attempting to maintain a logical analysis of this argument, these cases could be interpreted as indicating that, from a logical point of view, the conditional rule is simply false – precisely because it succumbs to counterexamples (Politzer & Braine 1991). But this is an excessively rigorous standpoint, from which almost all everyday conditionals will be discarded as false. But how could a plethora of false conditional statements provide a useful basis for thought and action. From a logical point of view, after all, we can only make inferences from *true* premises; a logical argument tells us nothing, if one or more of its premises are false.

In sum, there appears to be a fundamental mismatch between the nonmonotonic, uncertain character of everyday reasoning, and the monotonicity of logic; and this mismatch diagnoses the fundamental problem with logic-based theories of reasoning and logicist cognitive science more broadly. In *BR*, we draw a parallel with a similar situation in the philosophy of science, where there has been a gradual retreat from early positive claims that theoretical claims somehow logically follow from observable premises, to Popper's (1935/1959) limitation of logical deduction, to the process of drawing predictions from theories, to the abandonment of even this position, in the light of the nonmonotonicity of predictive inference (there are always additional forces, or factors, that can undo any prediction; Putnam 1974). Indeed, modern philosophy of science has taken a resolutely Bayesian turn (e.g., Bovens & Hartmann 2003; Earman 1992; Horwich 1982; Howson & Urbach 1993). *BR* also considers attempts to deal with the apparent mismatch by attempting to deal with uncertainty by developing nonmonotonic logics (e.g., Reiter 1980), a project that rapidly became mired in difficulties (see, e.g., Oaksford & Chater 1991). Perhaps it is time to shift our attention to a calculus that deals directly with uncertainty: probability theory.

4. The probabilistic turn

We have seen how uncertainty, or nonmonotonicity, is a ubiquitous feature of everyday reasoning. Our beliefs, whether arising from perception, commonsense thought, or scientific analysis, are tentative and provisional. Our expectation that the car will start, that the test tube will turn blue, or that one arrow is longer than another, are continually being confounded by faulty batteries, impure chemicals, or visual illusions.

Interestingly, Aristotle, the founder of logic, was keenly aware of the limits of the logical enterprise. After all, he was interested not only in mathematical and philosophical reasoning, but also with the scientific description and analysis of the everyday world, and with practical affairs and human action. An often quoted passage from the *Nicomachean Ethics* (1094b, Aristotle 1980, p. 3) notes that

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

"it is the mark of an educated man to look for precision in each class of things just so far as the nature of the subject admits: it is evidently equally foolish to accept probable reasoning from a mathematician and to demand from a rhetorician demonstrative reasoning."

Indeed, one key motivation for developing a theory of probability was closely connected with Aristotle's rhetorician. The goal in rhetoric, in its traditional sense, is to provide reasoned arguments for why people should hold certain opinions concerning matters about which certainty is impossible. Thus, in deciding court cases by jury, a different piece of evidence (e.g., eye-witness testimony, forensic evidence, evidence of previous good character) must somehow be combined to yield a degree of belief concerning the likely guilt of the defendant. Here, probability is interpreted subjectively, in terms of a person's strength of opinion, rather than concerning an assumption about the external world. Indeed, the very word "probability" initially referred to the degree to which a statement was supported by the evidence at hand (Gigerenzer et al. 1989). Jakob Bernoulli explicitly endorsed this interpretation when he entitled his definitive book *Ars Conjectandi*, or the *Art of Conjecture* (Bernoulli 1713). This *subjectivist*, or *Bayesian*, conception of probability ran through the eighteenth and into the nineteenth centuries (Daston 1988), frequently without clear distinctions being drawn between probability theory as a model of actual thought (or more usually, the thought of "rational", rather than common, people [Hacking 1975; 1990]) or as a set of normative canons prescribing how uncertain reasoning should be conducted. As with logic, early probability theory itself was viewed as a model of mind.

Over the latter part of the twentieth century, the Bayesian perspective has been increasingly influential across the cognitive sciences and related disciplines. Chapter 4 of *BR* surveys some of these developments. For example, if everyday inference is inherently probabilistic, this raises the possibility that natural language statements should be interpreted as making probabilistic, rather than logical, claims. So, for example, Adams (e.g., 1975; 1998) directly imports probability into logical theory, arguing that the conditional *If A then B* should, roughly, be interpreted as saying that *B* is probable, if *A* is true. Later we shall see how this, and other probabilistic analyses of familiar "logical" structures (e.g., concerning the quantifiers *All*, *Some*, etc.), cast new light on the empirical reasoning data.

It is, we suggest, significant that three key domains in which uncertain inference is ubiquitous, philosophy of science, artificial intelligence, and cognitive psychology, have all embraced the Bayesian approach. *BR* reviews some of the key developments: the application of Bayes' theorem to hypothesis confirmation (e.g., Earman 1992); the development of graphical models for knowledge representation and causal reasoning (Pearl 1988; 2000); and the application of Bayesian methods in rational models of cognitive phenomena (Chater & Oaksford 2008b; Oaksford & Chater 1998b) in areas as diverse as categorization (Anderson 1991b; Anderson & Matessa 1998), memory (Anderson & Milson 1989; Anderson & Schooler 1991), conditioning (Courville et al. 2006; Kakade & Dayan 2002), causal learning (Griffiths & Tenenbaum 2005; Novick & Cheng 2004), natural language processing (Chater et al. 1998; Chater & Manning 2006), and vision (Knill & Richards 1996; Yuille & Kersten 2006).

There has, in short, been a "probabilistic turn" across a broad range of domains – a move away from the attempt to apply logical methods to uncertain reasoning, and towards dealing with uncertainty by the application of probability theory. In Chapters 5–7, we illustrate how the switch from logical to Bayesian rationality leads to a radical re-evaluation of the psychology of human reasoning – so radical, in fact, that even apparently paradigmatic "logical" reasoning tasks turn out to be better understood from a probabilistic point of view.

5. Does the exception prove the rule? How people reason with conditionals

In Chapters 5–7 of *BR*, we describe Bayesian probabilistic models for the three core areas of human reasoning research: conditional inference (Ch. 5), data selection (Ch. 6), and quantified syllogistic inference (Ch. 7). The key idea behind all these models is to use conditional probability, $P(q|p)$, to account for the meaning of conditional statements, *if p then q* (e.g., *if you turn the key then the car starts*). The aim is to show that what appear to be "errors and biases" from a logicist standpoint are often entirely rational from a Bayesian point of view. In this Précis, for each area of reasoning, we introduce the task, the standard findings, and existing logicist accounts. We then introduce a Bayesian rational analysis for each problem, show how it accounts for the core data, and provide a snapshot of some of the further data that we discuss in *BR*. Finally, for each area of reasoning, we summarise and describe one or two outstanding problems confronting the Bayesian approach.

Chapter 5 of *BR* begins with conditional inference, that is, inferences directly involving the conditional *if p then q*. In the conditional, p is called the "antecedent" and q is called the "consequent." Four inference patterns have been extensively studied experimentally (see Fig. 1). Each inference consists of the *conditional* premise and one of four possible *categorical* premises, which relate either to the antecedent or consequent of the conditional, or their negations (p, $\neg p$, q, $\neg q$ where "$\neg$" = not). For example, the inference *Modus Ponens* (MP) combines the conditional premise *if p then q* with the categorical premise p; and yields the conclusion q.

According to standard logic, two of these inferences are logically valid (MP and *Modus Tollens* [MT], see Fig. 1), and two are fallacies (Denying the Antecedent [DA] and

$$\text{(MP)}\ \frac{p \Rightarrow q,\ p}{\therefore q} \qquad \text{(MT)}\ \frac{p \Rightarrow q,\ -q}{\therefore \neg p}$$

$$\text{(DA)}\ \frac{p \Rightarrow q,\ -p}{\therefore \neg q} \qquad \text{(AC)}\ \frac{p \Rightarrow q,\ q}{\therefore p}$$

Figure 1. The four inference patterns investigated in the psychology of conditional inference: *Modus Ponens* (MP) and *Modus Tollens* (MT) are logically valid. *Denying the Antecedent* (DA) and *Affirming the Consequent* (AC) are logically fallacious. These inference schemata read that if the premises above the line are true then so must be the conclusion below the line. "$p \Rightarrow q$" signifies the "material conditional" of standard logic, which is true unless p is true and q is false.

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

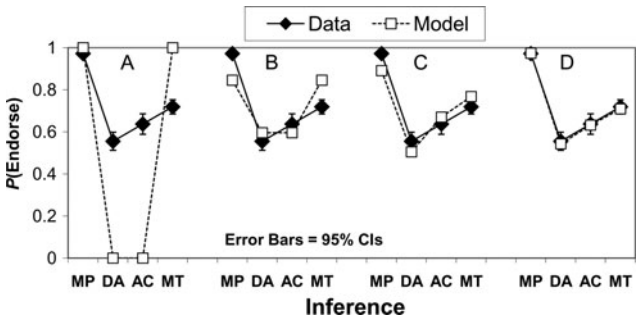


Figure 2. The fits to the experimental data (Schroyens & Schaeken 2003) of standard logic (Panel A), standard logic plus the biconditional interpretation and error (Panel B), the original probabilistic model (Panel C), and the probabilistic model adjusted for rigidity violations (Panel D).

Affirming the Consequent [AC], see Fig. 1). Figure 2 (Panel A) shows the results of a meta-analysis of experiments where people are asked whether they endorse each of these four inferences (Schroyens & Schaeken 2003). Panel A also shows the predictions of the standard logical model, revealing a large divergence.

From a logicist standpoint, this divergence may be reduced by assuming that some people interpret the conditional as a *biconditional*, that is, that *if p then q* also means that *if q then p*. This move from conditional to biconditional is, of course, logically invalid. For example, *if a bird is a swan, then it is white* clearly does not entail that *if a bird is white, then it is a swan*. Nonetheless, the biconditional interpretation may be pragmatically reasonable, in some cases. For example, promises such as *if you mow the lawn, I will pay you* £5 do seem to allow this pragmatic inference; it seems reasonable to assume that I will *only* pay you £5 if you mow the lawn (or, at least, that I will not pay you £5 if you refuse). By assuming that people make this pragmatic inference for the stimuli used in experimental tasks and by making some allowance for random error, the best fit that standard logic can provide is shown in Figure 2 (Panel B) (see Oaksford & Chater 2003a).

To further close the gap with the data in Figure 2, logicist theories of conditional inference typically assume not only that people adopt the pragmatic inference to the biconditional interpretation, but also that they fail to represent logic completely in their cognitive system. For example, mental logic (e.g., Rips 1994) is typically assumed to involve an MP inference rule, but no MT rule. This means that MT inferences must be drawn in a more complex way, often leading to error. Similarly, according to mental models theory, people do not initially represent the full meaning of the conditional (Johnson-Laird & Byrne 2002). To draw an MT inference, they must "flesh out" their representations to fully capture the meaning of the conditional. In both cases, logically unwarranted pragmatic inferences and assumptions about cognitive limitations are invoked to explain the data.

In contrast, the Bayesian approach only invokes probability theory. There are four key ideas behind the probabilistic account of conditional inference. First, the probability of a conditional is the conditional probability, that is, $P(\textit{if } p \textit{ then } q) = P(q|p)$. In the normative literature, this identification is simply called "The Equation" (Adams 1998; Bennett 2003; Edgington 1995). In the psychological literature, the Equation has been confirmed experimentally by Evans et al. (2003) and by Oberauer and Wilhelm (2003). Second, as discussed earlier, probabilities are interpreted "subjectively," that is, as degrees of belief. It is this interpretation of probability that allows us to provide a probabilistic theory of inference as belief updating. Third, conditional probabilities are determined by a psychological process called the "Ramsey Test" (Bennett 2003; Ramsey 1931/1990b). For example, suppose you want to evaluate your conditional degree of belief that *if it is sunny in Wimbledon, then John plays tennis*. By the Ramsey test, you make the hypothetical supposition that *it is sunny in Wimbledon* and revise your other beliefs so that they fit with this supposition. You then "read off" your hypothetical degree of belief that *John plays tennis* from these revised beliefs.

The final idea concerns standard conditional inference: how we reason when the categorical premise is not merely *supposed*, but is actually believed or known to be true. This process is known as *conditionalization*. Consider an MP inference, for example, *If it is sunny in Wimbledon, then John plays tennis*, and *It is sunny in Wimbledon*, therefore, *John plays tennis*. Conditionalization applies when we know (instead of merely supposing) that *it is sunny in Wimbledon*; or when a high degree of belief can be assigned to this event (e.g., because we know that it is sunny in nearby Bloomsbury). By conditionalization, our new degree of belief that *John plays tennis* should be equal to our prior degree of belief that *if it is sunny in Wimbledon, then John plays tennis* (here "prior" means before learning that *it is sunny in Wimbledon*). More formally, by the Equation, we know that $P_0(\textit{if it is sunny in Wimbledon, then John plays tennis})$ equals $P_0(\textit{John plays tennis}|\textit{it is sunny in Wimbledon})$, where "$P_0(x)$" = prior probability of x. When we learn *it is sunny in Wimbledon*, then $P_1(\textit{it is sunny in Wimbledon}) = 1$, where "$P_1(x)$" = posterior probability of x. Conditionalizing on this knowledge tells us that our new degree of belief in *John plays tennis*, $P_1(\textit{John plays tennis})$, should be equal to $P_0(\textit{John plays tennis}|\textit{it is sunny in Wimbledon})$. That is, $P_1(q) = P_0(q|p)$, where $p = \textit{it is sunny in Wimbledon}$, and $q = \textit{John plays tennis}$.[1] So from a probabilistic perspective, MP provides a way of updating our degrees of belief in the consequent, q, on learning that the antecedent, p, is true.

So, quantitatively, if you believe that $P_0(\textit{John plays tennis}|\textit{it is sunny in Wimbledon}) = 0.9$, then given you discover that *it is sunny in Wimbledon* ($P_1(\textit{it is sunny in Wimbledon}) = 1$) your new degree belief that *John plays tennis* should be 0.9, that is, $P_1(\textit{John plays tennis}) = 0.9$. This contrasts with the logical approach in which believing the conditional premise entails with *certainty* that the conclusion is true, so that $P_0(\textit{John plays tennis}|\textit{it is sunny in Wimbledon}) = 1$. This is surely too strong a claim.

The extension to the other conditional inferences is not direct, however. Take an example of AC, *if it is sunny in Wimbledon, John plays tennis* and *John plays tennis*, therefore, *it is sunny in Wimbledon*. In this case, one knows or strongly believes that *John play tennis* (perhaps we were told by a very reliable source), so $P_1(q) = 1$. But to use Bayesian conditionalization to infer one's new degree of belief that *it is sunny in Wimbledon*, $P_1(p)$, one needs to know one's conditional degree of belief

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

that *it is sunny in Wimbledon* given *John plays tennis*, that is, $P_0(p|q)$. However, the conditional premise of AC, like that of MP, is about $P_0(q|p)$ *not* about $P_0(p|q)$ (Sober 2002). The solution proposed by Oaksford et al. (2000; see also Wagner 2004) is that that people also know the prior marginal probabilities (at least approximately). That is, they know something about the probability of a sunny day in Wimbledon, $P_0(p)$, and the probability that John plays tennis, $P_0(q)$, *before* learning that it is in fact a sunny day in Wimbledon. With this additional information, $P_0(p|q)$ can be calculated from the converse conditional probability, $P_0(q|p)$, using Bayes' Theorem.[2] The same approach also works for DA and MT where the relevant probabilities are $P_0(\neg q|\neg p)$ and $P_0(\neg p|\neg q)$, respectively. The fact that the conditional premises of AC, DA, and MT do not determine the appropriate conditional probability marks an important asymmetry with MP. For these inferences, further knowledge is required to infer the relevant conditional degrees of belief.

The rest of Chapter 5 in *BR* shows how the errors and biases observed in conditional inference are a consequence of this rational probabilistic model. The first set of "biases" relates directly to the data in Figure 2. These are what, in *BR*, we call "the inferential asymmetries." That is, MP is drawn more than MT and AC is drawn more than DA (MT is also drawn more than AC). Figure 2, Panel C shows how well a probabilisitic account can explain these asymmetries. Here we have calculated the values of $P_0(q|p)$, $P_0(p)$, and $P_0(q)$ that best fit the data, that is, they minimize the sum of squared error between the data and the models predictions ("model" in Fig. 2). As Panel C shows, a probabilistic account can capture the asymmetries without pragmatic inference or appeal to process limitations. Panel C also shows, however, that this probabilistic model (Oaksford et al. 2000) does not capture the magnitudes of the inferential asymmetries (Evans & Over 2004; Schroyens & Schaeken 2003). It underestimates the MP – MT asymmetry and overestimates the DA – AC asymmetry.

In *BR*, we argue that this is because learning that the categorical premise is true can have two inferential roles. The first inferential role is in conditionalization, as we have described. The second inferential role is based on the pragmatic inference that *being told that the categorical premise is true* often suggests that there is a counterexample to the conditional premise. For example, consider the MT inference on the rule: *If I turn the key, the car starts*. If you were told that *the car did not start*, it seems unlikely that you would immediately infer that the *key was not turned*. Telling someone that *the car did not start* seems to presuppose that an attempt has been made to start it, presumably by turning the key. Consequently, the categorical premise here seems to suggest a counterexample to the conditional itself, that is, a case where the key was turned but the car did not start. Hence, one's degree of belief in the conditional should be reduced on *being told* that the car did not start. Notice, here, the contrast between being told that the car did not start (and drawing appropriate pragmatic inferences), and merely *observing* a car that has not started (e.g., a car parked in the driveway). In this latter situation, it is entirely natural to use the conditional rule to infer that the key has not been turned.

Where the second, pragmatic, inferential role of the categorical premise is operative, this violates what is called the *rigidity condition* on conditionalization, $P_0(q|p) = P_1(q|p)$ (Jeffrey 1983). That is, learning the categorical premise alters one's degree of belief in the conditional premise. In *BR*, we argue that taking account of such rigidity violations helps capture the probability of the conditional; and that, for MT this modified probability is then used in conditionalization. Furthermore, we argue that DA and AC also suggest violations of the rigidity condition, concerning the case where the car starts without turning the key. These violations lead to reductions in our degree of belief that the cars starts, given that the key is turned ($P_0(q|p)$). Using this lower estimate to calculate the relevant probabilities for DA, AC, and MT can rationally explain the relative magnitudes of the MP – MT and DA – AC asymmetries (see Fig. 2, Panel D).

We now turn to one of the other biases of conditional inference that we explain in Chapter 5 of *BR*: *negative conclusion* bias. This bias arises when negations are used in conditional statements, for example, *If a bird is a swan, then it is not red*. In Evans' (1972) *Negations Paradigm*, four such rules are used: *If p then q*; *if p then not-q*; *if not-p then q*; *if not-p then not-q*. The most robust finding is that people endorse DA, AC, and MT more when the conclusion contains a negation (see Fig. 3). So, for example, DA on *if p then q* (see Panel A in Fig. 3) yields a negated conclusion, *not-q*. Whereas, DA on *if p then not-q* (see Panel B in Fig. 3) yields an affirmative conclusion, *q* (because *not-not-q* = *q*). In Figure 3, it is clear that the frequency with which DA is endorsed for *if p then q* is much higher than for *if p then not-q*.

To explain negative conclusion bias, we appeal to the idea that most categories apply only to a minority of objects (Oaksford & Stenning 1992). Hence, the probability of an object being, say, red is lower than the probability of it not being red, that is, $P_0(Red) < P_0(\neg Red)$. Consequently, the marginal probabilities ($P_0(p)$ and $P_0(q)$) will take on higher values when *p* or *q* are negated. Higher values of the prior probabilities of the conclusion imply higher values of the relevant conditional probabilities for DA, AC, and MT, that is, to higher values of the posterior probability of the conclusion. So, for example, for our rule *if a bird is a swan, then it is white*, the prior probability of the conclusion of the DA inference ($P_0(\neg White)$) is high. This means that the conditional probability ($P_0(\neg White|\neg$

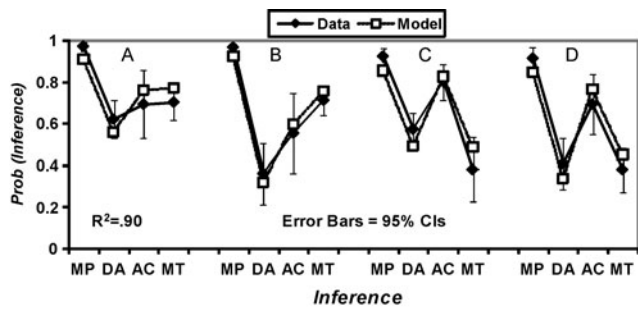


Figure 3. The results of Oaksford et al.'s (2000) meta-analysis of the negations paradigm conditional inference task for *if p then q* (Panel A), *if p then* $\neg q$ (Panel B), *if* $\neg p$ *then q* (Panel C), and *if* $\neg p$ *then* $\neg q$ (Panel D), showing the fit of the original probabilistic model.

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

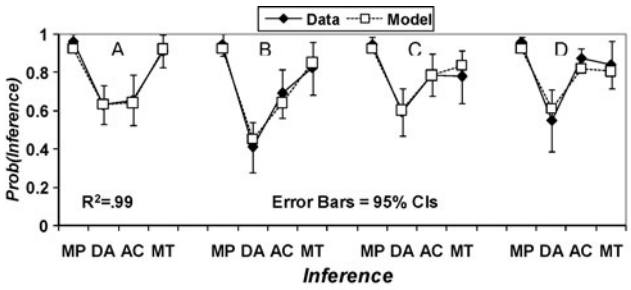


Figure 4. The results of Oaksford et al.'s (2000) Experiment 1 for the *low* $P(p)$, *low* $P(q)$ rule (Panel A), the *low* $P(p)$, *high* $P(q)$ rule (Panel B), the *high* $P(p)$, *low* $P(q)$ rule (Panel C), and the *high* $P(p)$, *high* $P(q)$ rule (Panel D), showing the fit of the original probabilistic model.

Swan)) is also high and, consequently, so is the probability of the conclusion ($P_1(\neg White)$). Therefore, an apparently irrational negative conclusion bias can be seen as a rational "high probability conclusion" effect. Oaksford et al. (2000) tested this explanation by manipulating directly $P_0(p)$ and $P_0(q)$ rather than using negations and showed results closely analogous to negative conclusion bias (see Fig. 4).

To conclude this section on conditional inference, we briefly review one of the most cited problems for a probabilistic account. Like any rational analysis, this account avoids theorising about the specific mental representations or algorithms involved in conditional reasoning. This may seem unsatisfactory. We suggest, by contrast, that it is premature to attempt an algorithmic analysis. The core of our approach interprets conditionals in terms of conditional probability, that is, using the Equation; and our current best understanding of conditional probability is given by the Ramsey test (Bennett 2003). But there is currently no possibility of building a full algorithmic model to carry through the Ramsey test, because this involves solving the notorious frame problem, discussed in Chapter 3. That is, it involves knowing how to update one's knowledge-base, in the light of a new piece of information – and this problem has defied 40 years of artificial intelligence research.

Nonetheless, an illustrative small-scale implementation of the Ramsey test is provided by the operation of a constraint satisfaction neural network (Oaksford 2004a). In such a model, performing a Ramsey test means clamping on or off the nodes or neurons corresponding to the categorical premise of a conditional inference. Network connectivity determines relevance relations and the weight matrix encodes prior knowledge. Under appropriate constraints, such a network can be interpreted as computing true posterior probabilities (McClelland 1998). A challenge for the future is to see whether such small-scale implementations can capture the full range of empirically observed effects in conditional inference.

6. Being economical with the evidence: Collecting data and testing hypotheses

Chapter 6 of *BR* presents a probabilistic model of Wason's selection task. In this task, people see four double-sided cards, with a number on one side and a letter on the other. They are asked which cards they should turn over, in order to test the hypothesis that *if there is an A* (p) *on one side of a card, then there is a 2* (q) *on the other*. The upturned faces of the four cards show an A (p), a K ($\neg p$), a 2 (q), and a 7 ($\neg q$) (see Fig. 5). The typical pattern of results is shown in Figure 6 (Panel A, Data).

As Popper (1935/1959) argued, logically one can never be *certain* that a scientific hypothesis is true in the light of observed evidence, as the very next piece of evidence one discovers could be a counterexample. So, just because all the swans you have observed up until now have been white, is no guarantee that the next one will not be black. Instead, Popper argues that the only logically sanctioned strategy for hypothesis testing is to seek *falsifying* cases. In testing a conditional rule *if p then q*, this means seeking out p, $\neg q$ cases. This means that, in the standard selection task, one should select the A (p) and the 7 ($\neg q$) cards, because these are the only cards that could potentially falsify the hypothesis. Figure 6 (Panel A, Model) shows the logical prediction, and, as for conditional inference, the divergence from the data is large. Indeed, rather than seek *falsifying* evidence, participants seem to select the cases that *confirm* the conditional (p and q). This is called "confirmation bias."

The range of theories of the selection task parallels the range of accounts of the conditional inference task described earlier. Mental logic theories (e.g., Rips 1994) assume that people attempt to perform conditional inferences, using the upturned face as the categorical premise to infer what is on the hidden face. Again, a biconditional interpretation is invoked: that *if A then 2* may pragmatically imply *if 2 then A*. If people perform an MP inference on both conditionals, this will yield a confirmatory response pattern. To infer that the 7 card should be turned, involves considering the hidden face. If people consider the possibility that the hidden face is not an A, then the complex inference pattern required for MT can be applied. A problem for mental logic is that, on this explanation, selection task performance

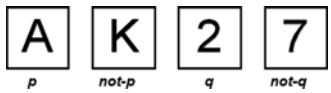


Figure 5. The four cards in Wason's Selection Task.

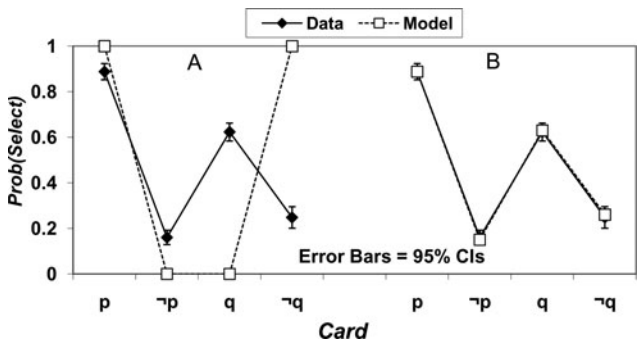


Figure 6. The fits to the experimental data on the Wason Selection Task of standard logic (Panel A) and of the optimal data selection model (Oaksford & Chater 1994) (Panel B).

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

should look like conditional inference task performance where selecting the 2 (q) card corresponds to AC and selecting the 7 ($\neg q$) card corresponds to MT. However, in conditional inference, MT is endorsed more than AC, but in the selection task this is reversed, that is, q (AC) is selected more than $\neg q$ (MT).[3] For mental models, similar predictions are made if people initially represent the conditional as a biconditional and do not "flesh out" this representation.

The optimal data selection (ODS) model of this task (Oaksford & Chater 1994; 1996; 2003b) is a rational analysis derived from the normative literature on optimal experimental design in Bayesian statistics (Lindley 1956). The idea again relies on interpreting a conditional in terms of conditional probability. For example, the hypothesis, *if swan* (p) *then white* (p), is interpreted as making the claim that the probability of a bird being white given that it is a swan, $P(q|p)$, is high, certainly higher than the base rate of being a white bird, $P(q)$. This hypothesis is called the *dependence* hypothesis (H_D). Bayesian hypothesis testing is comparative rather than exclusively concentrating on falsification. Specifically, in the ODS model, it is assumed that people compare H_D with an *independence* hypothesis (H_I) in which the probability of a bird being white, given it is a swan, is the same as the base rate of a bird being white, that is, $P(q|p) = P(q)$. We assume that, initially, people are maximally uncertain about which hypothesis is true ($P(H_D) = P(H_I) = 0.5$) and that their goal in selecting cards is to reduce this uncertainty as much as possible while turning the fewest cards.

Take, for example, the card showing *swan* (p). This card could show *white* on the other side (p, q) or another color (p, $\neg q$). The probabilities of each outcome will be quite different according to the two hypotheses. For example, suppose that the probability of a bird being white, given that it is a swan is 0.9 ($P(q|p, H_D) = 0.9$) in the dependence hypothesis; the marginal probability that a bird is swan is 0.2 ($P(p) = 0.2$); and the marginal probability that a bird is white is 0.3 ($P(q) = 0.3$). Then, according to the dependence hypothesis, the probability of finding *white* (q) on the other side of the card is 0.9, whereas according to the independence hypothesis it is 0.3 (as the antecedent and consequent are, in this model, independent, we need merely consult the relevant marginal probability). And, according to the dependence hypothesis, the probability of finding a color other than white ($\neg q$) on the other side of the card is 0.1, whereas according to the independence hypothesis it is 0.7. With this information it is now possible to calculate one's new degree of uncertainty about the dependence hypothesis after turning the *swan* card to find *white* on the other side ($P(H_D|p, q)$). According to Bayes' theorem (see Note 2), this probability is 0.75. Hence, one's new degree of belief in the dependence model should be 0.75 and one's degree of belief in the independence model should be 0.25. Hence, the degree of uncertainty about which hypothesis is true has been reduced. More specifically, the ODS model is based on *information gain*, where information is measured in *bits* as in standard communication theory. Here, the initial uncertainty is 1 bit (because $P(H_D) = P(H_I) = 0.5$, equivalent to the uncertainty of a single fair coin flip) and in this example this is reduced to 0.81 bits (because now $P(H_D) = 0.75$ and $P(H_I) = 0.25$). This is an *information gain* of 0.19 bits.

In Wason's task, though, participants do not actually turn the cards, and hence they cannot know how much information they will gain by turning a card before doing so. Consequently, they must base their decision on *expected* information gain, taking both possible outcomes (p, q and p, $\neg q$) into account. The ODS model assumes that people select each card in direct proportion to its expected information gain.

The ODS model also makes a key assumption about the task environment – that is, Step 2, in rational analysis: The properties that occur in the antecedents and consequents of hypotheses are almost always rare and so have a low base rate of occurrence. For example, most birds are not swans and most birds are not white. This assumption has received extensive independent verification (McKenzie et al. 2001; McKenzie & Mikkelsen 2000; 2007).

The ODS model predicts that the two cards that lead to the greatest expected information gain are the p and the q cards. Figure 6 (Panel B) shows the fit of the model to the standard data (Oaksford & Chater 2003b). The value of $P(q|p, H_D)$ was set to 0.9 and the best fitting values of $P(p)$ and $P(q)$ were 0.22 and 0.27 respectively, that is, very close to the values used in the earlier example. The ODS model suggests that performance on the selection task displays rational hypothesis testing behavior, rather than irrational confirmation bias. Taking rarity to an extreme provides a simple intuition here. Suppose we consider the (rather implausible) conditional: *If a person is bitten by a vampire bat* (p), *they will develop pointed teeth* (q). Clearly, we should check people who we know to have been bitten, to see if their teeth are pointed (i.e., turn the p card); and, uncontroversially, we can learn little from people we know have not been bitten (i.e., do not turn the $\neg p$ card). If we see someone with pointed teeth, it is surely worth finding out whether they have been bitten – if they have, this raises our belief in the conditional, according to a Bayesian analysis (this is equivalent to turning the q card). But it seems scarcely productive to investigate someone without pointed teeth (i.e., do not turn the $\neg q$ card) to see if they have been bitten. To be sure, it is *possible* that such a person might have been bitten, which would disconfirm our hypothesis, and lead to maximum information gain; but this has an almost infinitesimal probability. Almost certainly, we shall find that they have not been bitten, and learn nothing. Hence, with rarity, the expected informativeness of the q card is higher than that of the $\neg q$ card, diverging sharply from the falsificationist perspective, but agreeing with the empirical data.

It has been suggested, however, that behaviour on this task might be governed by what appears to be a wholly non-rational strategy: *matching bias*. This bias arises in the same context as negative conclusion bias that we discussed earlier, that is, in Evans' (1972) negations paradigm. Take, for example, the rule *if there is an A on one side, then there is not a 2 on the other side* (*if p then* $\neg q$). The cards in this task are described using their logical status, so for this rule, 2 is the *false consequent* (FC) card and 7 is the *true consequent* (TC) card. For this negated consequent rule, participants tend to select the A card (TA: *true antecedent*) and the 2 card (FC). That is, participants now seem to make the falsifying response. However, as Evans and Lynch (1973) pointed out, participants may simply ignore the negations entirely and *match*

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

the values named in the conditional, that is, A and 2. Prima facie, this is completely irrational. However, the "contrast set" account of negation shows that because of the rarity assumption – that most categories apply to a minority of items – *negated* categories are high probability categories (discussed earlier). Having a high probability antecedent or consequent alters the expected information gains associated with the cards. If the probability of the consequent is high, then the ODS model predicts that people *should* make the falsifying TA and FC responses, because these are associated with the highest information gain. Consequently, matching bias is a rational hypothesis testing strategy after all.

Probabilistic effects were first experimentally demonstrated using the *reduced array* version of Wason's selection task (Oaksford et al. 1997), in which participants can successively select up to 15 q and 15 $\neg q$ cards (there are no upturned p and $\neg p$ cards that can be chosen). As predicted by the ODS model, where the probability of q is high (i.e., where rarity is violated), participants select more $\neg q$ cards and fewer q cards. Other experiments have also revealed similar probabilistic effects (Green & Over 1997; 2000; Kirby 1994; Oaksford et al. 1999; Over & Jessop 1998).

There have also been some failures to produce probabilistic effects, however (e.g., Oberauer et al. 1999; 2004). We have argued that these arise because of weak probability manipulations or other procedural problems (Oaksford & Chater 2003b; Oaksford & Moussakowski 2004; Oaksford & Wakefield 2003). We therefore introduced a *natural sampling* (Gigerenzer & Hoffrage 1995) procedure in which participants sample the frequencies of the card categories while performing a selection task (Oaksford & Wakefield 2003). Using this procedure, we found probabilistic effects using the *same* materials as Oberauer et al. (1999), where these effects were not evident.

In further work on matching bias, Yama (2001) devised a crucial experiment to contrast the matching bias and the information gain accounts. He used rules that introduced a high and a low probability category, relating to the blood types Rhesus Negative (Rh–) and Positive (Rh+). People were told that one of these categories, Rh–, was rare. Therefore, according to the ODS model and the rule *if* p *then* $\neg Rh+$ should lead participants to select the rare Rh– card. In contrast, according to matching bias they should select the Rh+ card. Yama's (2001) data were largely consistent with the information gain model. Moreover, this finding was strongly confirmed by using the natural sampling procedure with these materials (Oaksford & Moussakowski 2004).

Alternative probabilistic accounts of the selection task have also been proposed (Evans & Over 1996a; 1996b; Klauer 1999; Nickerson 1996; Over & Evans 1994; Over & Jessop 1998). Recently, Nelson (2005) directly tested the measures of information underpinning these models, including Bayesian diagnosticity (Evans & Over 1996b; McKenzie & Mikkelsen 2007; Over & Evans 1994), information gain (Hattori 2002; Oaksford & Chater 1994; 1996; 2003b), Kullback-Liebler distance (Klauer 1999; Oaksford & Chater 1996), probability gain (error minimization) (Baron 1981; 1985), and impact (absolute change) (Nickerson 1996). Using a related data selection task, he looked at a range of cases in which these norms predicted different orderings of informativeness, for various data types. Nelson found the strongest correlations between his data and information gain (.78). Correlations with diagnosticity (–.22) and log diagnosticity (–.41) were actually *negative*. These results mirrored Oaksford et al.'s (1999) results in the Wason selection task. Nelson's work provides strong convergent evidence for information gain as the index that most successfully captures people's intuitions about the relative importance of evidence.

There has been much discussion in the literature of the fact that selection task results change dramatically for conditionals that express rules of conduct, rather than putative facts, about the world (Cheng & Holyoak 1985; Manktelow & Over 1991). In such tasks, people typically do select the p and $\neg q$ cards – the apparently "logical" response. One line of explanation is that reasoning is domain-specific, rather than applying across-the-board; a further claim, much discussed in evolutionary psychology, is that such tasks may tap basic mechanisms of social reasoning, such as "cheater-detection" (Cosmides 1989), which enables "correct" performance.

A Bayesian rational analysis points, we suggest, in a different direction – that such *deontic* selection tasks (i.e., concerning norms, not facts) require a different rational analysis. In the deontic selection task, participants are given conditionals describing rules concerning how people *should* behave, for example, *if you enter the country, you must have an inoculation against cholera*. The rule is not a hypothesis under test, but a regulation that should be obeyed (Manktelow & Over 1987). Notice, crucially, that it makes no sense to confirm or disconfirm a rule concerning how people *should* behave: People entering the country *should* be inoculated, whether or not they actually *are*. The natural interpretation of a deontic task is for the participant to check whether the rule is being disobeyed – that is, to look for p, $\neg q$ cases (people who enter the country, but are not inoculated); and indeed, in experiments, very high selections of the p and $\neg q$ cards are observed. This is not because people have suddenly become Popperian falsifiers. This is because the task is no longer about attempting to gain information about whether the conditional is true or false. The conditional now concerns how people *should* behave, and hence can neither be confirmed nor disconfirmed by any observations of actual behavior.

We adopted a decision theoretic approach to these tasks (Oaksford & Chater 1994; Perham & Oaksford 2005). Violators are people who enter the country (p) without a vaccination ($\neg q$). Thus, we assume that participants whose role it is to detect violators attach a high utility to detecting these cases, that is, U(p, $\neg q$) is high. However, every other case represents a cost, as it means wasted effort. We argue that people calculate the expected utility associated with each card. So, for example, take the case where someone does not have an inoculation ($\neg q$). She could be either entering the country (p, $\neg q$) or not entering the country ($\neg p$, $\neg q$). Just as in calculating expected information gain, both possible outcomess have to be taken into account in calculating *expected* utility (EU(x)):

$$EU(\neg q) = P(p|\neg q)U(p, \neg q) + P(\neg p|\neg q)U(\neg p, \neg q)$$

We argue that people select cards in the deontic selection task to maximise expected utility. As only the utility of detecting a violator – someone trying to enter without an

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

inoculation – is positive, this means that only the p and the $\neg q$ cards will have positive utilities (because only these cards could show a violator). This model can account for a broad range of effects on deontic reasoning, including the effects of switched rules (Cosmides 1989), perspective changes (Manktelow & Over 1991), utility manipulations (Kirby 1994), and probability manipulations (Manktelow et al. 1995).

Recently, we have also applied this model to rules that contain emotional content, for example, *if you clean up blood, then you must wear gloves* (Perham & Oaksford 2005). With the goal of detecting cheaters (Cosmides 1989), you will look at people who are not cleaning up blood but who are wearing gloves ($\neg p$, q). With the goal of detecting people who may come to harm, you will want to check people who are cleaning up blood but who are not wearing gloves (p, $\neg q$). Perham and Oaksford (2005) set up contexts in which cheater detection should dominate, but in which the goal of detecting people who may come to harm may still be in play. That is, $U(\neg p, \ q) > U(p, \ \neg q) > 0$. The threatening word "blood" can appear for either the p, q case or the p, $\neg q$ case. In calculating *generalized expected utility* (Zeelenberg et al. 2000), a *regret* term (*Re*) is subtracted from the expected utility of an act of detection, if the resulting state of the world is anticipated to be threatening. For example, by checking someone who is not wearing gloves ($\neg q$), to see if they are at risk of harm, one must anticipate encountering blood (p). Because "blood" is a threatening word, the utility for the participant of turning a $\neg q$ card is reduced; that is, the utility of encountering a p, $\neg q$ card is now $U(p, \neg q) - Re$, for regret term *Re*. Consequently, selections of the "not wearing gloves" card ($\neg q$) should be lower for our *blood* rule than for a rule that does not contain a threatening antecedent, such as, *if you clean up packaging, then you must wear gloves*.

In two experiments, Perham and Oaksford (2005) observed just this effect. When participants' primary goal was to detect cheaters, their levels of $\neg p$ and q card selection were the same for the threat (blood rule) as for the no-threat rule. However, their levels of p and $\neg q$ card selection were significantly lower for the threatening than for the non-threatening rules. This finding is important because it runs counter to alternative theories, in particular the evolutionary approach (Cosmides 1989; Cosmides & Tooby 2000), which makes the opposite prediction, that p and $\neg q$ card selections should, if anything, increase for threat rules.

Of the models considered in *BR*, the optimal data selection and expected utility models have been in the literature the longest, and have been subject to most comment. In the rest of Chapter 6, we respond in detail to these comments, pointing out that many can be incorporated into the evolving framework, and that some concerns miss their mark.

7. An uncertain quantity: How people reason with syllogisms

Chapter 7 of *BR* presents a probabilistic model of quantified syllogistic reasoning. This type of reasoning relates two quantified premises. Logic defines four types of quantified premise: *All*, *Some*, *Some. . .not*, and *None*. An example of a logically valid syllogistic argument is:

Some Londoners (*P*) *are soldiers* (*Q*)
All soldiers (*Q*) *are well fed* (*R*)
Therefore *Some Londoners* (*P*) *are well fed* (*R*)

In this example, *P* and *R* are the *end terms* and *Q* is the *middle term*, which is common to both premises. In the premises, these terms can only appear in four possible configurations, which are called *figures*. When one of these terms appears before the copula verb ("are") it is called the *subject* term (in the example, *P* and *Q*) and when one appears after this verb it is called the *predicate* term (*Q* and *R*). As the premises can appear in either order, there are 16 combinations, and as each can be in one of four figures, there are 64 different syllogisms.

There are 22 logically valid syllogisms. If people are reasoning logically, they should endorse these syllogisms and reject the rest. However, observed behavior is graded, across both valid and invalid syllogisms; and some invalid syllogisms are endorsed more than some valid syllogisms. Table 1 shows the graded behaviour over the 22 logically valid syllogisms. There are natural breaks dividing the valid syllogisms into three main groups. Those above the single line are endorsed most, those below the double line are endorsed least, and those in between are endorsed at an intermediate level.

Table 1. *Meta-analysis of the logically valid syllogisms showing the form of the conclusion, the number of mental models (MMs) needed to reach that conclusion, and the percentage of times the valid conclusion was drawn, in each of the five experiments analyzed by Chater and Oaksford (1999b)*

Syllogism	Conclusion	MMs	Mean
All(Q,P), All(R,Q)	*All*	1	89.87
All(P,Q), All(Q,R)	*All*	1	75.32
All(Q,P), Some(R,Q)	*Some*	1	86.71
Some(Q,P), All(Q,R)	*Some*	1	87.97
All(Q,P), Some(Q,R)	*Some*	1	88.61
Some(P,Q), All(Q,R)	*Some*	1	86.71
No(Q,P), All(R,Q)	*No*	1	92.41
All(P,Q), No(R,Q)	*No*	1	84.81
No(P,Q), All(R,Q)	*No*	1	88.61
All(P,Q), No(Q,R)	*No*	1	91.14
All(P,Q), Some. . .not(R,Q)	*Some. . .not*	2	67.09
Some. . .not(P,Q), All(R,Q)	*Some. . .not*	2	56.33
All(Q,P), Some. . .not(Q,R)	*Some. . .not*	2	66.46
Some. . .not(Q,P), All(Q,R)	*Some. . .not*	2	68.99
Some(Q,P), No(R,Q)	*Some. . .not*	3	16.46
No(Q,P), Some(R,Q)	*Some. . .not*	3	66.46
Some(P,Q), No(R,Q)	*Some. . .not*	3	30.38
No(P,Q), Some(R,Q)	*Some. . .not*	3	51.90
Some(Q,P), No(Q,R)	*Some. . .not*	3	32.91
No(Q,P), Some(Q,R)	*Some. . .not*	3	48.10
Some(P,Q), No(Q,R)	*Some. . .not*	3	44.30
No(P,Q), Some(Q,R)	*Some. . .not*	3	26.56

Note. The means in the final column are weighted by sample size.

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Alternative theories of syllogistic reasoning invoke similar processes to explain these data as for conditional inference and the selection task. However, here both mental logic and mental models have to introduce new machinery to deal with quantifiers. For mental logic (Rips 1994), this requires new logical rules for *All* and *Some*, and a guessing mechanism to account for the systematic pattern of responses for the invalid syllogisms. For mental models, dealing with quantifiers requires reinterpreting the lines of a mental model as objects described by their properties (P, Q, and R) rather than as conjunctions of propositions. For the different syllogisms different numbers of mental models are consistent with the truth of the premises. Only conclusions that are true in all of these possible models are logically valid. As Table 1 shows, for the most endorsed valid syllogisms, there is only one model consistent with the truth of the premises, and so the conclusion can be immediately read off. For the remaining valid syllogisms, more than one model needs to be constructed. If people only construct an initial model, then errors will occur. As Table 1 shows, mental models theory provides a good qualitative fit for the valid syllogisms, that is, the distinction between 1, 2, and 3 model syllogisms maps on to the key qualitative divisions in the data.

The probabilistic approach to syllogisms was developed at both the computational and the algorithmic levels in the *Probabilistic Heuristics Model* (PHM, Chater & Oaksford 1999b). One of the primary motivations for this model was the hypothesis that, from a probabilistic point of view, reasoning about *all* and *some* might be continuous with reasoning about more transparently probabilistic quantifiers, such as *most* and *few*. By contrast, from a logical standpoint, such *generalised quantifiers* require a different, and far more complex, treatment (Barwise & Cooper 1981), far beyond the resources of existing logic-based accounts in psychology. Perhaps for this reason, although generalised quantifiers were discussed in early mental models theory (Johnson-Laird 1983), no empirical work on these quantifiers was carried out in the psychology of reasoning.

In deriving PHM, the central first step is to assign probabilistic meanings to the central terms of quantified reasoning using conditional probability. Take the universally quantified statement, *All P are Q* (we use capitals to denote predicates; these should be applied to variables x, which are bound by the quantifier, e.g., $P(x)$, but we usually leave this implicit). Intuitively, the claim that *All soldiers are well fed* can naturally be cast in probabilistic terms: as asserting that the probability that a person is well fed given that they are a soldier is 1. More generally, the probabilistic interpretation of *All* is straightforward: because its underlying logical form can be viewed as a conditional, that is, *All*(x)(*if* $P(x)$ *then* $Q(x)$). Thus, the meaning is given as $P(Q|P) = 1$, as specifying the conditional probability of the predicate term (Q), given the subject term (P).

Similar constraints can be imposed on this conditional probability to capture the meanings of the other logical quantifiers. So, *Some P are Q* means that $P(Q|P) > 0$; *Some P are not Q* means that $P(Q|P) < 1$; and *No P are Q* means that $P(Q|P) = 0$. Thus, for example, "*Some Londoners are soldiers*" is presumed to mean that the probability that a person is a soldier given that he or she is a Londoner is greater than zero, and similarly for the other quantifiers. Such an account generalises smoothly to the generalised quantifiers *most* and *few*. *Most P are Q* means that $1 - \Delta < P(Q|P) < 1$ and *Few P are Q* means that $0 < P(Q|P) < \Delta$, where Δ is small. So, for example, *Most soldiers are well fed* may be viewed as stating that the probability that a person is well fed, given that they are a soldier, is greater than, say, 0.8, but less than 1.

At the level of rational analysis, these interpretations are used to build very simple graphical models (e.g., Pearl 1988) of quantified premises, to see if they impose constraints on the conclusion probability. For example, take the syllogism:

	Some P are Q	
	All Q are R	$P \rightarrow Q \rightarrow R$
Therefore	*Some P are R*	

The syllogistic premises on the left define the dependencies on the right because of their *figure*, that is, the arrangement of the middle term (Q) and the end terms (P and R) in the premises. There are four different arrangements or *figures*. The different figures lead to different dependencies, with different graphical structures. Note that these dependency models all imply that the end terms (P and R) are *conditionally independent*, given the middle term, because there is no arrow linking P and R, except via the middle term Q. Assuming conditional independence as a default is a further assumption about the environment (Step 2 in rational analysis). This is an assumption not made in, for example, Adams' (1998) probability logic.

These dependency models can be parameterised. Two of the parameters will always be the conditional probabilities associated with the premises. One can then deduce whether the constraints on these probabilities, implied by the earlier interpretations, impose constraints on the possible conclusion probabilities, that is, $P(R|P)$ or $P(P|R)$. In this example, the constraints that $P(Q|P) > 0$, and $P(R|Q) = 1$, and the conditional independence assumption, *entail* that $P(R|P) > 0$. Consequently, the inference to the conclusion *Some P are R* is probabilistically valid (p-valid). If each of the two possible conclusion probabilities, $P(R|P)$ or $P(P|R)$, can fall anywhere in the [0, 1] interval given the constraints on the premises, then no p-valid conclusion follows. It is then a matter of routine probability to determine which inferences are p-valid, of the 144 two premise syllogisms that arise from combining *most* and *few* and the four logical quantifiers (Chater & Oaksford 1999b).

In the PHM, however, this rational analysis is also supplemented by an algorithmic account. We assume that people approximate the dictates of this rational analysis by using simple heuristics. Before introducing these heuristics, though, we introduce two key notions: the notions of the *informativeness* of a quantified claim, and the notion of *probabilistic entailment* between quantified statements.

According to communication theory, a claim is informative in proportion to how surprising it is: informativeness varies inversely with probability. But what is the probability of an arbitrary quantified claim? To make sense of this idea, we begin by making a rarity assumption, as in our models of the conditional reasoning and the selection task, that is, the subject and predicate terms apply to only small subsets of objects. On this assumption, if we selected

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

subject term *P*, and predicate term, *Q*, at random, then it is very likely that they will not cross-classify any object (this is especially true, given the hierarchical character of classification; Rosch 1975). Consequently, $P(Q|P) = 0$ and so *No P are Q* is very likely to be true (e.g., *No toupees are tables*). Indeed, for any two randomly chosen subject and predicate terms it is probable that *No P are Q*. Such a statement is therefore quite uninformative. *Some P are not Q* is even more likely to be true, and hence still less informative, because the probability interval it covers includes that for *No P are Q*. The quantified claim least likely to be true is *All P are Q*, which is therefore the most informative. Overall, the quantifiers have the following order in informativeness: I(*All*) > I(*Most*) > I(*Few*) > I(*Some*) > I(*None*) > I(*Some-not*) (see Oaksford et al. 2002, for further analysis and discussion).

Informativeness applies to individual quantified propositions. The second background idea, probabilistic entailment, concerns inferential relations *between* quantified propositions. Specifically, the use of one quantifier frequently provides evidence that another quantifier could also have been used. Thus, the claims that *All swans are white* is strong evidence that *Some swans are white* – because $P(white|swan) = 1$ is included in the interval $P(white|swan) > 0$ (according to standard logic, this does not follow logically, as there may be no swans). Thus, we say that *All* probabilistically entails (or *p*-entails) *Some*. Similarly, *Some* and *Some…not* are mutually *p*-entailing because the probability intervals $P(Q|P) > 0$ and $P(Q|P) < 1$ overlap almost completely.

With this background in place, we can now state the probabilistic heuristics model (PHM) for syllogistic reasoning. There are two types of heuristic: *generate* heuristics, which produce candidate conclusions, and *test* heuristics, which evaluate the plausibility of the candidate conclusions. The PHM account also admits the possibility that putative conclusions may also be tested by more analytic test procedures such as mental logics or mental models. The generate heuristics are:

(G1) *Min*-heuristic: The conclusion quantifier is the same as that of the least informative premise (*min*-premise)

(G2) *P-entailments*: The next most preferred conclusion quantifier will be the *p*-entailment of the *min*-conclusion

(G3) *Attachment*-heuristic: If just one possible subject noun phrase (e.g., *Some R*) matches the subject noun phrase of just one premise, then the conclusion has that subject noun phrase.

The two test heuristics are:

(T1) *Max*-heuristic: Be confident in the conclusion generated by G1 – G3 in proportion to the informativeness of the most informative premise (*max*-premise)

(T2) *Some_not*-heuristic: Avoid producing or accepting *Some_not* conclusions, because they are so uninformative.

We show how the heuristics combine in the following example:

	All P are Q	(*max*-premise)
	Some R are not Q	(*min*-premise)
Therefore	*Some_not*	(by *min*-heuristic)
	Some R are not P	(by *attachment*-heuristic)

and a further conclusion can be drawn:

	Some R are P	[by *p-entailment*]

In *BR*, we compare the results of these heuristics with probabilistic validity, and show that where there is a *p*-valid conclusion, the heuristics generally identify it. For example, the idea behind the *min*-heuristic is to identify the most informative conclusion that validly follows from the premises. Out of the 69 *p*-valid syllogisms, the *min*-heuristic identifies that conclusion for 54; for 14 syllogisms the *p*-valid conclusion is less informative than the *min*-conclusion. There is only one violation; that is, where the *p*-valid conclusion is more informative than the *min*-conclusion.

In turning to the experimental results, in *BR* we first show how all the major distinctions between standard syllogisms captured by other theories are also captured by PHM. So, returning to Table 1, all the syllogisms above the double line have the most informative *max*-premise, *All* (see heuristic T1). Moreover, all the syllogisms below the single line have uninformative conclusions, *Some-not* (see heuristic T2), and those below the double line violate the *min*-heuristic (heuristic G1) and require a *p*-entailment (heuristic G2), that is, *Some…not* ↔ *Some*. Consequently, this simple set of probabilistic heuristics makes the same distinctions among the valid syllogisms as the mental models account.

In this Précis, we concentrate on novel predictions that allow us to put clear water between PHM and other theories. As we discussed earlier, the most important feature of PHM is the extension to *generalised quantifiers*, like *most* and *few*. No other theory of reasoning has been applied to syllogistic reasoning with generalised quantifiers. Table 2 shows the *p*-valid syllogisms involving generalised quantifiers showing the conclusion type and the percentage of participants selecting that conclusion type in Chater and Oaksford's (1999b) Experiments 1 and 2. The single lines divide syllogisms with different *max*-premises, showing a clear ordering in levels of endorsements dependent on heuristic T1. All those above the double line conform to the *min*-heuristic (heuristic G1), whereas those below it do not and require a *p*-entailment (heuristic G2). As Chater and Oaksford (1999b) pointed out, one difference with experiments using standard logical quantifiers was that the *Some…not* conclusion was not judged to be as uninformative, that is, heuristic T2 was not as frequently in evidence. However, in general, in experiments using generalised quantifiers in syllogistic arguments the heuristics of PHM predict the findings just as well as for the logical quantifiers (Chater & Oaksford 1999b).

Many further results have emerged that confirm PHM. The *min*-heuristic captures an important novel distinction between strong and weak possible conclusions introduced by Evans et al. (1999). They distinguished conclusions that are necessarily true, possibly true, or impossible. For example, taking the syllogism discussed earlier (with premises, *Some P are Q, All Q are R*), the conclusion *Some P are R* follows *necessarily*, *No P are R* is *impossible*, and *Some P are not R* is *possible*. Some possible conclusions are endorsed by as many participants as the necessary conclusions (Evans et al. 1999). Moreover, some of the possible conclusions were endorsed by as few participants as the impossible conclusions. Evans et al. (1999) observed that possible conclusions that are commonly endorsed all conform to the *min*-heuristic, whereas those which are rarely endorsed violate the

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Table 2. *The* p*-valid syllogisms less the syllogisms that are also logically valid (shown in Table 1), showing the form of the conclusion and the proportion of participants picking the* p*-valid conclusion in Chater and Oaksford's (1999b) Experiments 1 and 2*

Syllogism	Conclusion	Mean
All(Q,P), Most(R,Q)	*Most*	85
Most(Q,P), All(R,Q)	*Most*	65
All(P,Q), Most(Q,R)	*Most*	70
Most(P,Q), All(Q,R)	*Most*	55
Few(P,Q), All(R,Q)	*Few*	80
All(P,Q), Few(R,Q)	*Few*	85
Few(P,Q), All(R,Q)	*Few*	85
All(P,Q), Few(Q,R)	*Few*	75
Most(Q,P), Most(R,Q)	*Most*	65
Most(P,Q), Most(Q,R)	*Most*	50
Few(Q,R), Most(R,Q)	*Few*	60
Most(Q,R), Few(R,Q)	*Few*	75
Most(P,Q), Few(Q,R)	*Few*	70
Most(Q,P), Some. . .not(R,Q)	*Some. . .not*	80
Some. . .not(Q,P), Most(R,Q)	*Some. . .not*	60
Some. . .not(Q,P), Most(Q,R)	*Some. . .not*	75
Most(Q,P), Some. . .not(Q,R)	*Some. . .not*	65
Most(P,Q), Some. . .not(Q,R)	*Some. . .not*	75
Some. . .not(P,Q), Most(Q,R)	*Some. . .not*	75
Few(Q,P), Some. . .not(R,Q)	*Some. . .not*	60
Some. . .not(Q,P), Few(R,Q)	*Some. . .not*	40
Some. . .not(Q,P), Few(Q,R)	*Some. . .not*	30
Few(Q,P), Some. . .not(Q,R)	*Some. . .not*	60
Few(P,Q), Some. . .not(Q,R)	*Some. . .not*	60
Some. . .not(P,Q), Few(Q,R)	*Some. . .not*	40
All(P,Q), Most(R,Q)	*Some. . .not*	35
Most(P,Q), All(R,Q)	*Some. . .not*	35
Few(Q,P), Few(R,Q)	*Some. . .not*	35
Few(P,Q), Few(Q,R)	*Some. . .not*	30
Few(P,Q), Most(Q,R)	*Some. . .not*	30

Note. This table excludes the eight MI, IM, FI, and IF syllogisms, which have two *p*-valid conclusions only one of which was available in Chater and Oaksford's (1999b) Experiment 2.

min-heuristic (with one exception). Hence, PHM captures this important new set of data.

Some experiments designed to test the claim that syllogism difficulty is determined by the number of alternative mental models can also be interpreted as confirming PHM (Newstead et al. 1999). Participants wrote down or drew diagrams consistent with the alternative conclusions they entertained, during syllogistic reasoning. No relationship was found between the number of models a syllogism requires (according to mental models theory) for its solution and the number of conclusions or diagrams participants produced. This suggests that sophisticated analytical procedures, such as those described in mental models, play, at most, a limited role in the outcome of syllogistic reasoning. By contrast, participants' responses agreed with those predicted by the *min*- and *attachment*-heuristics. Furthermore, no differences in task difficulty dependent on syllogistic figure were observed, a finding consistent with PHM, but not mental models.

Recent work relating memory span measures to syllogistic reasoning has also confirmed PHM (Copeland & Radvansky 2004). PHM makes similar predictions to mental models theory because the number of heuristics that need to be applied mirrors the one, two, and three model syllogism distinction (see Table 1). For one model syllogisms, just the *min*-heuristic and *attachment* is required (two heuristics). For two model syllogisms, the *some__not*-heuristic is also required (three heuristics). In addition, for three model syllogisms a *p*-entailment is required (four heuristics). The more mental operations that need to be performed, the more complex the inference will be, and the more working memory it will require. Copeland and Radvansky (2004) found significant correlations between working memory span and strategy use, for both mental models and PHM. While not discriminating between theories, this work confirmed the independent predictions of each theory for the complexity of syllogistic reasoning and its relation to working memory span.

As with Chapters 5 and 6, Chapter 7 of *BR* closes by addressing the critiques of PHM that have arisen since the theory first appeared. One criticism is that PHM does not generalise to *cardinal* quantifiers (Geurts 2003) such as *Exactly three P are Q*, which have no probabilistic interpretation. Yet, such quantifiers can, nonetheless, naturally mesh with the generalized quantifiers, to yield interesting inferences. For example, suppose you are told that exactly three birds in the aviary are black. If there are twenty birds in the aviary, then *few* of the birds are black; if there are four, then *most* of the birds are black; and, in either case, further inferences from these generalized quantifiers can be drawn, as appropriate.

8. Conclusion

As we have seen, Chapters 5 to 7 of *BR* provide the empirical support for taking a probabilistic approach to human reasoning and rationality. The final chapter provides further arguments for pursuing this research strategy in the form of a dialogue between an adherent of the probabilistic approach and a sceptic. In this Précis, we concentrate on two key issues that emerge from that debate.

The first topic we consider is whether the brain is a probabilistic inference machine. *BR* focuses primarily, as we have seen, on providing rational analyses of human reasoning – and we have noted that rational analysis does not make direct claims about underlying computational operations. But, to what extent can the mind or brain be viewed as a probabilistic (or for that matter, a logical) calculating engine? Although not the primary focus in this book, this is nonetheless a fundamental question for the behavioural and brain sciences. We suspect that, in general, the probabilistic problems faced by the cognitive system are too complex to be solved by direct probabilistic calculation. Instead, we suspect that the cognitive system has developed relatively computationally "cheap" methods for reaching solutions that are "good enough" probabilistic

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

solutions to be acceptable. In particular, where we propose a specific processing theory (in our account of syllogistic reasoning) this account consists of simple, but surprisingly effective, heuristics – heuristics that, however, lead to errors, which we argue are revealed in the empirical data. Moreover, in related work on the topic of rational choice and decision making, which we do not consider here, we and others have proposed models that solve probabilistic/decision making problems, but do so using relatively cheap, and hence approximate, methods (Gigerenzer & Goldstein 1996; Gigerenzer et al. 1999; Stewart et al. 2006).

To the degree that algorithmic models can be formulated, is rational analysis simply redundant? We argue that it is not. Rational analysis is essential because it explains *why* the particular algorithms used by the cognitive system are appropriate. That is, without a characterization of what problem the cognitive system solves, we cannot ask, let alone answer, the questions of why the algorithm has its particular form, or how effectively it works. Moreover, it may be that a good deal of empirical data about human reasoning (and indeed, human cognition more generally) can be understood as arising from the structure of the problem itself – that is, the nature of the problem drives any reasonable algorithmic solution to have particular properties, which may be evident in the data. This idea is a core motivation for the rational analysis approach (Anderson 1990; 1991a); and we have seen that a broad spectrum of data on human reasoning can be understood purely at the rational level – that is, without formulating an algorithmic theory of any kind.

The second topic we consider is the importance of qualitative patterns of probabilistic reasoning, rather than precise numerical calculations. Suppose, for concreteness, we consider a person reasoning about a game of dice. If the dice are unbiased, then it is easy, of course, for the theorist to formulate a probabilistic model specifying that each throw is independent, and that each face has a probability of 1/6. But this model is both too strong and too weak. It is too strong because it generates all manner of subtle mathematical predictions, concerning, say, the relative probabilities of rolling at least one six out of six dice rolls versus rolling at least two sixes out of twelve dice rolls, predictions that are not available to everyday intuition. And it is too weak because it ignores many factors of crucial importance in everyday reasoning. For example, watching a dice being thrown, we have not only a model of the probability that each face will be uppermost, but a rough model of where it will land, how likely it is to fall off the table, how loud that impact is likely to be, how another player is likely to react to a particular outcome, given their temperament, the gamble they have placed, and so on.

This observation implies that, if the cognitive system is indeed building probabilistic models of the world, then it is building models of considerable complexity – models that can take into account any aspect of knowledge, from naïve physics to folk psychology. This implies that the probabilistic turn does not resolve the difficulty of representing knowledge – rather it provides a framework into which this knowledge must be integrated. The advantage of the probabilistic viewpoint, though, is that it provides a powerful framework for dealing with an uncertain world; and, indeed, for assessing competing explanations of observed phenomena (rival interpretations of perceptual input; competing grammars; alternative interpretations of sentences, stories, or court-cases). Moreover, probabilistic models of complex domains do not need to be fully specified, at a numerical level – most critical is that the *functional* relationships between pieces of information are represented. What tends to cause what? What is evidence for what? The direction and existence of functional dependencies between pieces of information may be mentally represented, even though precise numerical probabilities may be unknown. Thus, probability theory can provide a framework for *qualitative* reasoning, without using numerical values (e.g., Pearl 2000). We tentatively suggest that much of the power, and limitations, of human reasoning about the everyday world flows from this qualitative style of reasoning. From this point of view, it is perhaps not surprising that people are not good at explicit reasoning with probabilities – indeed, they fall into probabilistic fallacies just as readily as they fall into logical contradictions (e.g., Kahneman et al. 1982).

The probabilistic mind is not, of course, a machine for solving verbally or mathematically specified problems of probability theory. Instead, we suggest, the mind is a qualitative probabilistic reasoner, in the sense that the rational analysis of human reasoning requires understanding how the mind deals qualitatively with uncertainty. As we have stressed, this does not imply that the mind is a probabilistic calculating machine (although it may be); still less does it imply that the mind can process probabilistic problems posed in a verbal or mathematical format. Nonetheless, the concepts of probability are, we suggest, as crucial to understanding the human mind as the concepts of aerodynamics are in understanding the operation of a bird's wing.

ACKNOWLEDGMENTS

The book *Bayesian Rationality* owes huge debts to more people and institutions than we can summarize here. We would especially like to express our gratitude to the University of Edinburgh and the Centre for Cognitive Science (formerly the School of Epistemics, and now part of the Division of Informatics) for starting us on the adventure of understanding human reasoning from a formal and empirical standpoint, which has engaged so much of our lives for the last twenty years.

NOTES

1. The case where the categorical premise is uncertain can be accommodated using a generalization of this idea, Jeffrey conditionalization (Jeffrey 1983). The new degree of belief that *John plays tennis* (q), on learning that *it is sunny in Bloomsbury* (which confers only a high probability that *it is sunny in Wimbledon* [p]), is:

$$P_1(q) = P_0(q|p)P_1(p) + P_0(q|\neg p)P_1(\neg p)$$

2. Bayes' theorem is an elementary identity of probability theory that allows a conditional probability to be calculated from its converse conditional probability and the priors: $P(p|q) = (P(q|p)P(p))/P(q)$.

3. However, this may be because of the different way that negations are used in each task (see Evans & Handley 1999; Oaksford 2004b).

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Open Peer Commentary

Classical logic, conditionals and "nonmonotonic" reasoning

doi:10.1017/S0140525X09000296

Nicholas Allott and Hiroyuki Uchida

Department of Linguistics, University College London, London WC1N 1PF, United Kingdom.

n.allott@ucl.ac.uk **uclyhuc@ucl.ac.uk**

http://www.phon.ucl.ac.uk/home/nick/

Abstract: Reasoning with conditionals is often thought to be non-monotonic, but there is no incompatibility with classical logic, and no need to formalise inference itself as probabilistic. When the addition of a new premise leads to abandonment of a previously compelling conclusion reached by *modus ponens*, for example, this is generally because it is hard to think of a model in which the conditional and the new premise are true.

We doubt two claims made by Oaksford & Chater (O&C), one linguistic and one about inference, as they relate to conditionals (*Bayesian Rationality*, Oaksford & Chater 2007, henceforth *BR* as in target article). We do not think that the case has been made that sentences of the form "If A then B" have a semantics diverging from the material implication of propositional logics. We focus here on the related claim that human inference is probabilistic.

Classical Propositional Logic (CPL) is often claimed to be inadequate for explaining our spontaneous propositional inferences. The claim is based on the observation that, whereas CPL is monotonic, human inferences seem to be non-monotonic. For example, it is argued that the validity of some inferences, such as the one in (1a), may be cancelled by addition of another premise, such as the premise *R* in (1b). The claim is that now that John has broken his left leg, John will no longer run, even if the weather is fine. This is claimed to show the inadequacy of CPL as a tool to explain our propositional inferences, because in CPL, addition of another premise does not influence the provability of the sequent, as is shown in the inference from top to bottom in (1c).

(1a) (Premise 1) $P \rightarrow Q$: If the weather is fine, John runs for a mile.

(Premise 2) *P*: The weather is fine.

(Conclusion) 3. *Q*: John runs for a mile.

The valid inference: Whenever Premise 1 and Premise 2 are true, 3 is also true.

(1b) *R*: John breaks his left leg.

(1c) $P, P \rightarrow Q \vdash Q$

$P, P \rightarrow Q, R \vdash Q$ Weakening

However, it is not clear that the addition of the premise *R* in (1b) to the set of premises in (1a) actually cancels out the validity of the initial sequent, $P, P \rightarrow Q \vdash Q$, in our propositional inferences. The point becomes clearer if we add the new proposition *R* to the set of premises, as in $P, P \rightarrow Q, R \vdash Q$. With the specification of the premise propositions as in (1a) to (1b), we do not find this revised sequent to be valid, but this is only because we no longer find the proposition "If the weather is fine, John runs for a mile" ($= P \rightarrow Q$) to be true when we reevaluate the inference. In contrast, if we force ourselves to assume that all the three premise propositions $P, P \rightarrow Q, R$ are true in a model of interpretation, then in that model, we have to conclude *Q*. Some might find it difficult to think of such an interpretation model, because common-sense knowledge tells us that a person normally does not run with a broken leg, but we may sometimes make a claim such as, "If the weather is fine, John (always) runs for a mile. It does not matter if he gets injured. He always runs for a mile." Thus, it is not impossible to force ourselves to think of models in which the proposition, "If the weather is fine, John runs for a mile" is true despite the fact that John has broken his leg. In such models, whenever *P* is true, *Q* is also true: that is, the inference goes through.

As is clear from the foregoing discussion, typical interpretation data that allegedly show that monotonic logic cannot capture our propositional inferences include a change of models in which we evaluate the propositional sequents. The valid inference in (1) can be re-stated as "In each model in which *P* and $P \rightarrow Q$ are both true, *Q* is also true." The cancellation of the truth of *Q* arises because in the new model in which we "re-evaluate" the sequent, the premise $P \rightarrow Q$ is no longer true (or we find it more difficult to think of a model in which both $P \rightarrow Q$ and *R* are true). Because the initial valid inference in (1a) concludes *R* as a true proposition only on condition that $P \rightarrow Q$ and *P* are both true, this revision does not really involve the cancellation of the validity of the initial inference.

Given that specification of models in which logical formulas/sequents are evaluated is not part of either the syntax or the semantics of propositional logic languages, it is not clear that alleged non-monotonicity of reasoning, which arises because of the revision of models, requires formalising logical inference as non-monotonic or probabilistic. No doubt people hold beliefs with varying degrees of strength, and a result of reasoning is that these degrees of strength are changed. But one can agree with O&C on this point without thinking that the formal inference system itself is probabilistic.

Space here does not permit discussion of how some propositions are accepted and others rejected, but we are sympathetic to O&C's claim that heuristics that are sensitive to information gain must be involved, with the caveat that it cannot simply be information that is sought, but information that is important to the reasoner at a reasonable processing cost. This recalls discussion of relevance in Gricean pragmatics and Sperber and Wilson's relevance theory.

As Oakford & Chater note, for one's belief in the conditional in (2) it matters whether one discovers, for example, an unstarted car or is told that a car did not start.

(2) If the key is turned the car will start.

A pragmatic explanation in terms of the tendency of speakers to produce utterances relevant to their audience is natural. Effects of the order in which information is presented (see *BR*, pp. 157ff) also require such an explanation, we believe.

This raises a methodological point. To understand human reasoning, both classical logic and O&C's probabilistic account of conditionals and of inference must be supplemented by accounts of processing, and of the pragmatics of utterance interpretation. Thus, it is not obvious that the probabilistic account is more parsimonious.

Identifying the optimal response is not a necessary step toward explaining function

doi:10.1017/S0140525X09000302

Henry Brighton and Henrik Olsson

Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, 14195 Berlin, Germany.

hbrighton@mpib-berlin.mpg.de

olsson@mpib-berlin.mpg.de

Abstract: Oaksford & Chater (O&C) argue that a rational analysis is required to explain why a functional process model is successful, and that, when a rational analysis is intractable, the prospects for

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

understanding cognition from a functional perspective are gloomy. We discuss how functional explanations can be arrived at without seeking the optimal response function demanded by a rational analysis, and argue that explaining function does not require optimality.

Oaksford & Chater (O&C) argue in *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) that a rational analysis is an essential step toward understanding process models from a functional perspective and that "[d]oing this requires developing an account of the optimal behaviour" (p. 268). We argue that relative uses of rational principles of inductive inference can be used to explain function without knowing the optimal response function, and propose that multiple forms of functional analysis are required to understand the cognitive system from a functional perspective.

Rational principles of inductive inference such as Bayesian and simplicity principles are perhaps most frequently used as criteria in relative statements of the form "P is a better response than Q for problem X" For instance, the use of rational principles as model selection criteria (Kearns et al. 1997; Pitt et al. 2002). In contrast, step four of Anderson's (1991a) rational analysis excludes anything but statements of the form "P is the optimal response to task X." Because rational principles can be used to compare the behavior of process models without knowing the optimal response, rational analysis is an instance of a broader class of methodologies adopting rational principles to understand function (e.g., Gigerenzer et al. 1999). Given this, does knowledge of the optimal response offer any intrinsic advantage when explaining why a process is successful; and what price do we pay by demanding knowledge of the optimal response function?

Knowledge of the optimal solution is one way of establishing *that* a particular process or organism is successful; but explaining *why* the process is successful is not implied by this finding. For example, knowing the optimality conditions of the naïve Bayes classifier does not by itself tell us why it is successful. One explanation for why naïve Bayes is successful might be that the independence assumption results in fewer parameters. In certain contexts which violate this independence assumption, the assumption nevertheless causes a reduction in the variance component of error relative to a learning algorithm that assumes that the features are dependent (Domingos & Pazzani 1997). This causal explanation for why naïve Bayes is successful does not require knowledge of the optimal response. It can be established using relative uses of rational principles when the optimal response is incalculable. Furthermore, for realistic contexts of sparse exposure the optimality conditions for naïve Bayes, despite its simplicity, are not fully known (Kuncheva 2006). Although typically unavailable, knowing the optimality conditions for an algorithm would undoubtedly provide a good starting point to understand its function; but optimality conditions are neither a required starting point, nor do they by themselves offer a causal explanation for *why* the algorithm is functional.

Functional explanations arising from Bayesian rational analyses also aim to tell us why a pattern of behavior is rationally justified given that the environment has a certain probabilistic structure. The optimal response function provides a reference point against which to measure behavior without committing to how that behavior is achieved. Because relative uses of rational principles are order relations over theories, or models, of problem solving, they cannot be used in this way. However, abstracting from the process level limits what causal explanation one can offer for why the organism does what it does. One can say *that* it is successful, but whether or not such a finding provides a satisfactory causal explanation for *why* it is successful is not so clear (Danks 2008).

Problems arising from the intractability of reliably identifying the optimal response can also make it less desirable. O&C contemplate the possibility that when a rational analysis is intractable, cognitive science may also be intractable (*BR*, p. 283). Alternatively, this makes complementary forms of functional analysis all the more necessary. The distinction between relative and absolute uses of rational principles mirrors the distinction between, and relative difficulty of, verification and search problems in complexity theory. For example, for an instance of the traveling salesperson problem, the comparative statement "Tour P is shorter than tour Q" is trivial to verify, but the absolute statement "Tour P is the shortest tour" will often be intractable to establish. Many problems take this form and are NP complete (Nondeterministic Polynomial time), including the computation of optimal Bayesian responses and approximations in many settings (e.g., Cooper 1990).

Functional analyses based on rational principles of induction rest on several idealizations: (a) Although different rational principles of induction point to a coherent theoretical picture of what makes a good inference, their practical implementations are often inconsistent, and point to different conclusions (Kearns 1997); (b) functional models will not capture all forms of uncertainty impacting on the problem, some of which may change the character of a functional response (Bookstaber & Langsam 1985); (c) functional models always consider local goals, which only partially inherit the properties of the global goal being examined; (d) rational principles of inductive inference are approximate models of function, which do not consider functional pressures arising from, for example, processing (Brighton & Gigerenzer 2008; Todd & Gigerenzer 2003). These are unavoidable realities of modeling, and apply to both relative and absolute uses of rational principles of induction.

An explanation requiring an optimal response function must also consider that: (e) for problems of inductive inference, the optimal response is often analytically intractable to determine with exact methods, and will not be unique; (f) behavioral responses are typically approximately optimal, revealing a tendency rather than a correspondence; (g) successfully optimizing a local goal does not necessarily take us toward the global optimal when other dependencies are known to be only approximately fulfilled (Lipsey & Lancaster 1956). These additional factors lead to increased flexibility in what behaviors we choose to label optimal. Our point is that these further assumptions are a choice rather than a necessity, and are only required to support certain forms of explanation. We agree with (O&C) on the importance of understanding the ecological function of cognitive processes. We also agree that rational analysis represents a powerful move in this direction. But functional analysis can also proceed without seeking to establish optimality with respect to the organism.

Explaining norms and norms explained

doi:10.1017/S0140525X09000314

David Danks[a] and Frederick Eberhardt[b]

[a]*Department of Philosophy, Carnegie Mellon University and Institute for Human and Machine Cognition, Pittsburgh, PA 15213;* [b]*Institute of Cognitive and Brain Sciences, University of California, Berkeley, Berkeley, CA 94720, and Department of Philosophy, Washington University in St. Louis, St. Louis, MO 63130.*

ddanks@cmu.edu

http://www.hss.cmu.edu/philosophy/faculty-danks.php

fde@berkeley.edu

http://www.phil.cmu.edu/~fde/

Abstract: Oaksford & Chater (O&C) aim to provide teleological explanations of behavior by giving an appropriate normative standard: Bayesian inference. We argue that there is no uncontroversial independent justification for the normativity of Bayesian inference, and that O&C fail to satisfy a necessary condition for teleological explanations: demonstration that the normative prescription played a causal role in the behavior's existence.

In *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) we understand Oaksford & Chater (O&C) as providing the

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

following argument for Bayesian models as teleological explanations in psychology:

1. *Normative Justification:* The normativity of Bayesian inference can be established independently of empirical observations. (p. 31 of their book)

2. *Normative Force:* Appropriate normative principles lead to behavior conforming to those principles. (p. 33)

3. *Default Assumption:* People are pretty good at achieving their goals (i.e., they are "everyday rational"). (p. 19) Therefore, people's behavior is a guide to the appropriate normative principles. (p. 30)

4. *Empirical Results:* People's behavior conforms to the prescriptions of Bayesian models. (Chs. 6 & 7)

Conclusion: Therefore, Bayesian models explain *why* that behavior occurs. (Ch. 8)

The first three premises are *formal* claims that are instances of necessary parts of proper teleological explanations. Without independent justification (Premise 1), there is no normative power behind Bayesian inference. Without influence on behavior (Premise 2), normative principles provide no explanation. And if the normative prescription cannot be reasonably interpreted in terms of desirable human goals (Premise 3), then the explanation is implausible. We suggest, however, that neither Premise 1 nor Premise 2 is supported by O&C's arguments.

Premise 1 is usually grounded in two standard arguments. Dynamic Dutch book arguments aim to establish Bayesian inference as the unique belief change procedure that avoids irrational, between-time inconsistencies, understood as acceptance of bets (potential, not actual) over time that result in guaranteed loss (Teller 1973). Standard forms of these arguments, however, also imply absurd prescriptions, including: reasoners should not protect themselves against predictable irrationalities (Maher 1992); reasoners should not retrospect on their past beliefs (Levi 1988); and reasoners should never change their conditional commitments (Levi 1988; 2002). If these arguments are weakened to avoid these unwanted implications, then Bayesian inference is only one of infinitely many ways to avoid dynamic Dutch book; it has no particularly special status.

The second standard argument is long-run convergence: Roughly, any non-zero degree of belief in the truth will converge to 1 (using Bayesian inference) as one collects more evidence (Savage 1972), and no other reliable method always converges to the truth faster (Schulte 1999). However, the convergence arguments have the unreasonable requirement that the Bayesian reasoner be logically omniscient. There are also problems that can be solved by a naïve falsificationist, but which the Bayesian can solve only if she can "compute" uncomputable functions (Juhl 1993; Osherson et al. 1988). Long-run convergence thus cannot provide a conceptual justification for the normativity of Bayesian inference.

There might be other arguments for Premise 1, but we have not seen them, nor do O&C provide them. We agree with O&C that the normative principles underlying rationality may vary between situations. But independent justification of normativity must be provided for *whatever* principles are appropriate, else one cannot have an adequate teleological explanation. The dependence of the normative principle on the situation is inversely correlated with the explanatory power provided by the normative principle. A situation-independent justification of Premise 1 (or some similar premise) is necessary for teleological explanations.

Our second worry is that, even if Premise 1 holds, O&C require only an empirical match between behavior and normative prescription. Their argument thus neglects the requirement for a teleological explanation that the normative principle must have played a causal role – ontogenetic, phylogenetic, or both – in the behavior's existence or persistence. "Origin stories" are required for teleological explanation, but are never provided by O&C. Behavior B could be optimal for task T even though behavior B results from considerations that are independent of task T; B's optimality might be coincidental. In this case, the claim "Behavior B because it is optimal for task T" is an incorrect explanation, even though B conforms (empirically) to the normative prescription.

O&C seem to have two lines of response. First, if there is widespread alignment between the normative prescription and people's behavior, then it is arguably improbable that the behavior is only coincidentally optimal. Consequently, research confirming correspondence in a wide range of conditions is evidence in favor of the normative principle. This response fails, however, because widespread alignment is actually to be expected for Bayesian models, given the many "free parameters" in such models: hypothesis space, prior probabilities, utilities, likelihood functions, and various plausible computational approximations. In particular, for any evidence and any behavior (represented as a probability distribution over possible choices), there exists a prior probability distribution such that the predicted posterior distribution after inference on the evidence matches the observed behavior.[1] That is, for (almost) any psychological experiment, we know a priori that a Bayesian model will be capable of fitting the observed behavior, whatever it might be. Repeatedly developing *sui generis* Bayesian models for each task does not compensate for a lack of "origin stories," even if the models successfully predict.

O&C's second line of response is more promising: successful model predictions for experiments that vary features of one particular task make it plausible that the normative constraints played a causal role in shaping the observed behavior. We support this line of response to the lack of "origin stories," but are doubtful about how much support the evidence provides for the *normativity* of Bayesian inference. O&C here primarily cite manipulations of the base-rate (*BR*, pp. 146ff, 178ff), and as they recognize, there are several competing models with similar normative appeal. Moreover, there is substantial controversy about which model provides the best fit of the data. There is a potential response, but it is currently only potential.

O&C admirably search for teleological explanations of human behavior. We are in complete agreement with them that such explanations are desirable, but we believe that their enterprise requires a stronger foundation. They have neither properly established Bayesian inference as a normative principle, nor shown – directly or indirectly – that the optimality of Bayesian inference (if it *is* optimal in some respect!) is actually a cause of people approximating those prescriptions. They offer only an existence proof – "Behavior *could* be caused by a *potentially* normative principle" – and such accounts are too weak to provide teleological explanations.

ACKNOWLEDGMENT

The second author, Frederick Eberhardt, is funded by a fellowship from the James S. McDonnell Foundation.

NOTE

1. There are other technical assumptions (e.g., nonextremal likelihoods), but they are irrelevant to our core point.

Beyond response output: More logical than we think

doi:10.1017/S0140525X09000326

Wim De Neys

Lab for Experimental Psychology, University of Leuven, 3000 Leuven, Belgium.
wim.deneys@psy.kuleuven.be
http://ppw.kuleuven.be/reason/wim/

Abstract: Oaksford & Chater (O&C) rely on a data fitting approach to show that a Bayesian model captures the core reasoning data better than its logicist rivals. The problem is that O&C's modeling has focused

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

exclusively on response output data. I argue that this exclusive focus is biasing their conclusions. Recent studies that focused on the processes that resulted in the response selection are more positive for the role of logic.

It is well established that in a wide range of reasoning tasks people fail to select the response that standard logic dictates. Oaksford & Chater (O&C) in *Bayesian Rationality* (Oaksford & Chater 2007) nicely demonstrate that a Bayesian model fits the available response selection data better than its logicist rivals. However, it is crucial to stress that O&C's modeling has not moved beyond this mere response output data. Other measures such as basic latencies or brain-imaging data that can help to examine the processes that led to the outputs are not being considered. This limitation is bound to bias their conclusion. It is clear that people's final response selection does not tell us how they arrived at this response. The fact that people do not select the predicted logical response does not suffice to discard the logicist view. It is always possible that people attempt to reason in line with standard logic but fail to do so because of specific processing difficulties. Addressing this issue requires that one digs bellow the output surface.

In O&C's defense, it may be noted that the fixation on response output is characteristic for the whole psychology of reasoning (e.g., Hoffrage 2000). In the vast majority of classic reasoning studies, response accuracy has been the sole dependent variable of interest. From a pragmatic point of view, one might argue that O&C were simply focusing on the most prominent data and might move to more process-related measures in the future. However, the key point is that in the meantime caution is needed when drawing any strong conclusions about the role of logic in reasoning. I will illustrate the point with some examples from recent brain-imaging and latency studies. This processing data suggests that although people very often fail to select the predicted logical response, they are nevertheless trying to adhere to the logical norm.

Houdé et al. (2000), for example, scanned people's brain activation while they were trying to solve the selection task. Participants were scanned before and after they received training aimed at increasing their inhibitory control capacities. Although the training did not instruct people about the standard logical response (i.e., *P* and not-*Q* card selection), it did boost the selection rate of this pattern and resulted in an increased activation of prefrontal brain areas believed to be involved in inhibitory processing. This suggests that the initial low selection rate of the predicted logical response does not result from a lack of logical knowledge, but from a failure to inhibit the intuitive appealing matching response. If people did not know what the logical response was or considered it to be irrelevant, merely training their inhibitory capacities should not affect its selection rate. Likewise, numerous studies on belief bias in syllogistic reasoning have shown that people typically fail to respond logically to reasoning problems in which intuitive beliefs conflict with the logical response. Nevertheless, latency and eye-tracking studies indicate that these problems take longer to respond to and are more thoroughly inspected compared to problems in which beliefs and logic do not conflict (e.g., Ball et al. 2006). If people were not trying to adhere to the logical norm, violating it should not affect their reasoning behavior. Furthermore, consistent with the latency findings, De Neys and Franssens (2007) observed that after solving syllogisms in which beliefs and logic conflicted, reasoners showed an impaired access to words that were associated with the cued beliefs. Such an impaired memory access is a key marker of inhibitory processing. Even people who were always biased by their beliefs showed a minimal impairment, indicating that they had attempted to inhibit the intuitive beliefs, but failed to complete the process. If people were not trying to adhere to the logical norm, there would be no reason for them to block the conflicting intuitive response. Bluntly put, although people did not manage to select the logical response, they were at least trying to do so.

Interestingly, brain-imaging of classic tasks from the judgment and decision-making literature is pointing towards a similar conclusion. As O&C note, people hardly ever reason in line with classic probabilistic norms in these tasks. People typically select responses that are cued by intuitive heuristics. Nevertheless, a recent fMRI study showed that a specific brain-area, the anterior cingulate cortex, involved in the detection of conflict between competing responses, is activated when people select the heuristic response and violate the norm (De Neys et al. 2008). This indicates that people detect that their heuristic response conflicts with the classic normative response. If people were not taking the norm into account, they would not experience such a conflict (see also De Neys & Glumicic 2008).

The general point is that although people's response selections in all these cases deviate from the predicted classic normative pattern, additional process data indicates that people are nevertheless trying to adhere to the norm. This should give pause for thought before discarding the logical perspective. O&C have modeled but a small, superficial subset of relevant data. If one moves beyond the static response output, it becomes clear that people can be more logical than their answers suggest.

ACKNOWLEDGMENT

The authors' research is supported by the Fund for Scientific Research Flanders (FWO-Vlaanderen).

Does rational analysis stand up to rational analysis?

doi:10.1017/S0140525X09000338

Jonathan St. B. T. Evans

Centre for Thinking and Language, School of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom.

j.evans@plymouth.ac.uk

Abstract: I agree with Oaksford & Chater (O&C) that human beings resemble Bayesian reasoners much more closely than ones engaging standard logic. However, I have many problems with their "rational analysis" framework, which appears to be rooted in normative rather than ecological rationality. The authors also overstate everyday rationality and neglect to account for much relevant psychological work on reasoning.

There is much that I can admire in *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) and the lengthy research programme that it describes. There is also much that I agree with: for example, that human beings resemble Bayesian reasoners much more closely than ones engaging standard logic. I also share part of the authors' research programme, in that I have argued against logicism and excessive use of the deduction paradigm in the psychology of reasoning (Evans 2002), as well as proposed that ordinary conditionals are much better modelled by probability logic than by standard propositional logic (Evans & Over 2004). So why is it that when I read the Socratic dialogue presented in Chapter 8 of *BR*, my sympathy is constantly with the sceptic?

The problems for me lie with "rational analysis" and the manner in which Oaksford & Chater (O&C) go about doing their science. Rarely do we see anything approaching a descriptive or algorithmic account of reasoning (the model of syllogistic reasoning being the closest) or even – in spite of their philosophy – a genuinely computational one. What we typically get is an *alternative norms* account. Thus, choices on the Wason selection task look erroneous from the viewpoint of standard logic, but can be seen as rational attempts to maximise information gain (given a misreading of the task instructions that makes it about categories instead of four individual cards). Similarly, we can predict people's conditional inference rates on the assumption

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

that they are somehow reflecting the probability of the conclusion given the minor premise (despite the oddity that uncertainty resides exclusively in the major premise), and so on. The object of the exercise seems to be to show that there is some normative account of behaviour that can be offered. As a cognitive psychologist, I find this unsatisfying, because I want to know what people are actually doing and how.

The basic philosophy of the authors is set out in Chapter 2, on which I will focus. They start out by describing essentially the same paradox as the one that inspired our earlier book on rationality (Evans & Over 1996a). Why is it that people's reasoning and decision making seems essentially rational, intelligent, and adaptive in the real world, but results in numerous violations of logic and other standard normative systems when they are tested in the laboratory? We suggested that while people have some limited capacity for explicitly following normative rules (rationality2), they could often achieve everyday goals by implicit processes such as associative learning (rationality1). O&C (*BR*, p. 29) object, however, to any form of rationality that cannot be justified by a normative system. But elsewhere in the book, they seem to be describing something very similar to rationality1. For example, they describe some examples of "rational" behaviour with the comment: "Such behaviour may be built in by evolution or acquired via a long process of learning – but need not require on-line computation of the optimal solution" (*BR*, p. 36). Examples discussed here include Gigerenzer's research programme on fast and frugal heuristics (Gigerenzer 2004). But Gigerenzer appeals to ecological and not normative rationality, so the link between the two programmes is unclear.

The authors overstate everyday rationality when they ask (*BR*, p. 30), "*why* do the cognitive processes underlying everyday rationality consistently work?" Well, they don't, and to say otherwise is to ignore a massive body of psychological literature. For example, in the study of judgement and decision making, many cognitive biases have been shown to occur reliably in everyday contexts and with expert judges: phenomena such as outcome bias, hindsight bias, overconfidence, and the planning fallacy, to name just a few (for evidence, see papers from two recent collections on these topics: Gilovich et al. 2002; Koehler & Harvey 2004). There is also a massive accumulation of evidence for dual-process accounts of reasoning and decision making (Evans 2007; 2008), something which plays no visible role at all in the account of human reasoning offered by O&C. Nor do O&C feel moved to explain why the achievement of *standard* normative solutions to the task for which they provide *alternative* normative accounts is associated with those of high general intelligence (Stanovich 1999). It is particularly disappointing that they make no reply to Stanovich's carefully constructed challenge to their position.

A final issue is whether their methodology provides computational or normative accounts of behaviour, because these are not the same thing (Elqayam 2008). The authors try to combine them as follows: "Formal rational principles spell out the optimal solution for the information processing problem that the agent faces. The assumption is that a well-adapted agent will approximate this solution to some degree" (*BR*, p. 38). But the adaptive mechanisms of nature – evolution and learning – do *not* optimise. If evolution had an optimal solution, we would not have billions of different species of every conceivable form and function. If learning was optimised, animals would not, for example, generally match the probability of their foraging to that of food sources. But O&C are quite explicit (*BR*, pp. 30–31) that they are not worried about the performance of a priori normative theories in providing such quasi-computational accounts. On the contrary, they are looking for *empirical rationality*, which means in essence: observe some behaviour, assume that it is rational, find a normative theory that deems it to be so, and then . . . nothing else, apparently.

I can understand the benefits of rational analysis when it is rooted in ecological rationality, so that by assuming that behaviour is adapted to the environment, we can look to the structure of that environment for cues as to how our cognitive mechanisms are designed. However, I really don't understand why O&C feel the need to fit a normatively rational model to human reasoning, and still less why that should put an end to the inquiry, with no algorithmic account to follow. Nevertheless, I agree strongly with them that the Bayesian model is a far more appropriate reference for real world reasoning than one based on truth-functional logic, and that it is a standard much more likely to be approximated in the inferences that people actually make.

The strengths of – and some of the challenges for – Bayesian models of cognition

doi:10.1017/S0140525X0900034X

Thomas L. Griffiths

Department of Psychology, University of California, Berkeley, Berkeley, CA 94720-1650.

tom_griffiths@berkeley.edu

http://cocosci.berkeley.edu

Abstract: *Bayesian Rationality* (Oaksford & Chater 2007) illustrates the strengths of Bayesian models of cognition: the systematicity of rational explanations, transparent assumptions about human learners, and combining structured symbolic representation with statistics. However, the book also highlights some of the challenges this approach faces: of providing psychological mechanisms, explaining the origins of the knowledge that guides human learning, and accounting for how people make genuinely new discoveries.

Oaksford & Chater's (O&C's) book *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) is at heart a review of the authors' research program exploring probabilistic models of human deductive reasoning. However, the review is nested within a set of chapters that make clear the ambitious goal of this research program: not just to explain how people reason, but to provide us with a new functional notion of rationality. This work, together with that of Anderson (1990) and Shepard (1987; 1995) was one of the early successes of rational analysis and Bayesian models of cognition (see also, Chater & Oaksford 1999a). As such, *BR* illustrates the strengths of rational analysis, and helps to highlight the challenges that Bayesian models face.

One of the strengths of rational analysis is producing models that explain human behavior as an adaptive response to problems posed by the environment. Establishing how the notion of rationality appealed to by these explanations relates to traditional normative analyses in psychology and economics is one of the themes of the early chapters of the book, building on previous discussion by Anderson (1990). Loosely speaking, the thesis is that rationality can be a methodological assumption that guides us to explanations of behavior: Rather than deciding that people solve a problem poorly, we should consider the possibility that they are solving another problem well, and try to determine what that problem might be. This analysis can be more satisfying than traditional uses of normative models, in that it produces a systematic picture of the assumptions guiding human inferences rather than a list of deviations. This approach is perhaps best exemplified in *BR* by the Wason selection task, in which the traditional normative account is a combination of falsificationism and deductive logic, and the alternative construes the problem as one of confirmation and induction.

Rational accounts of cognition have virtues that go beyond mere systematicity. One of the most important is the explicit treatment of the assumptions of learners that results from explaining behavior as an adaptive response to computational problems. Computational models of cognition typically focus on the processes that govern behavior. Their content and

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

parameters concern the properties of hypothetical cognitive mechanisms – similarity, strength of association, and learning rates. The content and parameters of rational models are expressed in a different language, being about the problems learners solve and the assumptions they make. As a consequence, these models are more abstract, but more explicit about their commitments, providing a transparent account of what learners should find easy or hard to learn (their "inductive biases") and what factors should influence behavior (such as the rarity of properties in the Wason task). Developing models formulated in these terms can be valuable, because many questions in cognitive science are posed at the same level of abstraction, concerning the constraints that guide human learning.

The treatment of human reasoning presented in *BR* illustrates another strength of Bayesian models: the capacity to combine symbolic representations with statistical inference. As O&C discuss, logic and probability have often been viewed as distinct. Researchers in philosophy and artificial intelligence have touted the expressive power of deductive logic, whereas probability theory is typically applied to simpler representations such as sequences of coin flips. The models developed in the book begin to combine the expressive power of logic with the inferential capacity of probability, following a trend towards probabilistic logics that is now a hot topic in artificial intelligence (e.g., Friedman et al. 1999; Milch et al. 2004). More generally, the idea of combining structured representations with statistical inference brings together two traditions in cognitive science, where symbolic models employ structured representations and connectionism demonstrates power of statistical learning. As discussed in the book, the potential of the resulting synthesis is great, with structured statistical models being useful when both representation and learning are key, such as inferring the existence of causal relationships (e.g., Griffiths & Tenenbaum 2005) and learning and using language (e.g., Chater & Manning 2006).

While the examples presented in *BR* illustrate the strengths of rational analysis and Bayesian models of cognition, they also highlight some of the challenges that this approach faces. One significant challenge is connecting rational models to more conventional psychological explanations – identifying mechanisms that can support the sophisticated probabilistic inferences required by structured statistical models. Another challenge is understanding where the knowledge that guides these inferences comes from – the source of the priors and hypothesis spaces for Bayesian inference. And finally, the parallels between induction and scientific reasoning drawn in the book raise another compelling question. Following Reichenbach (1938), philosophers of science distinguish between discovery – developing genuinely new hypotheses – and justification – explaining why one hypothesis should be preferred to another. Bayesian inference, at least in the form presented in most Bayesian models, is a logic of justification, with all possible hypotheses being evaluated. So, how might Bayesian models account for discovery, something that would seem to require going beyond the hypotheses available?

These are significant challenges, but they might be overcome by drawing on work in other disciplines that use Bayesian models. Probabilistic inference is not just hard for people, it is difficult for computers too. The resulting work on approximate inference in computer science and statistics provides a potential source of psychological mechanisms (e.g., Sanborn et al. 2006; Shi et al. 2008). Hierarchical Bayesian models, originally developed in statistics to capture both individual and population effects, provide a way for learners to start out with vague priors and to refine their beliefs over time to develop more informative expectations (Tenenbaum et al. 2006). Finally, nonparametric Bayesian models use stochastic processes to define priors over infinite, structured hypothesis spaces, allowing new observations to be explained through new hypotheses (e.g., Sanborn et al. 2006; Griffiths & Ghahramani 2006), and perhaps paving the way towards a deeper understanding of discovery. Set in this broader context, *BR* is a significant step towards a genuinely interdisciplinary science of induction, connecting human minds with algorithms from computer science and the abstract principles of statistics.

ACKNOWLEDGMENTS

I would like to thank J. Austerweil, F. Eberhardt, C. Lucas, F. Reali, J. Vuong, J. Williams, and other members of the Computational Cognitive Science Lab at Berkeley for helpful discussions of the book.

Explaining more by drawing on less

doi:10.1017/S0140525X09000351

Ulrike Hahn

School of Psychology, Cardiff University, Cardiff CF10 3AT, United Kingdom.
hahnu@Cardiff.ac.uk

Abstract: One of the most striking features of "Bayesian rationality" is the detail with which behavior on logical reasoning tasks can now be predicted and explained. This detail is surprising, given the state of the field 10 to 15 years ago, and it has been brought about by a theoretical program that largely ignores consideration of cognitive *processes*, that is, any kind of internal behavior that generates overt responding. It seems that an increase in explanatory power can be achieved by restricting a psychological theory.

It is undoubtable that over the last decade, Oaksford & Chater (O&C) have revolutionized the psychology of human reasoning. What were previously viewed as "logical reasoning" tasks, such as Wason's card selection task, were given a probabilistic reinterpretation. At the same time, their work has formed a corner-stone in the wider probabilistic revolution that has taken place over the last 10 to 15 years: Finding evidence for human sensitivity to probabilities in as unlikely a corner as "logical" reasoning tasks, lends support to the wider claim that human cognition is fundamentally about probabilities.

Given the current popularity of Bayesian approaches to cognition (see e.g., Chater et al. 2006; Chater & Oaksford 2008, for overviews), it is worth trying to remind oneself just how much has changed in cognitive science in that period. At the beginning of the 1990s, logic was taught extensively in philosophy and cognitive science courses, but little or no mention was given to probability theory; now core topics such as epistemology are incomprehensible without a basic understanding of probability. Cognitive science as a field was still in the throws of the symbolic versus connectionist debate (e.g., Fodor & Pylyshyn 1988; Smolensky 1990) and both Nick Chater and Mike Oaksford were involved in the connectionist project (e.g., Bullinaria & Chater 1995; Oaksford & Brown 1994). Arguably, the rise of connectionism was a major factor in cognitive science's statistical "turn" (e.g., Chater 1995), but connectionism's emphasis was firmly on "mechanism" and its (potentially) "brainstyle" processing.

However, since then, the connectionist project has stagnated by comparison to what has been achieved in other areas, such as the psychology of reasoning. Crucial to this latter success, it seems, was the step back, or better, "upwards," in terms of the degrees of abstraction, and levels of description at which explanation were sought.

Retreating from considerations of implementation, or, by and large, even algorithms, O&C have stuck resolutely to computational level descriptions, and, in so doing, have become key exponents of Anderson's project of rational analysis (Anderson 1990; Chater & Oaksford 2008; Oaksford & Chater 1998b). To this day, computational level descriptions provoke suspicion in mainstream experimental psychology. "But what is it that people are actually *doing?*", "what is going on in the head?" are standard questions to researchers focused on the computational level, as is exemplified in the dialogue between Sceptic and Advocate in the final chapter of O&C's new book, *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*).

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

All are agreed that the project of understanding human behavior is not over until processes and their implementation have also been specified, but the thing that comes across most strikingly in O&C's book is the level of *explanatory specificity* that has now been achieved in the psychology of reasoning. Before O&C's seminal paper on the selection task (Oaksford & Chater 1994), data in the psychology of logical reasoning were a (in many ways somewhat haphazard) collection of qualitative phenomena ("context effects," "supression effects," etc.). Reading O&C's summary of the state of affairs 14 years on, the psychology of reasoning has become an arena in which detailed *quantitative* predictions are evaluated. This is not only true of their own probabilistic approach; O&C have also forced the hand of rival approaches which have since taken on board quantitative model evaluation (Klauer et al. 2007; Oberauer 2006; Schroyens & Shaeken 2003).

It is worth pausing to consider how remarkable this is. The reasoning tasks in question are firmly "high-level cognition" of the kind that is characteristic of what Fodor (1983) considered to be "central processing," and hence an aspect of human cognition for which we would never have detailed theories and predictions due to the inherent open-endedness of high-level thought. That the field can capture subtle changes in behavior in response to changes in the content of high-level, verbal experimental materials in such detail is truly a success. Moreover, this specificity has been spreading through other aspects of human reasoning, as well (e.g., Tenenbaum et al. 2007).

That greater predictive power with regard to human behavior might be achieved by taking a step backwards to a higher level of abstraction that disregards processes might seem counterintuitive, but it seems to be exactly what has happened in the psychology of reasoning.

One further aspect of O&C's work deserves mention in this context. While it has, in one way, considered "less" than has other research in the psychology of reasoning, there is also a way in which it has consistently considered "more": Apparent throughout O&C's book is the multi-disciplinarity of their project which has drawn support from both philosophy and Artificial Intelligence (AI). We disregard developments in adjacent disciplines at our peril: The problems of associating natural language *if. . .then* statements with the material conditional were long known in philosophy, but ignored by psychologists of reasoning (see Edgington 1995; Evans & Over 2004). Likewise, developments in AI spoke to the feasibility of the "logicist" program. Yet interdisciplinarity continues to be less of an asset and more of a stumbling block for publication in many of the core psychology journals that have housed the psychology of reasoning. At the same time, cognitive science as a discipline seems, if anything, to have lost in appeal over the last decade. Certainly in the United Kingdom, the number of degree courses in cognitive science has gone down – a trend that contrasts with the booming interest in cognitive neuroscience. Cognitive neuroscience, of course, seems first and foremost concerned with processes and their implementation. Although it undoubtedly will lead to results that are new and exciting in their own right, it seems worth pointing out that this trend is a move in the opposite direction to that taken so successfully by O&C; this should give pause for thought to those interested in high-level cognition.

Complexity provides a better explanation than probability for confidence in syllogistic inferences

doi:10.1017/S0140525X09000363

Graeme S. Halford

School of Psychology, Griffith University, Nathan, Queensland, 4111, Australia.
g.halford@griffith.edu.au

Abstract: Bayesian rationality is an important contribution to syllogistic inference, but it has limitations. The claim that confidence in a conclusion is a function of informativeness of the max-premise is anomalous because this is the least probable premise. A more plausible account is that confidence is inversely related to complexity. Bayesian rationality should be supplemented with principles based on cognitive complexity.

The Bayesian account of reasoning proposed by Oaksford & Chater (O&C) in *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) extends the historic trend away from regarding logic as the science of thought, and toward seeing reasoning as adaptation to the environment. Piaget, who is cited on page 4 of the book as wholeheartedly adopting a logicist conception of the mind, was in fact part of this trend, and he expressly repudiated logicism (Piaget 1957, p. 1). The Bayesian account arguably has potential to revolutionise our understanding of reasoning, but it has limitations. I want to draw attention to some issues that are additional to those that have been identified in the literature.

The heuristics proposed by the Probability Heuristics Model (PHM), that is, min, max, and attachment, operate on an implicit ranking of the informativeness of the four premise moods (All, Some, Some not, None). According to the min-heuristic, individuals tend to draw conclusions that match the mood of the least informative premise, which by information theory is also the most probable proposition in a set of premises. The validity of the min-heuristic is that it reflects constraints between the mood of the min-premise and the mood of the valid conclusion (see Tables 7.2 and 7.3 in *BR*). Thus, PHM implies that reasoning depends on knowledge of constraints contained in world knowledge. PHM accounts for approximately the same proportion of variance in problem forms as Mental Models theory and Relational Complexity theory (Halford et al. 2007). PHM goes beyond these theories in dealing with premises "Most" and "Few," but it only predicts mood of the conclusion, not actual conclusions, and it depends on estimated parameters.

There is also a significant anomaly. The max-heuristic determines confidence in the conclusion in proportion to the informativeness of the max-premise, but the most informative premise is the least probable. There does not appear to be an explicit explanation for confidence being inversely related to probability, whereas selection of conclusions, by the min-heuristic, is directly related to probability. An alternative hypothesis is that high confidence is associated with lowest *complexity*, because the least probable forms according to PHM are also the least complex according to complexity metrics, including number of mental models, and the relational complexity metric. The simplest syllogisms according to the Relational Complexity metric (Halford et al. 2007), those based on binary or ternary relations, have at least one A (All) premise, or at least one I (Some) premise. A (All) premises are the most informative (least probable) and I (Some) premises are second most informative in PHM. Thus, the most informative, and least probable, syllogisms are of low complexity, which is a more plausible basis for confidence than low probability. Therefore, PHM might work better if it incorporated other principles that have been established in cognitive science, including those that define cognitive complexity of reasoning tasks.

Are stomachs rational?

doi:10.1017/S0140525X09000375

Elias L. Khalil

Department of Economics, Monash University, Clayton, Victoria 3800, Australia.
elias.khalil@buseco.monash.edu.au
www.eliaskhalil.com

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Abstract: Oaksford & Chater (O&C) would need to define rationality if they want to argue that stomachs are not rational. The question of rationality, anyhow, is orthogonal to the debate concerning whether humans use classical deductive logic or probabilistic reasoning.

In *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*), the authors offer a definitive answer to what seems obvious to them: Although our stomachs can be "optimally efficient" in digesting the food in our environment, we still would not describe them as rational (*BR*, p. 38). For Oaksford & Chater (O&C), the question of rationality of stomachs does not even arise: Stomachs are not information-processing devices, have no beliefs/desires/knowledge, and make no decisions/inferences. O&C use the stomach example in the context of pressing their distinction between rationality and optimization: Optimization is about a "local" end (such as digestion), whereas rationality "requires not just optimizing something but optimizing something reasonable" (p. 37). What is ironic about O&C's "rationality analysis" approach, borrowed from Anderson (1990), is that their instrumental approach prohibits them from offering such an answer about stomachs. In fact, in the same context, they explicitly state that they would not offer an answer to the "ultimate question of what rationality is" (*BR*, p. 38). That is, if they have no definition of rationality, how could they answer at the abstract level that organs (such as stomachs) can be efficient, but not rational, while human organisms can be rational?

To be clear, O&C recognize that humans exhibit rationality in everyday activity that allows us to discern between good and bad reasoning. O&C, therefore, want to avoid the empirical approach that elevates everyday behavior to the entry point of theorizing, as the literature expounding the Allais paradox attempts to do. O&C correctly note that the empirical approach leads to chaos: If everyday activity is taken as the entry point, there would be no ground to distinguish between good reasoning from bad reasoning.

On the other hand, O&C want to avoid a definition of rationality. They do not want to invite what I call here the "specter of Plato": O&C are afraid of invoking the pitfalls of the logical approach that postulates deductive logic without being informed by everyday, empirical activity. The logical approach, O&C argue, invites the questions of why humans should follow deductive logic and from where such logic arises. Of more importance to the authors' project, experimental results have shown that humans perform poorly in classical logical tasks such as conditionals, Wason selection task, and syllogisms. O&C instead argue that humans reason in a world full of uncertainties and, hence, humans use natural language conditionals and probability assessment of premises when they make inferences. Humans do not use meaningless logical conditionals and deductive reasoning with premises that stand with absolute certainty.

Let us agree with O&C that the probabilistic approach to human reasoning explains why humans, in laboratory settings, are bad at solving tasks formulated by classical logic. But, still, why does the advocacy of a probabilistic approach prevent one from providing a definition of rationality – a definition that is necessary if one wants to argue that stomachs are not rational?

It seems that the issue of defining rationality is *orthogonal* to the debate between the logical approach and the probabilistic approach to human reasoning. Both approaches are about resolving the issue of how people can arrive at *true* beliefs – whether the truth is defined in absolute certainty or in degrees of belief. But the issue of rationality is not about how reasoning can arrive at truth. Rather, rationality is about the use of resources in the most efficient way (Becker 1976, Ch. 1). Organisms are not mainly concerned with becoming professional scientists, where costs of the truth are disregarded. Rather, organisms are mostly interested in harnessing food/energy from their environment in the most effective way in order to enhance their well-being or, what is the same thing, to expand the quantity/quality of offspring. So, the main goal of organisms, with the exception of professional scientists, is not reasoning in order to arrive at true beliefs – even when truth is recognized as degrees of belief. Rather, the main goal of organisms is reasoning that enhances well-being (or fitness). Such a goal entails that organisms may settle for inaccurate hypotheses or beliefs given that more accurate ones can be unjustifiably costly. That is, if the processing of new data – which may upset an older hypothesis or belief – is too costly, given the expected benefit of a new hypothesis of belief, then humans would be better off with the old, incorrect hypothesis.

This elementary definition of rationality stands (or falls) *orthogonally* to the issue of whether agents use classical logical methods or probabilistic methods. And O&C can adopt a definition of rationality without adopting the classical logical method or invoking the specter of Plato. To wit, their insistence, on more than one occasion, that stomachs are not rational implies that they uphold a definition of rationality. O&C must implicitly view rationality as characteristic of the actor, and not merely, appealing to the work of Marr (1982), as an "explanatory tool" (*BR*, p. 35). The idea that rationality is merely a tool, called generally *instrumentalism* and advocated by van Fraassen (1980, Ch. 2; see also Friedman 1953), is critical of *realism*, which in this case means that rationality is characteristic of the decision-maker.

On the view of rationality as efficiency, we only need to assume, at least initially, that preferences are consistent (Kreps 1990), and that agents change their actions in response to a change in the environment (constraints). There is no need to suppose that the actor or entity has a cognitive ability informed by beliefs/desires/knowledge. As long as an entity changes behavior in response to the environment in order to economize on effort expenditure, it is rational. Behavioral ecologists, ethologists, and philosophers have recently started to realize that organisms do process information and respond to the environment in ways that qualify them as rational (e.g., Griffin 1992; Hurley & Nudds 2006; Raby et al. 2007; Vermeij 2004). Some of these organisms have no brains, such as plants and single-cell organisms. The same reasoning can be applied to organs (including stomachs) insofar as organs change pace of function, such as secretions or size, in response to changes in the environment (constraints). There is no need, therefore, to distinguish efficiency from rationality – when rationality is about optimizing "something reasonable." Digestion seems also a reasonable goal. In fact, the efficiency/rationality distinction would beg the question of how to distinguish a reasonable from an unreasonable goal, that is, it would invite circularity of argument, as O&C clearly admit (*BR*, p. 37).

Is the second-step conditionalization unnecessary?

doi:10.1017/S0140525X09000387

In-mao Liu

Department of Psychology, National Chung-Cheng University, Chia-Yi 621, Taiwan.

psyiml@ccu.edu.tw

Abstract: Because the addition of the conditional premise tends to increase modus ponens (MP) inferences, Oaksford & Chater argue that the additional knowledge is assimilated to world knowledge before the Ramsey test is carried out to evaluate $P(q|p)$, so that the process of applying the Ramsey test could become indistinguishable from the process of applying the second-step conditionalization.

In *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*), Oaksford & Chater (O&C) argue that the second-step

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

conditionalization is unnecessary in the MP (modus ponens) argument. However, based on experimental findings, I show that the second-step conditionalization in the MP argument is sometimes necessary and sometimes unnecessary. I further argue that the second-step conditionalization is unnecessary because the conditional probability hypothesis holds. The conditional probability hypothesis, which is implied by the Ramsey test (see Over et al. 2007), states that the subjective probability of a natural language conditional is the subjective conditional probability. I will show that the conditional probability hypothesis holds only when reasoners explicitly evaluate the probability of conditionals, but that it may not hold in the context of making MP inferences.

The second-step conditionalization is experimentally isolated from the first-step conditionalization by adding the reduced problem to the complete problem. The reduced MP problem is of the form: Given p, how probable is q? The complete MP problem is the original MP problem of the form: If p then q; given p, how probable is q? Thus, reduced MP problems measure the result of the first-step conditionalization, whereas complete MP problems measure the result of performing both steps in succession.

To argue that the second-step conditionalization is unnecessary is tantamount to arguing that complete MP problems can be replaced by reduced MP problems. In other words, the major premise does not play its role in a complete MP problem. Thus, only studies that administered both reduced and complete problems in the same experiments can provide evidence of whether the major premise does play a role in complete MP problems. There are several studies – for example, those of Chou (2007), Liu (2003), Liu and Chou (2008), Liu et al. (1996), and Wang (1999) – that manipulated several variables to study their effect on reduced and complete MP problems in the same experiments.

Three variables are known to affect reduced MP problems differently from complete MP problems. The first variable, perceived sufficiency, affects complete problems only by affecting the reduced problems. Because of a ceiling effect, reasoners' responses in solving the reduced and complete MP problems are generally identical in the high sufficiency condition. The second variable, problem content, affects the complete MP problem without affecting the reduced MP problems. The third one, age differences, is a quasi-variable: Primary school children raised in a rural area of Taiwan, when tested several years ago, exhibit identical responses in solving reduced and complete MP problems, because they are still unable to perform the second-step conditionalization.

O&C realize that adding the conditional premise tends to increase endorsement of the inferences. However, to account for the Ramsey test (1931/1990a), they argued that the additional knowledge the conditional premise provides is assimilated to world knowledge before the Ramsey test is carried out to evaluate $P(q|p)$.

The Ramsey test (1931/1990a) is compatible with recent experimental findings that tend to support the conditional probability hypothesis (Evans et al. 2003; Oberauer & Wilhelm 2003; Over et al. 2007). On the surface, however, the conditional probability hypothesis is incompatible with the vast literature on MP findings. In particular, the MP is nearly perfectly endorsed when the conditional statement involves abstract content (e.g., Evans 1977; Taplin 1971; Taplin & Staudenmayer 1973).

More specifically, Evans (1977) used the following conditional to generate MP problems: If the letter is G, then the number is 9. The conditional probability of the number 9, given the letter G, would be extremely low. Since Evans found that his participants perfectly endorsed MP problems generated by this type of abstract conditional, the conditional probability interpretation of a conditional statement becomes questionable. Working with conditionals phrased in terms of thematic materials, such as "If Mary has an essay to write, she will stay late in the library," Byrne et al. (1999) also observed a mean of 95% endorsement of simple MP inferences in their experiment. Again, the conditional probability of Mary staying late in the library, given that she has an essay to write, would generally be small.

Faced with this complex array of empirical findings, O&C argue that the conditional-statement premise provides additional evidence that q and p are related, thus increasing the assessment of $P(q|p)$. They then argue that this additional knowledge is assimilated to world knowledge before the Ramsey test is carried out to evaluate $P(q|p)$. Empirically, there is an increase in $P(q|p)$ in conditionalizing the result of the first-step conditionalization on the conditional-statement premise when reasoners are able to detach it from reality (Liu 2003). In other words, although $P(q|p)$ observed from the reduced MP problems of abstract and thematic contents could be the same, the response level observed from the complete MP problems is generally higher for abstract than for thematic content. On the basis of this empirical finding, it is impossible to know whether the process of applying the Ramsey test (1931/1990a) is the same as the process of applying the second-step conditionalization.

The fact that reasoners might consider the antecedent as sufficient for the consequent in reading the conditional-statement premise (e.g., Byrne et al. 1999; Evans 1977) may not contradict the conditional probability hypothesis. This is because the former finding is observable in the context of MP inferences, whereas the conditional probability hypothesis receives its support through the reasoners' direct evaluation of the conditional statement.

ACKNOWLEDGMENTS

This study was supported by NSC Grant 96-2413-H-194-010-MY2. I thank Peter Schonemann for reading an early draft.

Bayes plus environment

doi:10.1017/S0140525X09000399

Craig R. M. McKenzie

Rady School of Management and Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0553.

cmckenzie@ucsd.edu

http://psy.ucsd.edu/∼mckenzie/

Abstract: Oaksford & Chater's (O&C's) account of deductive reasoning is parsimonious at a local level (because a rational model is used to explain a wide range of behavior) and at a global level (because their Bayesian approach connects to other areas of research). Their emphasis on environmental structure is especially important, and the power of their approach is seen at both the computational and algorithmic levels.

Oaksford & Chater (O&C) are to be commended for their comprehensive account of deductive reasoning in *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*). They don't just explain Wason's selection task, or conditional reasoning, or syllogistic reasoning, but all of these. Furthermore, the fact that their account is Bayesian is important for several reasons. First, an inductive account of deductive behavior is interesting and novel. Second, because the pattern of data which they explain consists largely of deductive errors, they are providing an alternative rational view of these "errors." Third, because their account is both rational and explains behavior in these tasks, there is an important sense in which readers (such as myself) gain an understanding of *why* people respond as they do in these tasks. The explanation is, essentially, a teleological one. Finally, the approach connects deductive reasoning to other areas of psychological research that are otherwise disconnected, such as vision, categorization, and language, all of which are influenced by

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Bayesian approaches. In short, O&C offer an account that is parsimonious both locally, in the sense that it is a rational framework that can explain a wide range of deductive behavior, and globally, in that it ties in with a much broader collection of diverse research topics.

Especially important is the environmental aspect of the "rational analysis" that they pursue (Anderson 1990). Researchers tend to point to the many examples of purportedly irrational behavior as evidence against a rational approach (e.g., Gilovich et al. 2002; Kahneman et al. 1982; Kahneman & Tversky 2000). Often overlooked by researchers, though, is that people make strong and reasonable assumptions about the structure of their environment. Unaware of these assumptions, researchers can draw misleading conclusions about behavior. For example, O&C point out that people appear to make the "rarity assumption," that is, that named categories tend to be small (see also McKenzie & Amin 2002; McKenzie & Mikkelsen 2000; 2007; McKenzie et al. 2001). Without the rarity assumption, some behavior can appear irrational and even nonsensical, but with it, behavior is not only reasonable, it is consistent with a Bayesian approach.

Similarly, McKenzie and Mikkelsen (2007) recently argued that the rarity assumption explains why people, when assessing the correlation between two variables that can be either present or absent, are especially influenced by observing the joint presence of the variables. Traditionally, the large influence of joint presence observations has been seen as an error, but it makes sense from a Bayesian perspective combined with the rarity assumption. A completely different example of the importance of environmental structure comes from framing effects, which refer to situations in which listeners respond differently to logically equivalent utterances (or frames). Framing effects have been considered irrational, but it has been pointed out that speakers do not choose randomly among frames. Instead, speakers choose frames systematically, and listeners know this. Thus, although many frames are logically equivalent, they are not *information equivalent*, and responding differently to them is not irrational (McKenzie 2004; McKenzie & Nelson 2003; Sher & McKenzie 2006; 2008). Understanding the structure of the environment, and what makes sense in light of this structure, goes a long way toward understanding people's behavior.

The power of rational analysis is apparent at the computational level (using Marr's [1982] terminology), where the aim is to understand what problem the organism is trying to solve. O&C's accounts of Wason's selection task and conditional reasoning are essentially computational. But when O&C discuss syllogistic reasoning and their probability heuristics model, they also show what rational analysis can do at the algorithmic level, which is concerned with the *processes* used by the organism to solve the problems it faces. The rational model remains the same, but it becomes clear that behavior, at this level, diverges from the model. Relative to the rational model, people use simplified processes (i.e., heuristics) that lead to systematic errors (biases; e.g., Kahneman et al. 1982). One must be careful here because, as others have pointed out, the "heuristics and biases" paradigm in judgment and decision making has left that field with a "laundry list" of heuristics and biases rather than an overarching theory (Gigerenzer 1996; Krueger & Funder 2004). However, rational analysis severely constrains the set of potential heuristics that people might use, which, to a large extent, would guard against the explosion of heuristics that has plagued the judgment and decision making literature. Furthermore, because this constrained set of plausible heuristics would come from a rational analysis, the adaptiveness of heuristics would be less likely to be a contentious topic, as it often is in the field of judgment and decision making. It would be fascinating (and a little ironic) if rational analysis were to pave the way for the next step in the heuristics and biases paradigm.

Naïve optimality: Subjects' heuristics can be better motivated than experimenters' optimal models

doi:10.1017/S0140525X09000405

Jonathan D. Nelson

Max Planck Institute for Human Development, Berlin 14195, Germany.
jnelson@salk.edu
http://jonathandnelson.com/

Abstract: Is human cognition best described by optimal models, or by adaptive but suboptimal heuristic strategies? It is frequently hard to identify which theoretical model is normatively best justified. In the context of information search, naïve subjects' heuristic strategies are better motivated than some "optimal" models.

Bayesian Rationality (Oaksford & Chater 2007) nicely synthesizes the growing body of research. This approach offers promise to give principled explanation not only of human higher cognitive processes, but also much of perception and animal cognition. Is cognition best described by optimal models, or by adaptive but suboptimal heuristic strategies? Oaksford & Chater (O&C) explain that heuristic strategies approximate human syllogistic reasoning better than optimal Bayesian models. This commentary illustrates, in the context of information search, that it is frequently possible to be mistaken about what model is best motivated; and that naïve heuristic strategies can perform better than "optimal" models!

Consider the task of deciding which of two medical tests to order, assuming cost constraints only allow one test, to best diagnose a patient's disease. We assume here that the patient either has Disease 1 or Disease 2, with equal (50 %) prior probability, and that Test 1 and Test 2 are each either positive or negative. How should a diagnostician decide which test to order?

A great deal of cognitive psychological and statistical thinking (since I. J. Good's [1950; 1975] work) claims that the optimal strategy is to conduct the test with highest *expected Bayesian diagnosticity* (*expected weight of evidence*) or *expected log diagnosticity*. [$T1 = p$ denotes that Test 1 is positive, $T1 = n$ that Test 1 is negative, $D = d1$ that the patient has disease 1, etc.] As shown in Figure 1, the expected Bayesian diagnosticity for Test 1 (i.e., its expected utility, as measured with Bayesian diagnosticity) is:

$$eu_{BD}(T1) =$$
$$P(T1=p) * max\left(\frac{P(T1=p \mid D=d1)}{P(T1=p \mid D=d2)}, \frac{P(T1=p \mid D=d2)}{P(T1=p \mid D=d1)}\right)$$
$$+$$
$$P(T1=n) * max\left(\frac{P(T1=n \mid D=d1)}{P(T1=n \mid D=d2)}, \frac{P(T1=n \mid D=d2)}{P(T1=n \mid D=d1)}\right)$$

Figure 1.

and, as shown in Figure 2, its expected log diagnosticity is:

$$eu_{LogD}(T1) =$$
$$P(T1=p) * \log\ max\left(\frac{P(T1=p \mid D=d1)}{P(T1=p \mid D=d2)}, \frac{P(T1=p \mid D=d2)}{P(T1=p \mid D=d1)}\right)$$
$$+$$
$$P(T1=n) * \log\ max\left(\frac{P(T1=n \mid D=d1)}{P(T1=n \mid D=d2)}, \frac{P(T1=n \mid D=d2)}{P(T1=n \mid D=d1)}\right)$$

Figure 2.

However, many subjects follow the *feature difference strategy* (Nelson 2005; Skov & Sherman 1986; Slowiaczek et al. 1992). This strategy involves calculating the absolute difference in feature likelihoods for each test; for example, as shown in Figure 3:

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

$$fDiff(T1) = \left| P(T1=p \mid D=d1) - P(T1=p \mid D=d2) \right|$$

Figure 3.

and ordering the test with the highest fDiff.

Thus, Skov and Sherman and Slowiaczek et al. concluded that many subjects use a suboptimal heuristic strategy that is highly correlated with the optimal strategy. Remarkably, however, both the claims (1) that Bayesian diagnosticity (and/or log diagnosticity) are theoretically optimal, and (2) that the feature difference strategy only imperfectly approximates optimal behavior, are in disrepute.

Both expected Bayesian diagnosticity and expected log diagnosticity are poorly behaved as optimal models. To illustrate, suppose that Test 1 were positive in 99% of people with Disease 1, and in 100% of the people with Disease 2. Suppose further that Test 2 were positive in 1% of people with Disease 1, and 99% of people with Disease 2. Test 1 leads, on average, to 50.5% probability of identifying the correct disease; Test 2 leads, on average, to 99% probability of correctly identifying the true disease. Clearly, Test 2 would be more helpful than Test 1 to differentiate between the diseases. Yet diagnosticity and log diagnosticity maintain that Test 1 is infinitely more useful than Test 2! Both diagnosticity measures hold that any test that offers greater-than-zero probability of obtaining 100% certainty of the true disease is infinitely useful. This bizarre claim is not a desirable property of an "optimal" model. (In Nelson [2005; 2008] I discuss these and other theoretical flaws with the diagnosticity measures, and how redefining a single point cannot fix them.)

Better-motivated theoretical models of the value of information, such as information gain-KL distance (Lindley 1956; Oaksford & Chater 1994), probability gain (error reduction; cf. Baron's 1981 talk at the Psychonomic Society Meeting, as cited in Baron 1985), and impact (Klayman & Ha 1987, pp. 219–20; Nelson 2008; Nickerson 1996; Wells & Lindsay 1980) behave reasonably in this medical diagnosis scenario, and do not suffer from the diagnosticity measures' aforementioned theoretical flaws.

Does the feature difference strategy also approximate these better-motivated theoretical models? In fact, it *exactly* corresponds to impact! The highest fDiff feature also has the highest impact, irrespective of the prior probabilities of the diseases and the specific feature probabilities (Nelson 2005, footnote 2).

Closer analysis of the supposedly optimal theoretical models used by some experimenters, and the supposedly suboptimal heuristics used by some subjects, showed that the subjects' heuristic strategy corresponds to a normative model (impact) that is theoretically superior to the normative model that the experimenters had in mind! Put in the context of Marr's (1982) levels of analysis, consideration of subjects' behavior at the algorithmic level can inform thinking about the kinds of computational-level models (normative theories) that are most appropriate (also see Chater et al. 2003; Cohen 1981).

Do all subjects use the feature difference strategy? No. As O&C discuss, the means with which information is presented is important. Different people use a variety of strategies, especially when environmental probabilities are presented in the *standard probability format*, with explicit prior probabilities and likelihoods. The standard probability format is not the most meaningful to subjects; frequency formats better facilitate Bayesian reasoning (Cosmides & Tooby 1996; Gigerenzer & Hoffrage 1995). Personal experience of environmental probabilities may be even more effective. When environmental probabilities are learned through personal experience, the vast majority of subjects maximize the probability of a correct guess (*probability gain*), rather than impact or information gain (Nelson et al., submitted). Note that impact (which the feature difference strategy implements) more reliably approximates probability gain than do Bayesian diagnosticity or log diagnosticity (Nelson 2005), and impact is easily calculated when the standard probability format is used.

Is cognition optimal? Adaptation can be impressive. Insects' flight length distributions appear well-calibrated to natural environments (Viswanathan et al. 1999). But the modern world is evolutionarily novel. For instance, sugar, fat, and salt are available in unprecedented abundance. Similarly, modern media may exaggerate the incidence of plane crashes versus car crashes, or terrorism versus heart disease. The increasing rate of human genetic evolution (Hawks et al. 2007) may facilitate adaptation to some modern environments, over phylogenetic time.

Among topics of interest in Bayesian and rational analysis, such as perception (e.g., Hoffman, in press), memory, information search, and category formation, the correct function to optimize is seldom clear. Baron (2004) noted that utilities are formed on the basis of reflection, and are constantly being modified. As a pragmatic matter, cognitive science would be wise to treat candidate normative models in similar fashion (also see McKenzie 2003). When there are clear and robust discrepancies between human behavior and a particular theoretical model, the normative status of the theoretical model should be reconsidered, as well as the rationality or adaptiveness of the human behavior.

ACKNOWLEDGMENTS

The author thanks Mike Oaksford and Nick Chater for their insightful suggestions. The writing of this commentary was supported by NIH MH57075 (Garrison W. Cottrell, P.I), when the author was at the Computer Science and Engineering Department, UCSD, and at the Computational Neurobiology Lab, Salk Institute.

Oaksford & Chater's theory of reasoning: High prior, lower posterior plausibility

doi:10.1017/S0140525X09000417

Klaus Oberauer

Department of Experimental Psychology, University of Bristol, Bristol BS8 ITU, United Kingdom.

k.oberauer@bristol.ac.uk

http://psychology.psy.bris.ac.uk/people/klausoberauer.htm

Abstract: Oaksford & Chater (O&C) subscribe to the view that a conditional expresses a high conditional probability of the consequent, given the antecedent, but they model conditionals as expressing a dependency between antecedent and consequent. Therefore, their model is inconsistent with their theoretical commitment. The model is also inconsistent with some findings on how people interpret conditionals and how they reason from them.

In *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) the authors present a strong theoretical case for the "probabilistic turn" in the psychology of reasoning. I agree with much of the general thesis of the book: People often reason from uncertain information, and they do so by drawing on probabilistic information. Conditionals, which form the backbone of much of our knowledge, express conditional probabilities. I disagree with Oaksford & Chater (O&C), however, in details of their models of how people reason, and I am less sanguine about the evidence supporting these models. I focus on reasoning with conditionals.

O&C's model of reasoning from conditionals is based on a contingency table of the antecedent (A) and the consequent (C). One axiom of their model is that the marginal probabilities, $P(A)$ and $P(C)$, must be constant when the degree of belief in the conditional changes. This is an unfortunate assumption, for two reasons. First, it is implausible. Assume a new drug X is tested, and it turns out that it causes headaches. Thus, we increase our belief in "If a person takes X then they get a headache." To accommodate the increase in $P(\text{headache}|X)$ in one's subjective contingency table, one can either revise $P(\text{headache}|\neg X)$

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

down to hold P(headache) constant, or else revise P(headache) up while holding P(headache|¬X) constant. The latter appears more reasonable – as drug X is taken by more people, the overall rate of headaches will increase, but the probability of headaches in those who refrain from taking X will not change. Revising P(headache|¬X) down would lead to the absurd conclusion that, when many people take X, those who don't will benefit because they get fewer headaches.

Second, holding P(C) constant links the conditional to the probabilistic contrast, that is, the difference between P(C|A) and P(C|¬A). With P(C) held constant, every increase in belief in the conditional, that is, every increase in P(C|A), must be accompanied by a decrease in P(C|¬A), resulting in an increased probabilistic contrast. As a consequence, there is an ambiguity in O&C's model on what a conditional means. Initially, O&C endorse "the Equation," that is, the probability of "If A then C" equals P(C|A), and is independent of P(C|¬A). But later, O&C seem to endorse the view that a conditional is believable to the degree that the probabilistic contrast is high. For instance, they argue that "it is possible to believe a rule strongly that has many exceptions" (*BR*, p. 190), as long as the probabilistic contrast is high, such as "If a child walks home from school, it is abducted." In line with this reasoning, O&C introduce the "independence model" as the alternative to a conditional hypothesis. The independence model means that $P(C|A) = P(C)$, which implies that the probabilistic contrast is zero. Since the independence model is meant to be the alternative to the conditional, they cannot both have high probability. If the conditional is defined by the Equation, however, P(C|A) can be high and at the same time be equal to P(C). For example, the probability of arriving safely on a flight, given one has a window seat, is very high, but not different from the unconditional probability of arriving safely. It follows that the independence model cannot, in general, be the alternative hypothesis to a conditional when the latter is defined by the Equation.

My colleagues and I tested whether people interpret conditionals as simply expressing a high P(C|A) or as expressing a high probabilistic contrast. We found that people's degree of belief in a conditional depended only on P(C|A), not on P(C|¬A), in agreement with the Equation but not with the probabilistic contrast model (Oberauer et al. 2007). This finding demands a revision of O&C's argument in defence of the MP-MT asymmetry (i.e., the finding that people endorse modus ponens more readily than modus tollens) and their explanation of the Wason selection task, which both assume that the independence model is the alternative to the conditional hypothesis.

The evidence for O&C's model of reasoning with conditionals is mixed at best. Evidence comes from three sources: (1) The model fits endorsement rates for the four basic inference forms. Fitting four data points with three free parameters is no convincing accomplishment, though. A richer database is provided by the frequencies of the 16 possible patterns of endorsement or rejection across the four inference forms. I applied seven formal models of reasoning to such pattern frequencies (Oberauer 2006). O&C's model (Oaksford et al. 2000) provided fits that were worse than all competitors. (2) The model can explain established findings such as negation and suppression effects. Other theories, however, can also explain these effects (Evans & Handley 1999; Markovits & Barrouillet 2002). Therefore, these findings do not support O&C's model over alternatives. (3) Direct manipulation of probabilities is arguably the most direct and stringent test, because no competing theory predicts the same effects as the O&C model. There are two series of experiments using this method. One provided support for the O&C model (Oaksford et al. 2000), whereas the other did not (Oberauer et al. 2004). O&C dismiss the latter evidence as difficult to interpret in light of "the large number of findings showing probabilistic effects in the conditional-inference task and in the selection task" (*BR*, p. 204). At least for the inference task, I fail to see this large number of confirmatory findings.

One difference between the experiments of Oaksford et al. (2000) and those of Oberauer et al. (2004) is that we used the standard deductive instruction, asking participants to judge whether the conclusion follows with logical necessity from the premises, whereas Oaksford et al. simply asked whether one can draw the conclusion. This difference points to a distinction between goals of reasoning, which I think is not sufficiently acknowledged by O&C. The goal of deductive reasoning is to evaluate whether an inference is valid, and in experiments investigating deduction people are instructed accordingly. Most experiments cited in support of probabilistic theories of reasoning, however, ask people to evaluate the soundness of inferences or the truth of conclusions. The (sparse) evidence from direct manipulations of probabilities suggests that people can ignore probabilities when asked to judge validity, whereas they draw on probabilities when asked to rate whether the conclusion is true. Such a modulation of reasoning processes by goals would be entirely rational.

To conclude, the probabilistic view on human reasoning has high a priori plausibility, but the version fleshed out by O&C is conceptually ambiguous and not well supported by the data.

Human reasoning includes a mental logic

doi:10.1017/S0140525X09000429

David P. O'Brien

Department of Psychology, Baruch College, City University of New York, New York, NY 10010.

david_obrien@baruch.cuny.edu

Abstract: Oaksford & Chater (O&C) have rejected logic in favor of probability theory for reasons that are irrelevant to mental-logic theory, because mental-logic theory differs from standard logic in significant ways. Similar to O&C, mental-logic theory rejects the use of the material conditional and deals with the completeness problem by limiting the scope of its procedures to local sets of propositions.

In *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) the authors reject the conception of human reasoning that focuses on logical inferences, arguing that probability theory should be used instead to account for rationality. Given space limitations, I here address only the two most prominent reasons Oaksford & Chater (O&C) present to reject logic, arguing that they fail to appreciate what mental-logic theory actually proposes.

First, mental-logic theory (e.g., Braine 1990; Braine & O'Brien 1991; 1998; O'Brien 1993; 2004; O'Brien & Manfrinati, in press) consistently has proposed that mental logic differs from standard logic. O&C equate the logical view of conditionals with the truth table for the material conditional (*if p then q* is true unless *p* is true and *q* is false). Indeed, Oaksford and Chater (2003a) stated that the Braine and O'Brien theory includes the material conditional for *if p then q*. The problem with their criticism is that Braine and O'Brien consistently argued that the material conditional does not capture psychological reality. Our theory of conditionals consists instead of two schemas: one for modus ponens (MP) and another for conditional proof. The conditional proof schema states that to derive or evaluate *if p then q*, first suppose *p*; when *q* follows from the supposition of *p* together with other information assumed, one may assert *if p then q*. The schema is applied with a reasoning program that supposes *p* and then treats *q* as a tentative conclusion to be evaluated. When one evaluates a conditional *if not p then not q* from the premise *p or q*, one concludes that the conditional is false, even though this evaluation would not follow when treating *if* as the material conditional (because *p* might be false). Thus,

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

O&C's criticisms of the logical semantics of the material conditional do not apply to our mental-logic theory.

Second, O&C reject logic as a model for human rationality because, they say, logic needs to provide a formal mechanism that could identify all and only those "common sense inferences that people endorse" (*BR*, pp. 89–90), and such a system would be either incomplete or intractable. (See Rosser [1939] on Gödel's 1931 theorems concerning incompleteness and undecidability.) After rejecting logic as a psychological model because of the incompleteness problem, O&C propose that cognitive science should instead rely on probability theory by applying it to particular domains (e.g., in reasoning about a game of dice), although they acknowledge that their suggested solution does not resolve the problem of knowledge representation identified in logic systems, because the computations also would be intractable if extended beyond working out probabilities for small problem sets. Furthermore, they fail to consider whether a mental logic might be able to provide a similar solution to their Bayesian solution by limiting the need to make decidability judgments to small sets of propositions. Indeed, Braine and O'Brien (1991) described just such a mental-logic approach.

To understand how mental-logic theory deals with which inferences are admissible in constructing lines of reasoning and which are not, consider the reasoning constraints described in Braine and O'Brien (1991). Unlike standard logic, in mental logic nothing follows from a contradiction except the realization that some assumption that led to the contradiction is wrong. Thus, one cannot suppose a proposition that contradicts other propositions already being used in an argument, and one cannot bring a proposition into an argument under a supposition that contradicts the supposition. The nub of the question concerns how one knows whether a supposition, or a proposition being imported into an argument under a supposition, is creating a contradiction (see Braine & O'Brien 1991, p. 184). Clearly, in making such decisions people do not check the whole set of propositions they carry in long-term memory. Further, it is not simply a matter of deciding whether a proposition being considered is true, but whether it still would be true in the argument being constructed under a supposition. O&C, on the one hand, and Braine and O'Brien, on the other, are in agreement that making a global judgment about the consistency of all possible propositions being held in long-term memory is intractable. The mental-logic proposal is to consider the matter only when preparing to introduce a proposition into a line of reasoning, in order to decide whether that proposition would still be true given the supposition. The resolution of this matter does not require one to consider all possible propositions that might be affected by the supposition, for example, to decide the details of the closest possible world, but rather, to consider only the relevant proposition one desires to import into the argument; and the decision criterion concerns whether the reasoner is satisfied that the proposition still would be true under the supposition. When two or more interlocutors are involved, the matter is up for negotiation.

O&C come to their conclusion because they view logic in terms of the sorts of systems that were developed by professional logicians beginning with Hilbert, Frege, Russell, and so on, who attempted to ground number theory in logic. A perusal of a history of logic, such as Kneale and Kneale (1962), reveals a more varied sense of what constitutes logic. The founder of Stoic logic, Chrysippus, suggested that even his dog was able to make simple logic inferences; for example, knowing that when it followed a scent on a road and came to a fork, if the scent did not go down one path, it must go down the other. Surely Chrysippus's dog did not compute all possible inferable propositions in a dog's possible world to make this disjunction-elimination inference, nor must we infer that the dog was computing Bayesian probabilities. Mental-logic theory assumes that the human logical reasoning repertory developed through bio-evolutionary history because our hunter/gatherer ancestors gained an advantage in making simple, direct logic inferences that did not include assessing the consistency of the set of all possible propositions they could have held. An interest in the decidability of intractable sets of propositions awaited the advent of logic as a profession and is not part of the universal logic repertory.

Uncertain premises and Jeffrey's rule

doi:10.1017/S0140525X09000430

David E. Over[a] and Constantinos Hadjichristidis[b]

[a]*Department of Psychology, Durham University, Durham DH1 3LE, United Kingdom;* [b]*Facoltà di Scienze Cognitive, Università degli Studi di Trento, 38068 Rovereto, Italy.*

david.over@durham.ac.uk

k.hadjichristidis@form.unitn.it

Abstract: Oaksford & Chater (O&C) begin in the halfway Bayesian house of assuming that minor premises in conditional inferences are certain. We demonstrate that this assumption is a serious limitation. They additionally suggest that appealing to Jeffrey's rule could make their approach more general. We present evidence that this rule is not limited enough to account for actual probability judgements.

In *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*), Oaksford & Chater (O&C) propose an important probabilistic theory of conditional reasoning. They identify the probability of a conditional, $P(\textit{if } p \textit{ then } q)$, with the conditional probability, $P(q|p)$. Following the normative literature, they call this identity the Equation, and we will also take it for granted. There is substantial direct evidence for the Equation as a description of people's probability judgements (Over et al. 2007). O&C take an impressive indirect approach: They show how much they can explain in the psychology of conditional reasoning using the Equation. However, their main account occupies the halfway Bayesian house of assuming that contingent minor premises are certain.

The simplest application of O&C's main account is to Modus Ponens (MP): inferring q from the major premise *if* p *then* q and the minor premise p. They assume that the probability of the minor premise is 1. They thus hold that the probability of q after MP, $P(q)$, is $P(q|p)$, with $P(p) = 1$. Their assumption is not fully Bayesian. For a full Bayesian, committed to strict coherence, a contingent proposition should not have a probability of 1. That would imply that it could never be disconfirmed. O&C also assume that the minor premise is certain in their treatment of other conditional inferences. To illustrate the limitations of this assumption, consider Modus Tollens (MT). This is inferring, for example, "Global warming will not continue" (*not-p*) from "If global warming continues then London will be flooded" (*if p then q*) as the major premise and "London will not be flooded" (*not-q*) as the minor premise. (Over et al. 2007, has evidence that conditionals like this satisfy the Equation.)

Looking in detail at O&C's description of MT in temporal reasoning, we begin (using, we hope, clearer symbols) with the prior conditional probability, $P\text{old}(q|p)$, and the prior marginal probabilities, $P\text{old}(p)$ and $P\text{old}(q)$. They derive a formula for MT, by elementary steps in probability theory, for inferring $P\text{old}(\textit{not-}p|\textit{not-}q)$ from those prior probabilities. They suppose that we later "learn" the minor premise, *not-q*, and so acquire a new belief state. They take "learn" in the extremely strong sense of meaning that $P\text{new}(\textit{not-}q) = 1$. This allows them to identify the probability of the conclusion of MT, $P\text{new}(\textit{not-}p)$, with $P\text{old}(\textit{not-}p|\textit{not-}q)$. They note that we may lower our belief in the major premise of MT when we learn *not-q* (Stevenson & Over 1995): learning the minor premise sometimes "alters" $P\text{old}(q|p)$ by lowering it. We use, they hold, this "modified" lower probability, along with $P\text{old}(p)$ and $P\text{old}(q)$, in their derived formula to recalculate $P\text{old}(\textit{not-}p|\textit{not-}q)$, and that is

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

then taken as a revised *P*new(*not-p*). However, when the probability of the minor premise, *not-q*, becomes 1, the probability of *q* becomes 0. The probability of the major premise, *if p then q*, should consequently collapse to 0 (because the conditional probability is then 0), and not merely be "modified" to varying relatively low values. Of course, there is no technical problem in putting *P*old(*q*|*p*)=0 into O&C's formula to recalculate *P*old(*not-p*|*not-q*). The problem is that making *P*old(*q*|*p*)=0 in every case is a serious limitation. (There is a technical problem if the probability of *p* becomes 0, since that makes the conditional probability undefined. We set this possibility aside, although it is a problem for O&C's account of denying the antecedent.)

Stevenson and Over (2001) found that people had more confidence in some minor premises than in others, depending on the level of expertise of the person who asserted the premise. Their results suggest that, even if people heard an expert on the Thames Barrier assert "London will not be flooded," they would not assign it a probability of 1, and might not even judge it to have a very high probability. Of course, people sometimes claim that they are "certain" or "sure" of a contingent proposition, but we should interpret them as saying that their confidence in it is relatively high, and not that it can never be disconfirmed. They might, in turn, interpret an experimenter who tells them to assume that a minor premise is "certain" as merely indicating that its probability is high. But what confidence do people have in the conclusion of a conditional inference as they become more (or less) confident of the minor premise? Rightly starting to move on from their halfway house, O&C suggest that a way to answer this question is to adopt Jeffrey's rule of conditionalization (Evans & Over 1996a; Jeffrey 1983).

The relevant form of Jeffrey's rule for MP is:

$$P\text{new}(q) = P\text{old}(q|p)P\text{new}(p) + P\text{old}(q|not-p)P\text{new}(not-p)$$

The normative status of this rule is debated (Howson & Urbach 1993), but to what extent do people generally conform to it? We and our collaborators investigated this question using categorical inferences (Evans et al. 2004; Hadjichristidis et al., in preparation). An example is inferring "Falcons have the neurotransmitter Dihedron" (*q*) from "Robins have the neurotransmitter Dihedron" (*p*). One way to look at these inferences is to see them as having an implicit conditional (*if p then q*) as a major premise. We used a plausible manipulation of *P*new(*p*), for example, "Scientists are 80% certain that Robins have the neurotransmitter Dihedron." We analysed our data with a regression that aimed to predict judgements of *P*new(*q*) using as predictors *P*old(*q*|*p*)*P*new(*p*) and *P*old(*q*|*not-p*)*P*new(*not-p*). Only *P*old(*q*|*p*)*P*new(*p*) had a consistent influence on participants' judgements about *P*new(*q*).

These results imply that O&C are too limited, in their main account, in making *P*new(*q*) depend only on *P*old(*q*|*p*). But the results also imply that they are not limited enough in later suggesting, more generally, that *P*new(*q*) will be determined by the full use of Jeffrey's rule. The evidence so far is that *P*new(*q*) will primarily depend on *P*old(*q*|*p*)*P*new(*p*) and so, on just the probabilities of the major and minor premises. However, these results are only for one kind of implicit MP inference. The descriptive adequacy of Jeffrey's rule should be much more extensively investigated.

Mental probability logic

doi:10.1017/S0140525X09000442

Niki Pfeifer and Gernot D. Kleiter

Department of Psychology, University of Salzburg, A-5020 Salzburg, Austria.
niki.pfeifer@sbg.ac.at
http://www.users.sbg.ac.at/~pfeifern/
gernot.kleiter@sbg.ac.at
http://www.users.sbg.ac.at/~gdkleiter/

Abstract: We discuss Oaksford & Chater's (O&C's) probabilistic approach from a probability logical point of view. Specifically, we comment on subjective probability, the indispensability of logic, the Ramsey test, the consequence relation, human nonmonotonic reasoning, intervals, generalized quantifiers, and rational analysis.

Probability logic investigates probabilistic inference and relates it to deductive and other inferential systems. It is challenging to relate human deductive reasoning to probability logic rather than to logic. Oaksford & Chater (O&C) were the first who systematically investigated human deductive reasoning within a probabilistic framework. Our approach to human reasoning is, in many respects, related to O&C's. However, our approach is closer to probability logic, especially with respect to the analysis of the experimental tasks. In commenting on *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*), we discuss a selection of questions arising from these different perspectives.[1]

Common everyday reasoning requires us to process incomplete, uncertain, vague, and imprecise information. Artificial Intelligence (AI) has developed many different approaches for uncertain reasoning. Typical examples are belief functions, possibilistic models, fuzzy systems, probabilistic description languages, many-valued logic, imprecise probabilities, and conditional independence models. Of these approaches only conditional independence models are considered in the book. Why do O&C consider probability and not, say, belief functions as a normative reference system?

The probabilities are interpreted as subjective probabilities. The theory of subjective probability was conceived by Bruno de Finetti (1974/1975), and further developed by Coletti and Scozzafava (2002), Gilio (2002), and many others. A central concept of subjective probability theory is *coherence*. A probability assessment is coherent if it cannot lead to sure losses. The theory does not require event structures that are closed under negation and conjunction. Conditional events are primitive. Conditional probabilities are not defined by absolute probabilities. When we talk about *P*(A|B), why should we assume that we also know *P*(A and B) and *P*(B)? These properties make the *coherence based probability logic* more appropriate as a framework for psychological research than other approaches, including the "pragmatic strategy of using the simplest probabilistic semantics" (*BR*, p. 75).

O&C argue that classical logic should be replaced by probability theory as a framework for human reasoning. This position is too radical. We should not "throw out the baby with the bath water." Probability theory presupposes logic for operations on propositions. Simple logical inferences like And-Introduction or Modus Ponens (MP) are endorsed by practically all subjects. We do not see a dichotomy between logic and probability.

We fully support O&C's hypothesis, that human subjects understand the indicative "if A, then B" in the sense of a conditional probability, *P*(B|A), and not as the probability of a material conditional, *P*(A ⊃ B). Many empirical studies corroborate this hypothesis (Evans & Over 2004).

In subjective probability theory, conditional events are not truth-functional. If the antecedent is false, then the truth value of the conditional is undetermined. This corresponds to what is called a "defective truth table." Considering *P*(B|A), we would not say the Ramsey Test adds *P*(A)=1 to one's stock of belief. Rather, A is *assumed* to be true. Probability 1 and the truth value TRUE are not the same. Ramsey's idea can be explained in the way Lindley (2006) and several others introduce conditional probability. Subjective probabilities are assessed relative to a knowledge base *K*. The absolute probability *P*(A) is shorthand for *P*(A|*K*), and *P*(B|A) is shorthand for *P*(B|A:*K*). The colon separates the supposition from the knowledge base. The change from supposition to fact does not change the conditional probability, *P*(B|AC:*K*) = *P*(B|A:C*K*).

The core of O&C's analysis of the conditional inferences (Modus Ponens [MP], Modus Tollens [MT], Affirming the Consequent [AC], and Denying the Antecedent [DA]) is "that the probability of the conclusion ... is equal to the conditional probability of the conclusion given the categorical premise" (*BR*, p. 119), *P*(conclusion|categorical premise). Thus, the consequence relation

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

(denoted by a horizontal line and three dots in the book) is *probabilistic*. In our approach the consequence relation is *deductive*. Each premise obtains a probability. The probabilities of the premises are propagated *deductively* to the conclusion. Because the MP, MT, AC, and DA arguments consist of only two premises, only an *interval probability* can be inferred (Pfeifer & Kleiter 2005b; 2006). A conclusion with a point probability would require three premises. In this case, however, the argument does not "mimic" the logical forms of MP, MT, AC, or DA. O&C mention probability intervals, but they do not use them.

In Chapter 4, O&C oppose nonmonotonic reasoning and probabilistic reasoning, and advocate probabilities. We do not see why both approaches are incompatible. *System P* (Kraus et al. 1990) is a basic and widely accepted nonmonotonic reasoning system. Several probabilistic semantics were developed for System P (Adams 1975; Biazzo et al. 2005; Gilio, 2002; Hawthorne & Makinson 2007; Lukasiewicz 2005). We observed good agreement between the predictions of the coherence-based semantics and actual human inferences (Pfeifer & Kleiter 2005a; in press a; in press b).

O&C were the first who realized the importance of *generalized quantifiers* in psychology. Murphree (1991) and Peterson (2000) developed logical systems for syllogisms that involve generalized quantifiers (see also Peters & Westerståhl, 2006). On p. 219 of *BR*, O&C note that "without a notion such as *p*-validity [not in Adams' sense!], there is no way of defining the correct answers to these generalized syllogisms." Peterson, however, investigates syllogisms that involve quantifiers like "Most," "Few," or fractionated quantifiers like "n/m." The semantics of the quantifiers works by comparisons of the cardinalities of appropriate sets and by the use of relative frequencies. Thus, Peterson's semantics can easily be related to a probabilistic interpretation. Moreover, *Atmosphere* or *Matching* are easily generalized within this framework.

O&C use Bayesian networks to model syllogisms. Each vertex in the network corresponds to a term in the syllogism. The directed arcs represent conditional probability distributions. We are uncertain about the statement that in Bayesian networks the conditional independence is a "standard assumption" (*BR*, p. 222). Moreover, under the assumption of conditional independence (X independent Z, given Y), there exists only one probabilistic model. Three models (Figure 1, 3, and 4 in Fig. 7.3) are Markov equivalent; only the vee-structure (Figure 2 in Fig. 7.3) has a different factorization. It encodes a marginal independence (X and Z are independent).

Rational analysis puts the quest for cognitive processes and representations in the second line. This is fine if the normative models fit the empirical data. In this case the models are both normative and descriptive. A good theory requires: (1) a thorough task analysis, (2) a minimum of generality to avoid "adhoceries," and (3) a connection with other theoretical concepts like language, memory, or attention.

ACKNOWLEDGMENTS

This work is supported by the European Science Foundation (EUROCORES program "LogICCC") and by the Austrian Research Fonds (project P20209 "Mental Probability Logic").

NOTE

1. We shall not comment on a number of misprints in formulae or inconsistencies in notation.

Popper's *Severity of Test* as an intuitive probabilistic model of hypothesis testing

doi:10.1017/S0140525X09000454

Fenna H. Poletiek

Department of Psychology, Leiden University, 2300RB Leiden, The Netherlands.

poletiek@fsw.leidenuniv.nl

Abstract: Severity of Test (SoT) is an alternative to Popper's logical falsification that solves a number of problems of the logical view. It was presented by Popper himself in 1963. SoT is a less sophisticated probabilistic model of hypothesis testing than Oaksford & Chater's (O&C's) information gain model, but it has a number of striking similarities. Moreover, it captures the intuition of everyday hypothesis testing.

Popper's philosophical principle of falsification is commonly represented as *seeking* and *finding* counter-examples to a hypothesis, as Oaksford & Chater (O&C) show in *Bayesian Rationality* (henceforth *BR*, Oaksford & Chater 2007, p. 167). Indeed, general laws can be definitively falsified, but not definitively confirmed, according to logic. Hence, counter-examples give us sure knowledge that the putative hypothesis is false. Large parts of Popper's work are directed at uncovering the failure of confirmation theory. However, philosophers and psychologists rejected the principle of falsification as obviously incompatible with scientific practice and everyday reasoning. In a natural environment, inside or outside the scientific lab, we cannot live with falsified beliefs alone. Falsified ideas are not publishable and they give poor cues about how to cope with the everyday environment (Poletiek 1996; Poletiek & Berndsen 2000). However, as O&C show, old theories of confirmation have also failed to provide a rational account of hypothesis testing. It was not before the introduction of the Bayesian approach in confirmation theory that a coherent alternative to the falsification method could be advanced.[1] O&C's information gain model provides such an account for hypothesis testing that predicts test preferences in accordance with the information they can provide for the truth and falsity of hypotheses.

There is, however, an interesting alternative version of the idea of falsification, hesitatingly proposed by Popper himself. It avoids the obvious problems of Popper's logical falsification principle, and fits in with O&C's Bayesian approach. (Poletiek 2001). This alternative definition of falsification is putting hypotheses to *severe tests*. I argue here that *Severity of Test* (SoT) provides an intuitive statistical model for what people do when they test hypotheses. As Figure 1 below shows, the formal measure for SoT [2] (Popper 1963/1978; for a discussion, see Poletiek 2001) for a piece of evidence e predicted by the hypothesis H is:

$$S(e,H) = \frac{P(e \mid H)}{P(e)}$$

Figure 1.

What distinguishes SoT from logical falsification, and to what behavior does it correspond? First, SoT is about evidence that *confirms* the hypothesis put to the test, and not evidence that is expected to falsify it. The severity of a test increases as the prior probability of confirmation e is low and its likelihood under assumption of the hypothesis is high. Minimizing the prior probability of a confirmation corresponds to a falsifying strategy. Second, it can be shown that the measure S is mathematically *equivalent* to the traditional Bayesian measure for degree of confirmation, that is, the revision of belief in H, after observing e[3] (see Poletiek 2001) (Fig. 2 below):

$$\frac{P(H \mid e)}{P(H)}$$

Figure 2.

This equivalence is probably unacceptable for a logical falsificationist, but for modeling hypothesis-testing strategies it is quite helpful. Indeed, it implies that increasing the severity of test,

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

prior to the actual test, will result in an equal amount of increased confirmation *after e* has been observed. Third, the definition of test severity shows that the *probability* of a predicted confirmation *e* and its *severity S* are inversely related. Hence, everything being equal, as a confirming prediction is less expected a priori, it will provide a better confirmation of the hypothesis. This is in line with the information gain perspective. But, of course, the risk of not finding that confirmation increases as well. Furthermore, if we assume a test with two possible outcomes (a confirmation *e* and a falsification), the probability of the falsifying outcome increases as $P(e)$ decreases. This uncovers a Popperian paradox: maximizing the *probability* of a falsification necessarily minimizes its *power* to reject the hypothesis (Poletiek 2001).

Finally, SoT models hypothesis testing without specifying the choices people make, because these choices are assumed to depend on utilities regarding the values of test outcomes. Let us suppose a hypothesis H that generates several predictions with likelihoods 1 ($P(e|H) = 1$). SoT suggests that choosing a test involves solving the trade-off between the chance of observing a prediction and the degree of confirmation this evidence may provide for H. Depending on considerations that are external to the test, people may either prefer low degrees of confirmation and avoid the risk of getting a falsification, or go for "risky" tests because they need to be very sure about H's truth. This representation of test choice has been shown to capture the intuition of everyday hypothesis testing (Poletiek & Berndsen 2000).

As another example of this intuition, consider how SoT would account for the following adapted selection task. A participant is presented with four cards revealing an A, a K, a 2, and a 7, respectively. Every card has a number on one side and a letter on the other. She is asked to test the rule, *if there is an A on one side of a card, then there is a 2 on the other side*. In addition, she is told that the cards are sampled from a huge number of cards having either the letter A or the letter K on one side, and any number between 1 and 100 on the other. Which cards would she turn over? SoT predicts that she is interested in confirming results only ("A and 2" cards). The K and 7 cards are not selected. Under reasonable assumptions about the distribution of the letters and the numbers on the cards, it can be calculated that turning the A and finding a 2 card is a more severe test of the rule than turning the 2 and finding an A. But this difference is *felt*, as well. Indeed, finding a 2 ($P(e)$) is much less likely (given 100 different numbers) than finding the A, beforehand. Only if the rule is true would we expect a 2. Observing an A (given only two possibilities) on the 2 card, however, may just be chance. The the SoT model not only predicts that we should be more convinced by the first test outcome, we actually *feel* we should.

In summary, O&C's probabilistic model of hypothesis testing, which provides a sophisticated alternative for Popper's problematic logical falsification theory of testing, has an interesting precursor developed by Popper himself. However, this alternative falsification theory got little attention, and has been somewhat played down, both by Popper himself and by other falsificationist philosophers of science. Probably because it turned out to resemble too much its rival: confirmation theory.

NOTES

1. See, however, the relevance ratio proposed by Carnap (1950), which basically expresses the Bayesian notion of revision of belief.

2. Notice that the probabilities are applied to evidence, not to the hypothesis, as can be expected in Popper's frequentist view on probability.

3. The confirming value of a test result is expressed by Popper in his concept of "degree of corroboration" (Popper 1963/1978).

Let us not put the probabilistic cart before the uncertainty bull

doi:10.1017/S0140525X09000466

Guy Politzer[a] and Jean-François Bonnefon[b]

[a]*Institut Jean Nicod, Centre National de la Recherche Scientifique (CNRS), Ecole Normale Supérieure, 75005 Paris, France;* [b]*Cognition, Langue, Langages, Ergonomie (CLLE), Centre National de la Recherche Scientifique (CNRS), Université de Toulouse, 31058 Toulouse, Cedex 9, France.*

politzer@univ-paris8.fr

http://www.institutnicod.org/notices.php?user=Politzer top

bonnefon@univ-tlse2.fr

http://w3.univ-tlse2.fr/ltc/bonnefon

Abstract: Although we endorse the primacy of uncertainty in reasoning, we argue that a probabilistic framework cannot model the fundamental skill of proof administration. Furthermore, we are skeptical about the assumption that standard probability calculus is the appropriate formalism to represent human uncertainty. There are other models up to this task, so let us not repeat the excesses of the past.

Although research on human reasoning has long been focused on formal tasks, the past 15 years have seen a shift towards more realistic tasks, closer to everyday life. These tasks feature uncertain premises, incomplete information, defeasible conclusions, and the exploitation of individual knowledge bases. Earlier theoretical approaches to deduction were not designed to function in this new paradigm. Oaksford & Chater's (O&C's) *Bayesian Rationality* (Oaksford & Chater 2007) provides a decisive contribution to the adoption of an alternative approach to the investigation of deductive reasoning, in a radical way, by eliminating deductive logic and adopting the Bayesian view of uncertainty. Although we welcome the paradigm shift, we take issue with O&C's radical solution, on at least two fronts.

O&C's radical probabilistic approach is silent about a fundamental question: Where do conclusions come from? Given a set of premises, how do reasoners produce their own conclusion? Whatever their shortcomings, logic-based theories such as mental models and mental rules have answers to these questions. For mental models, the procedure consists of reading off the models propositions that contain new information. For mental rules, it consists of applying strategies and executing routines that yield a chain of inferences. O&C's proposal does not offer any procedure that would play this role. The probabilistic model is apt to analyze how a proffered conclusion is evaluated, but not how it is produced in the first place. In other words, a purely probabilistic model lacks the power of executing one basic human skill, the generation of proofs.

We believe with O&C that theories of human deduction cannot be purely logical, otherwise they could not account for the nonmonotonic nature of everyday inferences. But, as we just argued, no account of deduction can be purely focused on uncertainty, either, because it would miss fundamental aspects of deductive competence. We believe that the solution to this quandary is to be found in a mixed model, accommodating both uncertainty and deduction.

Our second concern with O&C's Bayesian approach revolves around the nature of human uncertainty. Human beings have the precious metacognitive ability of being aware of their own uncertainty, with respect to the conjectures they entertain. If we wish to develop models of how individuals make judgments and inferences from uncertain information, we need, first, to characterize their metacognitive representation of uncertainty. We believe this question has been largely ignored, because the answer is all too often assumed right from the start. The temptation is strong to take it for granted that standard probability calculus is the only candidate to constitute a normative psychological model of reasoning with uncertainty. Accordingly, the debate has shifted towards the question of the extent to which lay reasoners perform in agreement with the calculus.

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

But should they? Consider the requirement of completeness. Are we ready to take it for granted that the credibilities of any two assertions can always be compared? Consider now complementarity. Can we take it for granted that anytime the credibility of a conjecture increases, the credibility of its negation decreases by the exact same amount? Finally, do we take it for granted that credibilities in the mind map onto real numbers on the [0,1] interval, rather than levels on an ordinal scale?

The hurried adoption of the probability calculus as a normative model preempts the examination of these built-in assumptions. We are concerned that such a move is putting the probabilistic cart before the uncertainty bull. Important evidence is lacking before the probability calculus (Bayesian or not) can be deemed an adequate formalism for psychological modeling of reasoning with uncertainty, be it at the normative or at the descriptive level.

We do not deny that the probabilistic framework is highly successful in many disciplines, but this success does not warrant electing the probability calculus as the adequate psychological model of uncertainty, without carefully examining other options, for there are many other options, as noted by O&C. Intriguingly, this multiplicity is framed as a shortcoming in *Bayesian Rationality* (p. 73), where the uniqueness of the probability calculus is favorably compared to the multiplicity of nonprobabilistic frameworks. But surely, every uncertainty formalism is unique in its own right, and it seems abusive to throw in the same "nonprobabilistic" bag alternative candidates such as, for example, belief functions (Shafer 1976), possibility measures (Dubois & Prade 1988), relative likelihood (Halpern 1997) and plausibility measures (Friedman & Halpern 1995).

In *Bayesian Rationality*, the main argument for preferring probability theory to all other frameworks is practical: Most psychologists are only aware of probability theory as a model of uncertainty, probability theory fares reasonably well, therefore psychologists should not waste time getting acquainted with other formalisms, which are not expected to fare much better.

This argument would make perfect sense in an applied setting, but its relevance is not as clear to fundamental research on human reasoning. First, we will not know whether probability theory fares reasonably well until we compare its performances to that of other uncertainty formalisms. Second, most scientists would probably be perplexed at the thought of adopting the first theory they encounter, without considering even briefly the merits of competing theories. Third, there actually exists a community of psychologists and experimental philosophers investigating the empirical plausibility of alternative uncertainty formalisms (e.g., Benferhat et al. 2005; Ford 2004; Pfeifer & Kleiter 2005). In brief, we know far too little about human uncertainty to definitely adopt a normative model; and this matter should be solved empirically, rather than a priori.

In conclusion, we are worried that O&C may repeat within the uncertainty paradigm the double move that doomed the logical paradigm. Part of this move is to adopt an exclusive view, and to reject indispensable concepts that belong to the other paradigm; the other part is to adopt one well-known formalism (formerly the propositional calculus, and currently the probability calculus) and to establish it too hastily as a model of competence.

On is an ought: Levels of analysis and the descriptive versus normative analysis of human reasoning

doi:10.1017/S0140525X09000478

Walter Schroyens

Department of Psychology, University of Gent, B-1000 Gent, Belgium.

Walter.Schroyens@Ugent.be

http://ppw.kuleuven.be/reason/schroyens

Abstract: Algorithmic-level specifications carry part of the explanatory burden in most psychological theories. It is, thus, inappropriate to limit a comparison and evaluation of theories to the computational level. A rational analysis considers people's goal-directed and environmentally adaptive rationality; it is not normative. Adaptive rationality is by definition non-absolute; hence, neither deductive logic nor Bayesian probability theory has absolute normative status.

In *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) Oaksford & Chater (O&C) present both positive arguments in support of a probabilistic approach and negative arguments against a non-probabilistic "deductive" approach to human reasoning. The negative arguments are comparative in nature and hinge on the presumed acceptability of a direct comparison between the theories' computational-level descriptions, without taking account of their algorithmic-level specifications. I argue that this central premise is fundamentally mistaken: It is potentially misguiding to compare and evaluate theories on the basis of just their computational-level descriptions.

Cognitive psychology is about explaining what people do. Ignoring the algorithmic level is only acceptable when there is a "triumphant cascade" through the levels of description (Dennett 1987, p. 227), that is, when there is a functional match between the input-output function computed at the computational level and the function computed at the algorithmic level. In these cases, there is no added explanatory value in going from the computational level to the algorithmic level and, thus, no loss in explanatory value by not considering the algorithmic level. Cognitive theories are, however, swamped with functional mismatches between the levels (Franks 1995; 1999). It follows that when one critiques the idealized computational-level description of such theories on the basis of their explanatory (in)adequacy, one creates a straw-man argument, that is, a negative argument aimed at an oversimplified straw-man version of a theory (Schroyens, in press).

One could try to justify a comparison at the computational-level descriptions of theories by assuming that they are normative. The underlying assumption is that a theory can be normatively justified on the basis of what people actually do: "the apparent mismatch between normative theories and reasoning behaviour suggests that the wrong normative theories may have been chosen; or that the normative theories may have been misapplied" (*BR*, p. 30). O&C take the poor explanatory adequacy of deductive logic to indicate it provides the wrong norm. (Of course, when one does not consider that the normative theories may have been misapplied, one has jumped to a conclusion). There are several problems with the "is-ought" argumentation in O&C's book.

Given that cognitive psychology is not concerned with how people should reason, the "is-ought" analysis in *BR* signifies that O&C are not engaged in cognitive psychology, but are making an epistemological, philosophical argument. As an argument about normative rationality, the "is-ought" analysis is fallacious. "This fallacy consists in assuming that because something is now the practice, it ought to be the practice. Conversely, it consists in assuming that because something is not now the practice, it ought not to be the practice" (Damer 2005, p. 127). That is, one must be careful not to commit the "naturalistic fallacy" (to use G. E. Moore's term).

The "is-ought" analysis is also fallacious in its appeal to common opinion. "This fallacy consists in urging the acceptance of a position simply on the grounds that a large number of people accept it or in urging rejection of a position on the ground that very few people accept it" (Damer 2005). Virtually all people were once ignorant about the currently accepted fact that the earth is a round celestial object that revolves around the sun; and even today there is a large segment of the world population that remains unenlightened with regard to this. Does this mean that Copernicus was wrong, and was rightly excommunicated by the Church? One can have a personal opinion about how people should reason, and I think one actually should have

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

such an opinion. However, when one is considering the psychology of human reasoning, it is irrelevant whether one can label 5% or 95% of the participants as "rational" because their behavior corresponds to some norm. Whether 5% or 95% of participants' responses can be labeled "correct" is irrelevant to the task of describing and explaining what they are actually doing. The cognitive psychologist's task is first and foremost that of developing descriptively adequate processing models that allow us to understand how people reason. All performance has to be accounted for, whether logically correct or incorrect.

O&C give body to their "is-ought" ideology, by adopting Anderson's (1990) rational analysis approach. However, it is a mistake to think that the "the essence of rational analysis [Anderson 1990] is to view cognitive processes as approximating some normatively justified standard of correct performance" (Oaksford & Chater 1998, p.18). As O&C stated more correctly elsewhere: "Anderson [1990] argues that we must distinguish normative from adaptive rationality" (Oaksford & Chater 1998, p. 174; see also, Chater et al. 2003). Indeed, "rationality is being used in two senses and that rationality in the adaptive sense, which is used here, is not rationality in the normative sense that is used in the studies of decision making and social judgement" (Anderson 1990, p. 31). There is no unconditional "should": there are no absolute normative standards of correct performance. This goes against the very notion of adaptive rationality, which is relative and "defined with respect to a specific environment" (Anderson 1990, p. 35).

O&C have fallen into the same conceptual trap many "mental logicians" fell into: a conceptual mistake – called "reasoning imperialism" by Rips (2002) – for which O&C, with much merit to their work, criticized such logicians. *Bayesian Rationality* provides prominent and valuable contributions to the literature that echoes Anderson's (1990) idea that deductively irrational behavior might not be adaptively irrational. That is, the book successfully questions the normative status of logic within the rational analysis approach of Anderson (1990). However, instead of adopting and following the notion of adaptive rationality to its full extent, by questioning the feasibility of a single theory (whether probabilistic or logical) to have absolute normative status, they simply replace one normative theory with another. Adaptation is defined with respect to a specific environment and a true rational analysis accordingly takes account of this environment. It does not concern itself with absolute normative issues, because this denies the environmental relativity of adaptively rational behavior.

"Nonmonotonic" does not mean "probabilistic"

doi:10.1017/S0140525X0900048X

Keith Stenning[a] and Michiel van Lambalgen[b]

[a]*Human Communication Research Center, University of Edinburgh, Edinburgh EH8 9LW, United Kingdom;* [b]*Institute for Logic, Language and Computation, Department of Philosophy, University of Amsterdam, Amsterdam, 1012CP, The Netherlands.*

keith@inf.ed.ac.uk

www.inf.ed.ac.uk/people/staff/Keith_Stenning

M.vanLambalgen@uva.nl

http://staff.science.uva.nl/ ~ michiell

Abstract: Oaksford & Chater (O&C) advocate Bayesian probability as a way to deal formally with the pervasive nonmonotonicity of common sense reasoning. We show that some forms of nonmonotonicity cannot be treated by Bayesian methods.

One argument that Oaksford & Chater (O&C) proffer in *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*) for a probabilistic approach to cognition is the pervasiveness of nonmonotonicity in reasoning: almost any conclusion can be overturned, if additional information is acquired. They claim that nonmonotonic inferences fall outside the scope of logical methods, and that probability theory must be preferred.

This judgment of the authors does not reflect the present state of play (see, e.g., Antoniou 1997, especially Part V). There are several good logical formalisms for nonmonotonic reasoning in existence, some of which even have a computational complexity that is vastly less than that of classical logic. More importantly, even in those cases where probabilistic modeling is in principle useful, one sometimes needs to supplement the model with a nonmonotonic logic to account for all the data. We illustrate this here by using an example from Chapter 5 of *BR*, not mentioned in the present *Précis*, but relevant throughout: the suppression task (Byrne 1989). When subjects are presented with the *modus ponens* material "If she has an essay she studies late in the library. She has an essay," they almost universally draw the conclusion: "She studies late in the library." When instead they are presented with the same two premises *plus* the premise: "If the library is open she studies late in the library," about half of them withdraw the inference. Logically speaking, we here have an instance of nonmonotonicity: addition of a premise leads to withdrawal of the original conclusion.

O&C argue that their probabilistic model – which represents the conditional "if p, then q" as a conditional probability $P(q|p)$ – can account for suppression effects. To explain the model and our criticism, we must first state the main Bayesian mechanism for updating probabilities, "Bayesian conditionalization": (BaCo). The absolute subjective probability of event D given that evidence E has been observed is equal to the conditional probability $P(D|E)$. Here E should encompass available evidence. In this context it is important to note that Bayesian conditionalization is a nonmonotonic principle: An extension of the evidence p may invalidate a previous posterior probability for q derived by (BaCo) applied to $P(q|p)$.

Informally, the O&C model works likes this: One is given the conditional probability $P(q|p)$ with value, say, $1 - e$ for some positive but small e. If few exceptions are salient, e is small, and, given p, (BaCo) yields that q can be concluded with high probability (namely, $1 - e$). The second conditional highlights a possible exception (the library being closed), which leads to an increase in e, and hence to a decrease in the a posteriori probability of q. But appealing though this picture is, it is *not* Bayesian (Stenning & van Lambalgen 2008b). Consider two putative Bayesian processes that can change the value of $P(q|p)$ when new possible exceptions – say, *not-r* with probability $1 - P(r)$ – become salient.

(1) The probability space is enlarged from $\{p, q\}$ to $\{p, q, r\}$ (where r stands for, say, "the library is open"), and this leads to a new representation of $P(q|p)$. One may write

$$P(q \mid p) = \frac{P(p \wedge q)}{P(p)} = \frac{P(p \wedge q \wedge r)}{P(p)} + \frac{P(p \wedge q \wedge \neg r)}{P(p)} = \\ = \frac{P(p \wedge q \wedge r)}{P(p)} = \frac{P(p \wedge q \wedge r)}{P(p \wedge r)} \frac{P(p \wedge r)}{P(p)} = P(q \mid p \wedge r) P(r)$$

Figure 1.

where the last equality follows under the plausible assumption that p and r are independent. O&C assume that the subject assigns a lower probability to $P(q|p)$ in the enlarged representation $P(q|p \ \& \ r)P(r)$, suggesting that this is because the subject lowers $P(r)$ from 1 to a value smaller than 1 when becoming aware of possible exceptions. In probability theory, $P(q|p \ \& \ r)$ $P(r)$ would simply be a different representation of $P(q|p)$, and the values of these expressions must be the same. There are no rationality principles governing changes in probabilities when enlarging the probability space, or rather there is one such

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

principle, that $P(q|p)$ remains the same when computed on an enlarged space. This is the only way in which one can guarantee that enlargements of the probability space in the limit lead to a coherent probability distribution – the starting point of Bayesian rationality.

(2) An orthodox Bayesian alternative would be a construction in which the probability spaces remain the same (namely, the universal space based on all possible propositions), but the probability distributions change. In our toy world, the probability space is in both cases {*p, q, r*}, but one could assume that the probability distribution first assigns probability *0* to *not-r*, and, upon becoming aware of the second conditional "if *r* then *q*," a nonzero probability. The trouble with such a suggestion is that from a Bayesian point of view, the transition from the a priori probability $P(not\text{-}r)=0$ to the a posteriori $P(not\text{-}r) > 0$ is not allowed, because this cannot be achieved via (BaCo): conditionalizing on more evidence cannot make a null probability positive. One thus needs an additional rationality principle (beyond [BaCo]) governing such transitions. In the absence of such a principle, one has to assume that the probabilities of all nonsalient exceptions (such as *not-r*) are initially very small but nonzero. This increases the computational complexity of probabilistic reasoning enormously: One requires massive storage and intricate computations to maintain consistency of the probability assignment.

These considerations show that in order to account for the data on the suppression task any probabilistic model needs to be supplemented with a theory about nonmonotonic and non-Bayesian, but still somehow rational, changes in degrees of belief. One may then question whether a probabilistic model is necessary at all; Stenning and van Lambalgen (2005; 2008a) provide a model cast entirely in terms of nonmonotonic logic.

The dynamics of development: Challenges for Bayesian rationality

doi:10.1017/S0140525X09000491

Nils Straubinger, Edward T. Cokely, and Jeffrey R. Stevens

Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, 14195, Germany.

straubinger@mpib-berlin.mpg.de
cokely@mpib-berlin.mpg.de
jstevens@mpib-berlin.mpg.de
http://www-abc.mpib-berlin.mpg.de/users/jstevens/

Abstract: Oaksford & Chater (O&C) focus on patterns of typical adult reasoning from a probabilistic perspective. We discuss implications of extending the probabilistic approach to lifespan development, considering the role of working memory, strategy use, and expertise. Explaining variations in human reasoning poses a challenge to Bayesian rational analysis, as it requires integrating knowledge about cognitive processes.

Bayesian rationality highlights the remarkable successes rather than failures of human reasoning by recasting seemingly erroneous reasoning in logical tasks using a probabilistic approach. However, in their book *Bayesian Rationality* (Oaksford & Chater 2007, henceforth BR), Oaksford & Chater (O&C) draw a rather static picture of human reasoning by focusing on typical patterns of responses from adults. We propose a more dynamic perspective, which considers that reasoning systematically varies within individuals over the lifespan (Howe & Rabinowitz 1996; Markovits & Barrouillet 2002) and between individuals with different levels of knowledge and expertise (Ericsson et al. 2006). Although O&C acknowledge the importance of considering reasoning data on individual differences (*BR*, p. 288) and on information processing capacities (p. 290), they do not adequately account for how variation influences a Bayesian rational analysis of reasoning. Anderson (1991a) and others have pointed out that perhaps the major potential limitation, the "Achilles heel," of rational analysis would be computational constraints that are too complex or arbitrary. We argue that our understanding of the mechanisms of change in reasoning can help us specify computational limitations for probabilistic modeling and assess whether a single model can capture the complexities of reasoning.

Many important aspects of cognition change over the lifespan, and reasoning is no exception (Baltes et al. 1999). According to Piaget, both logical reasoning and probabilistic reasoning emerge from adolescence to young adulthood at the highest stage of cognitive development (Piaget & Inhelder 1975). Subsequent research, however, has qualified these findings, showing that younger children understand aspects of such reasoning (Falk & Wilkening 1998; Galotti et al. 1997). Furthermore, reasoning continues to develop during adulthood with performance in specific domains increasing as individuals gain reasoning knowledge and expertise (Ericsson & Lehmann 1996; Sternberg 1999). Yet, overall across the adult lifespan, abstract reasoning (measured by intelligence tests) declines with age (Verhaeghen & Salthouse 1997). Thus, reasoning is a dynamic aspect of cognition that varies with age and experience and results from the interplay of biological processes (e.g., brain maturation) and enculturation (e.g., education) (Baltes et al. 1999).

A developmental perspective may inform Bayesian rational analysis by specifying computational limitations of the cognitive system. An important limitation faced by the human cognitive system is working memory capacity – a key determinant of reasoning performance (Kyllonen & Christal 1990). Like other cognitive capacities, working memory systematically changes across the lifespan by steadily increasing during childhood (Conlin et al. 2005) and declining across adulthood (Verhaeghen & Salthouse 1997). Working memory, therefore, poses a dynamic constraint on the rational analysis of reasoning.

Although O&C are currently silent on the role of developmental changes in working memory and reasoning, they do note that individuals with higher working memory capacities tend to exhibit more logical reasoning. To illustrate, in the Wason selection task, a subgroup of individuals (ca. 10%) consistently chooses the logically correct combination of cards, indicating that although most seem to adopt a probabilistic model, others clearly do not. O&C suggest that this variation in behavior primarily reflects deliberative strategy use and educational (training) differences, which are "not indicative of individual differences in the nature of the fundamental principles of human reasoning" (*BR*, p. 288). This claim seems problematic given what we know about the interplay between strategy use, training, and basic cognitive mechanisms. Of course, cognitive capacities can constrain the strategies that people use; however, specific strategy use and training may shape the basic cognitive mechanisms, as well. Differences in memory strategies (e.g., rehearsal, chunking) can also alter basic mechanisms of working memory capacity and its relationship to cognitive performance (Cokely et al. 2006). In addition, both extensive practice with specific strategies and the acquisition of knowledge and expertise dramatically expand working memory (Ericsson & Kintsch 1995). Indeed, as training changes deliberative strategies to automatic processes, the cortex can undergo functional neuroanatomical reorganization (Dick et al. 2006). Thus, it is possible that deliberative strategy use and training may influence reasoning precisely *because* they alter underlying cognitive mechanisms such as working memory. Given the complex relationship between strategies, training, and cognitive mechanisms, it seems premature to dismiss individual differences in strategy use as not fundamental to reasoning. A comprehensive model of human reasoning must account for these differences.

Variation in human reasoning has proven difficult to capture for probabilistic models (Shultz 2007), although recent research

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

has made some progress applying probabilistic models to individual differences (e.g., category learning; Navarro et al. 2006) and cognitive development (e.g., causal reasoning; Sobel et al. 2004). This work represents a step in the right direction; however, we expect that no single model can predict reasoning performance equally well across age groups and levels of experience. Indeed, systematic variations in peoples' behavior suggest that several different models (or modifications of a given model) may be required to explain developing behavior (Shultz 2007). Nevertheless, investigating differences between the models across age groups and skill levels may help us to understand more exactly "what differs" between and "what develops" within individuals.

In closing, we must emphasize O&C's comment that probabilistic models are often only functional level theories that should not be confused with algorithmic level theories (process models). Brighton and Gigerenzer (2008) have pointed out in their discussion of the limits of Bayesian models of cognition that the question of why the human mind does what it does (functional level) cannot be separated from the question of how the human mind does it (algorithmic level). Therefore, it is crucial that future Bayesian rational analyses specify how exactly their functional level models constrain theorizing about cognitive processes. This issue is especially relevant as the data connecting development, expertise, working memory, and reasoning imply that multiple strategies (and therefore processes) are at play. Though Bayesian rationality seems to provide a functional level account of prototypical adult reasoning, the development of cognitive capacities and expertise remains underappreciated.

ACKNOWLEDGMENTS

Preparation of this commentary was supported in part by a stipend from the International Max Planck Research School LIFE to NS. We are grateful to Henrik Olsson for comments on a previous draft.

How do individuals reason in the Wason card selection task?

doi:10.1017/S0140525X09000508

Eric-Jan Wagenmakers

Department of Psychological Methods, University of Amsterdam, Amsterdam 1018 WB, The Netherlands.

EJ.Wagenmakers@gmail.com

http://users.fmg.uva.nl/ewagenmakers/

Abstract: The probabilistic approach to human reasoning is exemplified by the information gain model for the Wason card selection task. Although the model is elegant and original, several key aspects of the model warrant further discussion, particularly those concerning the scope of the task and the choice process of individuals.

In the book *Bayesian Rationality* (Oaksford & Chater 2007, henceforth *BR*), Oaksford & Chater (O&C) present a summary and a synthesis of their work on human reasoning. The authors argue that formal logic and deduction do not explain how people reason in everyday situations. The deficiencies of the most simple forms of logic are obvious when one considers that they may assign "true" to absurd statements such as "if the moon is blue, than cows eat fish" (*BR*, p. 70). More importantly, the authors propose that, in contrast to formal logic, probability calculus does provide the right tools for an analysis of human reasoning. Thus, the authors argue that people solve deductive tasks by inductive methods. From this perspective, human reasoning can be characterized as Bayesian or rational.

Consider the Wason card selection task discussed in Chapter 6. Participants are confronted with four cards, showing an A, a K, a 2, and a 7. Participants are told that each card has a number on one side and a letter on the other. They are given a rule, "if there is an A on one side, then there is a 2 on the other side," and subsequently, have to select those cards that need to be turned over to assess whether the rule holds true or not. A moment's thought reveals that the cards that need to be turned over are the A card and the 7 card. Yet, the majority of participants do not choose the 7 card, but tend to choose the 2 card instead.

O&C propose an elegant Bayesian model – the information gain model – to account for people's performance in the Wason task. According to the model, people select the cards that reduce their expected uncertainty the most. Specific assumptions about the rarity of the information on the cards lead to the conclusion that selection of the 2 card might be rational after all.

The information gain model has been subjected to intense scrutiny (e.g., Oberauer et al. 1999). For non-experts, the details of this discussion are somewhat difficult to follow. A useful guideline is that a model should only be abandoned when it can be replaced with something better. And – criticisms raised against the information gain model notwithstanding – I have not come across a model that does a better job explaining how people make their card selections.

Despite its simplicity and elegance, some important details of the information gain model were not clear to me. First, O&C argue, on page 210, that their account only holds if participants regard the cards as a sample from a larger population. Perhaps the authors could spell out this argument in a bit more detail. Taking probability as a reflection of degree of belief, I did not immediately see what calculations are in need of adjustment. Second, the authors mention that participants who realize that the cards are *not* sampled from a larger population would always choose the A card and the 7 card. I do not know whether this prediction has been tested empirically, but I find it only slightly more plausible than cows eating fish. Note that in the Wason task a substantial proportion of participants do not even select the A card.

Another issue that warrants closer examination is the way the model's predictions relate to the data. In the information gain model, each card reduces the expected uncertainty to some extent. Why then does an individual participant not select all four cards, but generally only selects one or two? In other words, it was unclear to me how the model, from a consideration of expected uncertainty reduction, can predict card selections for an individual participant.

A fourth point concerns the role of individual differences. As the authors discuss on page 211, a subgroup of undergraduate students with high intelligence (about 10%) do select the A card and the 7 card. This result strengthened my initial belief that a motivated, intelligent person would always choose the A and 7 cards, when given sufficient time. In the spirit of falsification, I then tested this assumption on a colleague, who of course immediately selected the A and 2 cards. Perhaps she was not sufficiently motivated to think the problem through carefully – would incentives of time or money increase the selection of the 7 card?

O&C are to be admired for their principled approach to quantitative modeling, and for their courage to take on the unassailable dogma of human irrationality. It is unfortunate that much of the material in the book was already available elsewhere (e.g., Oaksford & Chater 2001; 2003b); therefore, it was not entirely clear to me what the book adds to our current knowledge base.

One final comment. It strikes me as paradoxical that researchers who argue for a coherent, rational approach to human reasoning then proceed to apply an incoherent, irrational approach to the statistical analysis of their experimental data. Throughout the book, the authors renounce Popper's stance on the importance of falsification, arguing that this is not how science works, nor how people reason. But then, in the very same work, the authors measure the validity of their models by means of p-values, and include statements such as "the model could not be rejected." Why?

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Authors' Response

The uncertain reasoner: Bayes, logic, and rationality

doi:10.1017/S0140525X0900051X

Mike Oaksford[a] and Nick Chater[b]

[a]*School of Psychology, Birkbeck College London, London, WC1E 7HX, United Kingdom;* [b]*Division of Psychology and Language Sciences & ESRC Centre for Economic Learning and Social Evolution, University College London, London, WC1E 6BT, United Kingdom.*

mike.oaksford@bbk.ac.uk
n.chater@ucl.ac.uk
www.bbk.ac.uk/psyc/staff/academic/moaksford
www.psychol.ucl.ac.uk/people/profiles/chater_nick

Abstract: Human cognition requires coping with a complex and uncertain world. This suggests that dealing with uncertainty may be the central challenge for human reasoning. In *Bayesian Rationality* we argue that probability theory, the calculus of uncertainty, is the right framework in which to understand everyday reasoning. We also argue that probability theory explains behavior, even on experimental tasks that have been designed to probe people's logical reasoning abilities. Most commentators agree on the centrality of uncertainty; some suggest that there is a residual role for logic in understanding reasoning; and others put forward alternative formalisms for uncertain reasoning, or raise specific technical, methodological, or empirical challenges. In responding to these points, we aim to clarify the scope and limits of probability and logic in cognitive science; explore the meaning of the "rational" explanation of cognition; and re-evaluate the empirical case for Bayesian rationality.

R1. Introduction

Bayesian Rationality (Oaksford & Chater 2007, henceforth *BR*) proposed that human reasoning should be understood in probabilistic, not logical, terms. In Part I, we discussed arguments from the philosophy of science, artificial intelligence, and cognitive psychology, which indicate that the vast majority of cognitive problems (outside mathematics) involve uncertain, rather than deductively certain, reasoning. Moreover, we argued that probability theory (the calculus for uncertain reasoning) is a more plausible framework than logic (the calculus for certain reasoning) for modeling both cognition in general, and commonsense reasoning in particular. In Part II, we considered a strong test of this approach, asking whether the probabilistic framework can capture human reasoning performance even on paradigmatically "logical" tasks, such as syllogistic reasoning or conditional inference.

The structure of this response is as follows. In section R2, we reflect on the ubiquity of uncertainty and address the theoretical attempts to preserve logic as a separate and core reasoning process. In section R3, we compare and evaluate Bayesian and logic-based approaches to human reasoning about uncertainty. Section R4 focuses on the methodology of rational analysis (Anderson 1990; 1991a; Oaksford & Chater 1998b) and its relationship to more traditional algorithmic and neuroscientific approaches. Section R5 discusses a variety of specific issues in the empirical data from the psychology of reasoning, and the modeling of that data. Finally, section R6 concludes the case for a "Bayesian turn" in the brain and cognitive sciences in general, and for the understanding of human reasoning in particular.

R2. The ubiquity of uncertainty: Distinctions that might preserve logic

Many commentators suggest ways to preserve a role for logic as a separate and core component in an account of human reasoning, despite the challenge provided by uncertainty (**Allott & Uchida**, **Evans**, **Politzer & Bonnefon**). We argue that logic does have an important role in modeling cognition; but we argue against the existence of cognitive processes dedicated to logical reasoning.

R2.1. Rationality 1 versus Rationality 2

Evans suggests that a distinction should be drawn between two types of rationality (Evans & Over 1996a). Rationality 1 relates to implicit, possibly associative, processes, operating over world knowledge, which Evans also terms "ecological rationality." This type of rationality arises from System 1 in Evans and Over's (2004) Dual Process Theory (see also Evans & Frankish, in press; Sloman 1996; Stanovich & West 2000). Rationality 2 involves explicitly following normative rules, and is the type of rationality achieved by Evans and Over's (2004) System 2. System 2 processes are logical, rule-governed, and conscious. Moreover, Evans has argued for a crucial asymmetry between the systems. It requires cognitive effort to ignore System 1, and to use System 2 for logical inference: that is, to infer only what follows from the structure of the given premises.

The fundamental problem with this Dual Process view is that these two systems must interact – and if the systems obey fundamentally different principles, it is not clear how this is possible. Consider the familiar example of inferring that *Tweety flies* from the general claim that *birds fly* and the fact that *Tweety is a bird*. On the Dual Process view, this inference could be drawn logically from the premises given by System 2, from the assumption that *birds fly* is a true universal generalization; System 1, by contrast, might tentatively draw this conclusion by defeasible, associative processes, drawing on general knowledge. But a lack of synchrony between the two systems, presumed to operate by different rational standards, threatens to cause inferential chaos. Consider, for example, what happens if we consider the possibility that *Tweety is an ostrich*. If System 2 works according to logical principles, the clash of two rules threatens contradiction: we know that birds fly, but that ostriches do not. To escape contradiction, one of the premises must be rejected: most naturally, *birds fly* will be rejected as false. But we now have two unpalatable possibilities. On the one hand, suppose that this retraction is *not* transferred to general knowledge and hence is not assimilated by System 1. Then the two systems will have contradictory beliefs (moreover, if System 2 reasoning cannot modify general knowledge, its purpose seems unclear). On the

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

other hand, if *birds fly* is retracted from world knowledge, along with other defeasible generalizations, then almost all of general knowledge will be stripped away (as *BR* notes, generalizations outside mathematics are typically defeasible), leading System 1 into inferential paralysis.

The centrality of logic for a putative System 2 is also brought into doubt by considering that one of its main functions is to consciously propose and evaluate *arguments*. Yet, argumentation, that is, the attempt to persuade oneself or others of a controversial proposition, is uniformly agreed not to be a matter of formal logic (Walton 1989), although aspects of argumentation may naturally be modeled using probability theory (Hahn & Oaksford 2007). Thus, perhaps the core human activity for which a logic-based System 2 is invoked may, ironically, be better explained in probabilistic terms.

People can, of course, be trained to ignore some aspects of linguistic input, and concentrate on others – for example, in the extreme, they can learn to translate natural language statements into predicate logic (ignoring further aspects of their content) and employ logical methods to determine what follows. But, for the psychology of reasoning, this observation is no more significant than the fact that people can learn the rules of chess and ignore most of the visual features of the pieces, the board, or indeed, their surroundings. Conscious application of logical principles is a learned skill built on top of non-logical machinery (and, indeed, is highly effortful, even for logicians); it does not involve, we suggest, tapping in to some underlying logical "engine."

It is this conscious application of logical concepts (and related notions from mathematics, philosophy, and computer science) that underpins, we suggest, the small but significant correlation between "logical" performance on some reasoning tasks (e.g., selecting the *p* and *not-q* cards, in Wason's selection task) and IQ (Stanovich & West 2000). Logical reasoning is a late and cognitively challenging cultural innovation, rather than a core component of our mental machinery.

Evans also expresses disappointment that we do not address individual differences (Stanovich 2008), which have been viewed as supporting a Dual Process account. But from the present perspective, individual differences concerning the application of learned logical rules are no different from individual differences in chess playing – that is, neither are directly relevant to the question of whether there are single or multiple reasoning systems. Indeed, we suggest that individual differences provide no stronger evidence that cognition involves core *logical* competence, than that cognition involves core *chess-playing* competence.

It may turn out, indeed, that there is no real incompatibility between Stanovich's account and ours. In particular, the distinction Stanovich draws between control processes and other autonomous systems is a distinction common to all theories of cognition (see Oaksford & Chater, in press). But as Kowalski's (1979) classic equation, "Algorithm = Logic + Control," reminds us, logic and control processes are very different (see, e.g., Anderson 1983). Hence, Stanovich may not really be committed to anything like **Evans**' logically competent System 2. (A further complication is that a distinction between processes of logic and control is now reflected in Evans [2007], who moots the possibility of a tri-process theory.)

R2.2. The split between semantics and pragmatics

Grice's (1975) theory of conversational implicature originally attempted to split off a "stripped down" logic-based natural language semantics, from the complex, knowledge-rich processes of pragmatic interpretation involved in inferring a speaker's intentions. In this way, he aimed to retain a logical core to semantics, despite apparently striking and ubiquitous clashes between the dictates of formal logic and people's intuitions about meaning and inference.

Within this type of framework, **Allott & Uchida** attempt to preserve the truth of potentially defeasible conditionals (*if it's a bird, then it flies*, or, as above, *birds fly*) despite the ready availability of counterexamples. They suggest that this conditional is true in one model, but not in the model that is considered when an additional premise giving a counterexample is added (e.g., when we consider the possibility that Tweety is an ostrich). But in classical logic, only an inference that holds in *all* models is deductively valid, by definition. Thus, accepting that this inference holds only in some models implies accepting that the inference is uncertain (contra, e.g., **O'Brien**). Indeed, in *BR*, we argue uncertainty is ubiquitous in human reasoning; outside mathematics, deductive reasoning, which *guarantees* the truth of a conclusion given the premises, is, to a first approximation, never observed.

Moreover, understanding reasoning involves working out pragmatic details about what default background assumptions are applicable in reasoning. Thus, for example, our accounts of specific reasoning phenomena, across conditional reasoning, the selection task, and syllogistic reasoning, involve default assumptions about the environment, for example, what is rare and what is common (cf. **McKenzie**; McKenzie et al. 2001) and when states are likely to be independent or conditionally independent. In this light, we agree with **Stenning & van Lambalgen**'s claim that "pure" Bayesian analysis, working from the premises alone, cannot capture suppression effects in conditional reasoning (see sect. R3.6) – we view this as illustrating the knowledge-rich character of reasoning, rather than challenging a Bayesian account.

The ubiquity of uncertain, knowledge-rich inference, argues for an alternative to invoking the semantics/pragmatics distinction to maintain a logical semantics for natural language: namely, that natural language semantics may be probabilistic "all the way down." Experiments in the psychology of reasoning, as reviewed in *BR*, find little support for the existence of a level of logic-based representation or inference. *BR* proposes a starting point for a probabilistic semantics: *If p then q* conditionals are assumed to express that the conditional probability $P(q|p)$ is high (following Adams 1975; 1998; Bennett 2003; and Edgington 1995, among others); the quantifiers *Some*, *Few*, *Most*, *All* are similarly assumed to express constraints on probabilities (e.g., *Some A are B* is rendered as $P(A, B) > 0$; *Most A are B* claims that $P(B|A)$ is high). Switching from a logical to a probabilistic semantics provides, we argue, a better fit with patterns of human reasoning. Of course, it remains possible that a logical core

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

interpretation might be maintained – but it seems theoretically unparsimonious to do so (Edgington 1995).

A shift from a logical to a probabilistic semantics for aspects of natural language may also allow a more integrated account of semantics and pragmatics. Indeed, **McKenzie** (e.g., Sher & McKenzie 2006) has powerfully demonstrated the importance of pragmatic factors, even within a purely probabilistic framework (but see, Hilton et al. 2005). Nonetheless, the core insight of Grice's program remains: that splitting apart semantic factors (concerning meaning) and pragmatic factors (concerning inferences about speaker intentions) is a prerequisite for constructing a tractable semantic theory, whether that theory be based on logic (as **Allott & Uchida** argue) or probability (as *BR* proposes).

R2.3. Proof and uncertainty and structure and strength

Politzer & Bonnefon argue that a key element missing from a purely probabilistic account is how premises can be used to construct proofs to derive conclusions. Thus, they argue that the probabilistic account allows the *evaluation* of the strength of the relationship between premises and conclusion, but not how the conclusion is *generated* in the first place. Note, though, that both logic and probability are theories of the nature of inferential relationships between propositions (Harman 1986). Neither specify how reasoning should be carried out, let alone how interesting conclusions should be generated. Moreover, for both logic and probability, a range of algorithms have been developed which can both evaluate given conclusions, and generate new conclusions (e.g., logic programming and Bayesian networks). From both perspectives, any set of information potentially generates an infinite set of possible conclusions; so that an immediate question is: What counts as an *interesting* conclusion? A natural suggestion from the probabilistic point of view is that conclusions with a low prior probability are, other things being equal, more surprising and hence more interesting (as employed in the account of syllogistic reasoning described in *BR*), although interesting logic-based measures of semantic information content have also been proposed (Johnson-Laird 1983).

More generally, the probabilistic approach is just as able as logic-based approaches to serve as the basis for algorithmic models of thought. For example, Oaksford & Chater (in press) use a constraint satisfaction neural network implementation of the probabilistic approach. The links in the network captures the conditional and default assumptions about *structural* relations between variables (in the causal context, involving alternative causes and defeaters); and the *strength* of each link is captured by a weight. A similar distinction between structure and strength has been invoked in causal reasoning using Bayesian networks (Griffiths & Tenenbaum 2005) and applied in Hahn and Oaksford's (2007) probabilistic account of argumentation.

R3. Logic, probability, and the challenge of uncertain reasoning?

In this section, we consider whether, as some commentators suggest, we have mischaracterized the scope of logic or chosen the wrong alternative calculus in order to reason about uncertainty. We deal with logic and probability in turn.

R3.1. How are logic and probability related?

Pfeifer & Kleiter observe that probability theory already includes classical propositional logic as a special case. Thus, one way of understanding the approach outlined in *BR* is as enriching conventional logic to give an *inductive* logic – a system of logic that extends deduction to less-than-certain inferences (Hawthorn 2008). To a good approximation, modern inductive logic just *is* Bayesian probability (Chater et al., in press; Earman 1992), with some additional discussion of the measure of the confirmation relation (see later discussion of **Poletiek** and **Nelson**). Since Carnap (1950), this Bayesian inductive logic includes classical logic – if a statement has a probability of 1, then any logical consequence of that statement also has a probability of 1. Similarly, if a statement has an implication with a probability of 0, then that statement has a probability of 0 (note, however, that probability theory does not readily represent the *internal* structure of atomic propositions, and has no general theory of, for example, quantification or modality). The Bayesian inductive perspective is required not because classic logic is incorrect, but because, outside mathematics, it rarely, if ever, applies (Oaksford & Hahn 2007) – inferential relations between propositions are relentlessly uncertain (Jeffrey 1967).

R3.2. Is relevance relevant?

O'Brien proposes a different enrichment of logic, drawing on his important work with Braine on mental logics (Braine & O'Brien 1991), which aims to capture a notion of *relevance* between antecedent and consequent (i.e., so that conditionals such as *if 2 is odd, then the sky is purple* are no longer automatically true, just in virtue of the false antecedent). Thus, Braine and O'Brien's work aims to go beyond the material conditional, which *BR* ascribed to mental logic as a whole (e.g., Rips 1994).

Adding a condition of relevance, while potentially important, does not help deal with the problem of uncertain reasoning, however. Indeed, **O'Brien**'s account of conditionals is, instead, a strictly deductive version of the Ramsey test (like, e.g., Gärdenfors 1986) – conditionals are only asserted if the consequent, q, follows with certainty from the antecedent p (and background knowledge). Thus, Braine and O'Brien's (1991) logical interpretation of the conditional suffers the same fundamental problem as material implication: an inability to capture the fact that generalizations outside mathematics are inevitably uncertain.[1]

Moreover, despite Braine and O'Brien's intentions, their system does not seem to enforce relevance between antecedent and consequent, either. The introduction rule for *if* p *then* q, used by **O'Brien**, and described in Braine and O'Brien (1991), states that *if* p *then* q can be inferred if q follows from the supposition of p together with background knowledge, B. If we know p is false (i.e., background knowledge B implies *not-*p), then supposing p and B implies p & *not-*p, which is a contradiction, from which any conclusion follows – including q. So conditionals

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

such as *if 2 is odd, then the sky is purple* can be asserted, after all. Similarly, any conditional whose conclusion is known to be true (i.e., *B* implies *q*) will automatically meet the condition that $p \& B$ implies *q* (because this is a monotonic logic – adding premises can never remove conclusions). Hence, conditionals such as *if the sky is purple, then 2 is even*, will also be asserted – again violating intuitions of relevance.

R3.3. Uncertain reasoning via nonmonotic logic?

Stenning & van Lambalgen argue that we misrepresent the scope of current logical methods, noting that a range of nonmonotonic logics, in which adding a premise may require withdrawing a previously held conclusion, might meet the challenge of uncertainty. As noted in *BR*, and elsewhere (e.g., Oaksford & Chater 1991; 2002), there are, however, fundamental problems for nonmonotonic logics in the crucial case where different "lines of argument" clash. Thus, *if it is sunny, John goes to the park*, and *it's sunny* appears to provide a powerful argument that *John goes to the park*. But adding the premise, *John is arrested by the police in a dawn raid*, together with background knowledge, appears to yield the conclusion that *John does not go to the park*.

From the perspective of classical logic, this situation is one of contradiction – and what is needed is a way of resolving which premise should be rejected. For example, one might claim that the conditional *if it's sunny, John goes to the park* is false, precisely because of the possibility of, among other things, arrest. But, as noted in section R2.1, it is then difficult to avoid the conclusion that all conditionals, outside mathematics, are false, because the possibility of counterexamples always exists. Reasoning from premises known to be false is not, of course, justified, whether in logic, or any other standard framework, and hence, the logical analysis of the original argument collapses.

The strategy of nonmonotonic logic attempts to solve this problem by treating the conditional as a *default* rule, which holds, other things being equal. Indeed, outside mathematics, almost all rules are default rules. Indeed, the implicit rule that allows us to infer that being arrested is incompatible with a trip to the park is itself a default rule, of course – for example, arrest may be extremely brief, or perhaps the police station is itself in the park. Thus, from this viewpoint, uncertain reasoning centrally involves resolving clashes *between* default rules. In *BR*, we argue that resolving such clashes is not typically possible by looking only at the structural features of arguments. Instead, it is crucial to differentiate stronger and weaker arguments, and degrees of confidence in the premises of those arguments. Logical methods provide no natural methods for expressing such matters of degree; but dealing with degrees of belief and strength of evidence is the primary business of probability theory.

R3.4. Is logic relevant to cognition?

Several commentators suggest that the powerful machinery of logic should not be jettisoned prematurely (**Allott & Uchida**, **De Neys**, **O'Brien**, **Politzer & Bonnefon**, **Stenning & van Lambalgen**). As we noted in section R3.1, probability theory (i.e., modern inductive logic) is a generalization of logic, allowing degrees of uncertainty. However, it is a generalization that is presently limited in scope. This is because how probability interacts with richer representations involving, for example, relations, quantification, possibility, deontic claims, tense and aspect, and so on, is yet to be worked out. *BR* has, as we have mentioned, some preliminary suggestions about the probabilistic representation of individual connectives (*if...then...*) and quantifiers (*Most*, *Few*, *Some*, etc.). But this is very far from a full probabilistic generalization of, for example, the predicate calculus, the workhorse of classical logic and natural language semantics. The formal challenges here are substantial. Nonetheless, much progress has been made, in a number of directions, in fusing together probabilistic and logical methods (e.g., see papers in Williamson & Gabbay 2003), thus advancing Carnap's (1950) program of building an inductive logic. **Pfeifer & Kleiter** apply logic in an interesting, but distinct, way: as providing a machinery for reasoning *about* probability, rather than using probability to generalize logic.

According to **De Neys**, concentrating on the computational level means that *BR* underplays the role of logic in human reasoning. De Neys argues that latency and brain imaging studies, investigating the mental processing involved in reasoning, rather than just the output of these processes, consistently reveal a role for logic. Yet all the cases that De Neys cites involve a conflict between belief and logic such that prior belief suggests one response, but logical reasoning from the given premises suggests another. However, the Bayesian approach can explain at the computational level why such conflicts might arise and therefore why inhibitory processes might need to be invoked (De Neys et al. 2008; Houdé et al. 2000). Oaksford and Hahn (2007) point out that probabilistic validity of an argument and its inductive strength can conflict. So, for example, Modus Ponens (MP) is probabilistically valid. However, if the probability of the conditional is low, then the inductive strength of the argument, that is, the probability of the conclusion given the premises, will also be low. The right computational level analysis may, therefore, remove the need to propose two special purpose cognitive systems operating according to different principles. This view seems consistent with the current state of imaging studies, which provide little evidence for a dedicated logical reasoning module (Goel 2007).

O'Brien describes Chrysippus' dog's ability to follow a scent down one path in a fork in the road, having eliminated the other as an application of the logical law of disjunction elimination – and hence, suggests that logic is cognitively ubiquitous. However, this logical law cannot uncritically be imported into a theory of canine cognition. For one thing, such patterns of behavior are at least as well modeled in probabilistic (Toussaint et al. 2006), as in logical, terms. Indeed, probabilistic methods are crucial in planning tasks in uncertain environments, which is, of course, the normal case, outside mathematically specified game-playing environments. In any case, just because a behavior can be *described* in logical or probabilistic terms does not directly imply that it is *governed* by logical or probabilistic processes. The issues here are complex (see the excellent introductory chapter to Hurley & Nudds 2006) and many possibilities would

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

need to be ruled out before abandoning Lloyd Morgan's canon: that lower-level explanations of animal behavior should be preferred.

In short, we believe that cognitive science ignores logic at its peril – logic provides powerful and much needed tools, just as do other branches of mathematics. It does not, however, readily capture patterns of human reasoning, or, we suggest, cognition at large, unless generalized into a probabilistic form able directly to deal with uncertainty.

R3.5. Why probability rather than other numerical measures?

Danks & Eberhardt and **Politzer & Bonnefon** ask why we use probability, rather than other numerical measures of degrees of belief, such as confidence intervals, Dempster-Shafer belief functions (Dempster 1968; Shafer 1976), or fuzzy logic (Zadeh 1975). In *BR*, our primary motivation is practical: Bayesian probabilistic methods provide a natural way to capture human reasoning data; and more generally, Bayesian methods have swept through the brain and cognitive sciences, from understanding neural coding (Doya et al. 2007), through vision, motor control, learning, language processing, and categorization. Even within research on reasoning, Bayesian methods have proved central to understanding inductive inference (Griffiths & Tenenbaum 2005; Tenenbaum et al. 2007), causal reasoning (Sloman 2005; Tenenbaum & Griffiths 2001), and argumentation (e.g., Hahn & Oaksford 2007), as well as the primarily deductive reasoning problems considered in *BR*.[2] Moreover, probabilistic methods connect with rich literatures concerning computational inference methods (e.g., based on graphical models, Lauritzen & Spiegelhalter 1988; Pearl 1988), machine learning (e.g., Jacobs et al. 1991), and normative theories of reasoning about causality (Pearl 2000). Finally, probability also has deep relationships to other powerful concepts in the brain and cognitive sciences, including information theory (e.g., Blakemore et al. 1991) and simplicity, for example, as captured by Kolmogorov complexity theory (e.g., Chater 1996; Chater & Vitányi 2002). Thus, our focus on probability is primarily pragmatic rather than, for example, depending on a priori justifications.

Danks & Eberhardt focus, nonetheless, on justification, arguing that doubt can be cast on justifications such as the Dutch Book argument and long run convergence theorems. We see the project of rational analysis as a *user* of probability, on a par with the rest of science, for example, statistical mechanics, Bayesian image restoration, or economics. We only need to be as concerned about justification as these other endeavors. Danks & Eberhardt's worries are analogous to Berkeley's objections to Newton's infinitesimals: of considerable conceptual importance, but with little direct impact on the practical conduct of science. Nonetheless, probability is at least *better* justified than alternative formalisms for modeling uncertainty.

Politzer & Bonnefon and **Danks & Eberhardt** raise the possibility that the assumptions of the probabilistic approach may be too strong. We instead believe that they are, if anything, too weak; that is, they define minimal coherence conditions on beliefs, which need to be supplemented with richer formalisms, including, as noted in section R3.4, the ability to represent relations and quantification, and to represent and manipulate causal relations (e.g., Pearl 2000).

R3.6. Are we Bayesian enough?

Other commentators (**Over & Hajichristidis**, **Pfeifer & Kleiter**, **Stenning & van Lambalgen**) have the opposite concern: that *BR* is not Bayesian enough. Over & Hadjichristidis argue that in conditional inference, not only is the conditional premise (e.g., *if p then q*) uncertain, but so is the categorical premise, *p*. In *BR* (p. 121), we mention this general case (implying Jeffrey's rule [Jeffrey 1983]), but point out that this extra element of uncertainty appears unnecessary to capture the conditional reasoning data.

Stenning & van Lambalgen and **Pfeifer & Kleiter** also argue, in different ways, that we are insufficiently Bayesian. Stenning & van Lambalgen argue that our account of suppression effects is not Bayesian because coherent Bayesian revision of the probability space assumes "rigidity": that is, the conditional probability $P(q|p)$ remains the same if we learn the truth of a categorical premise: *p*, *q*, *not-p*, or *not-q* (and no other information). We agree. But this does *not* imply that $P(q|p)$ remains the same if we are *told* about that *p*, because pragmatic factors allow us to infer a great deal of additional information; and this information can legitimately change $P(q|p)$. It is this latter case that is relevant for reasoning with verbal materials. Thus, suppose I believe *if the key is turned, the car starts*; and I am told: "the car didn't start this morning." This would be a pragmatically pointless remark, if the key had not been turned. I therefore infer that the key was turned, and the car didn't start for some other reason. Thus, I revise down the probability of the relevant conditional P(*car starts*|*key turned*) dramatically. So the violation of rigidity, notably in this type of Modus Tollens (MT) inference, does not violate Bayesian precepts, but merely applies them to the pragmatics of utterances (see *BR*, pp. 126–128; Sobel 2004; Sober 2002).

Pfeifer & Kleiter suggest that inference can proceed locally and deductively in a mental probability logic. In such a logic, the precise probability of a conclusion cannot typically be deduced from the probabilities of the premises – but a probability *interval* can be. We adopted a similar approach to probabilistic validity for syllogisms where, according to our probabilistic semantics, quantifiers describe probability intervals. Nonetheless, in line with Stanovich and West's (2000) "fundamental computational bias," we believe that people spontaneously contextualize and elaborate verbal input, by adding information from world knowledge. Indeed, it takes substantial cognitive effort *not* to do this. Consequently, we think it unlikely that people reason deductively about probability intervals.

R3.7. Measuring confirmation

People are not merely passive observers. They can actively search for information to help test hypotheses, or to

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

achieve specific goals. In *BR*, we outline "rational" accounts for both cases. Where people test between hypotheses, a natural objective is to search for data in order to maximize the expected amount of *information* that will be gained in the task (Shannon & Weaver 1949). This is "disinterested" inquiry. Where people gain information to help achieve specific goals, then a natural objective is to choose information to maximize expected utility (balancing costs of information search with the improved choices that may result from new information). This is "goal-directed" inquiry. In *BR*, we note that different variations of Wason's selection task are appropriately captured by versions of one or other model. In particular, we showed how attending to the goal-directed case avoids the postulation of specific machinery, such as "cheater-detection" modules (e.g., Cosmides 1989), to explain patterns of experimental data (e.g., *BR*, pp. 191–98).

Focusing on disinterested inquiry, **Nelson** notes that a wide range of normative and descriptive proposals for assessing the strength of information in a piece of data have been proposed. In testing these models against a wide range of psychological data (Nelson 2005), he finds that the information-theoretic measure implicit in our analysis stands up well against competitors, although it is not picked out uniquely by the empirical data.

Poletiek notes a further interesting link to philosophy of science, noting that Popper's measure of severity of test is equivalent to $P(e|H)/P(e)$, for data e and hypothesis H. And the logarithm of this quantity just *is* the amount of information carried by the evidence e about H – the quantity which we use in our model of disinterested inquiry in the selection task. This quantity is also used as a measure of the degree to which a theory is confirmed by the data in confirmation theory (Milne 1996). This is, as Poletiek notes, particularly interesting, given that Popper's measure of severity of test is part of a theoretical framework which aims to entirely avoid the notion of confirmation (see also Milne 1995). Thus, our account of the selection task could be recast, from a Popperian standpoint, as a rational analysis in which people attempt to choose data to provide the more severe possible tests for their hypotheses.

R4. Rational analysis, algorithmic processes, and neural implementation

BR is primarily concerned with the rational analysis of human reasoning (e.g., Anderson 1990; 1991a; Chater & Oaksford 2008a; Oaksford & Chater 1998b). In this section, we consider the role of rational analysis in the brain and cognitive science and whether this style of explanation is fundamentally flawed.

R4.1. The power of rational analysis

Hahn notes that the shift away from considerations of algorithms and representations, encouraged by rational analysis, has led to a substantial increase in explanatory power in cognitive science, in a number of domains. Where the underlying explanation for an aspect of cognition arises from the rational structure of the problem being solved, there focusing on specific algorithmic and neural mechanisms may be unhelpful. Therefore, building specific algorithmic models (e.g., connectionist networks) of a phenomenon may replicate the phenomenon of interest (by virtue of being an adaptive solution to the "rational" problem in hand), but may throw little light on *why* it occurs.

R4.2. Normativity and rational analysis

Evans and **Schroyens** are concerned about the normative aspect of rational analysis. Evans questions whether normativity is a proper part of a computational-level analysis of human reasoning, and by implication, cognition in general, and recommends a switch to an *ecological* notion of rationality. He suggests rationality should concern how well people are adapted to their environment, which may not require following the prescriptions of any normative theory of reasoning (cf. Gigerenzer & Goldstein 1996).

We suggest, however, that ecological rationality does not replace, but rather, complements normative rationality. Normative considerations are still required to explain *why* a particular algorithm works, given a particular environment; indeed, this is precisely the objective of rational analysis. Thus, for example, in arguing for the ecological rationality of various fast and frugal heuristics (Gigerenzer et al. 1999), Gigerenzer and colleagues appeal to a Bayesian analyses to explore the type of environmental structure for which their algorithms succeed (e.g., Martignon & Blackmond-Laskey 1999). Thus, rational analysis cannot be replaced by, but seeks to explain, ecological rationality.

Note, too, that rational analysis is goal-relative: it specifies how best to achieve a given goal, in a given environment, with given constraints (Anderson 1990; Oaksford & Chater 1998b). So, if your goal is to land a rocket on the moon, your guidance system ought to respect classical physics; if your goal is to avoid contradictions, you ought to reason according to standard logic; and if your goal is to avoid accepting bets that you are bound to lose, you ought to follow the rules of probability theory (see Ch. 2 of *BR*).

Ignoring the goal-relativity of rational analysis leads **Schroyens** to suggest that we have fallen into Moore's (1903) "naturalistic fallacy" in ethics: that we have attempted to derive an "ought" from an "is." Moore's concern is that no facts about human behavior, or the world, can justify an ethical theory. Ethics is concerned with non-relative notions of "ought": the aim is to establish universal principles of right behavior. But the goal-relativity of rational analysis makes it very different from the domain of ethics, because it is *conditional*. Rational analysis considers: *if* you have objective O, given an environment E, and constraints C, then the optimal action is A. Ethics, by contrast, considers whether O is a justifiable objective. And the nature of the solution to a well-specified optimization problem is itself firmly in the domain of facts.

Indeed, were **Schroyens**' concern valid, then its consequences would be alarming, sweeping away functional explanation in biology and rational choice explanation in economics. Yet in all cases, rational/optimality explanations are used to derive empirical predictions; and, as in any scientific enterprise, the assumptions of the rational/optimality accounts are adjusted, where

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

appropriate, to give a better fit with empirical predictions. Specifically, empirical data lead to revision of *empirical* assumptions in the rationality/optimality analysis – the empirical data does *not* lead to a revision of the laws of logic, probability, or any other rational theory.

Khalil raises the opposite concern: that we use rational explanation too narrowly. He argues that the style of optimality explanation that we advocate applies just as well in the explanation of non-cognitive biological structures as it does to cognitive processes – he argues that, in the sense of rationality used in *BR*, stomachs are just as rational as cognitive mechanisms. This concern appears purely terminological; we reserve "rationality" for information processing systems. But rational analysis is, indeed, parallel to optimality explanation in biology (Chater et al. 2003).

R4.3. Relevance of the algorithmic level

McKenzie and **Griffiths** note, however, that advocating rational analysis does not make the challenges concerning algorithmic, and indeed neural, implementation, disappear. Moreover, the mapping between levels of explanation need not necessarily be straightforward, so that a successful probabilistic rational analysis of a cognitive task does not necessarily require that the cognitive system be carrying out probabilistic calculations – any more than the bird is carrying out aerodynamic calculations in growing a wing perfectly adapted for flight. Nonetheless, in many contexts, it *is* natural to see cognition as carrying out probabilistic calculations; and a prior rational analysis (or, in Marr's [1982] terms, *computational* level of explanation) is extremely valuable in clarifying what calculations need to be carried out. Without a "rational analysis" for arithmetic calculations (i.e., a mathematical theory of elementary arithmetic), understanding which algorithms might be used by a pocket calculator, let alone how those algorithms might be implemented in silicon, would be impossible. Griffiths outlines key challenges for creating an algorithmic-level theory of cognition, viewed from a Bayesian perspective; and this perspective dovetails nicely with work viewing neural machinery as carrying out Bayesian inference (e.g., Ma et al. 2006; Rao et al. 2002), which we consider briefly further on.

BR is largely focused on rational level explanation (Anderson 1990; 1991a). Indeed, following Marr (1982), we argued that understanding the rational solution to problems faced by the cognitive system crucially assists with explanation in terms of representations and algorithms, as stressed by **Hahn** and **Griffiths**. In *BR*, this is illustrated by our model of syllogistic reasoning, which proposes a set of "fast and frugal" heuristics (Gigerenzer & Goldstein 1996) for generating plausible conclusions, rooted in a Bayesian rational analysis (Chater & Oaksford 1999b). More recently, we have suggested methods for causal and conditional reasoning, based on "mental mechanisms" (Chater & Oaksford 2006; Ali et al., in press) directly building on rational and algorithmic models inspired by the literature on Bayesian networks (Glymour 2001; Pearl 1988; 2000). Moreover, an explicit algorithmic implementation of our probabilistic account of conditional inference has been constructed using a constraint satisfaction neural network (Oaksford & Chater, in press). Moreover, there is a significant movement in current cognitive science that focuses on developing and employing Bayesian machine learning techniques to model cognition at both the rational and algorithmic levels (e.g., Griffiths et al. 2007; Kemp & Tenenbaum 2008).

Evans' concern that we ignore the algorithmic level is therefore puzzling. He worries that *BR* recommends that one should "observe some behaviour, assume that it is rational, find a normative theory that deems it to be so, and then . . . nothing else, apparently." We assume that the ellipsis should, in Evans' view, be fleshed out with an algorithmic, process-based explanation, which should then be subject to rigorous empirical test. The abovementioned list of algorithmic level proposals inspired by Bayesian rational analysis, both in the domain of reasoning and in cognitive science more generally, gives grounds for reassurance. Moreover, the extensive empirical testing of these models (Green & Over 1997; 2000; McKenzie & Mikkelsen 2000; 2007; McKenzie et al. 2001; Nelson 2005; Oaksford & Moussakowski 2004; Oaksford & Wakefield 2003; Oaksford et al. 1999; 2000; Tenenbaum 1999) should allay concerns that rational analysis provides no testable predictions. Ironically, the only theories in the psychology of reasoning that have been algorithmically specified, aside from those within the Bayesian tradition, are directly based on another rational level theory: logic (Johnson-Laird 1992; Rips 1994). Theorists who have instead focused primarily on heuristics for reasoning have couched their explanations in purely verbal terms (Evans 1989; Evans & Over 2004). This indicates, we suggest, that rational analysis assists, rather than impedes, algorithmic explanation.

R4.4. Relevance of neural implementation

Bayesian rational analysis is, moreover, appealing because it appears to yield algorithms that can be implemented in the brain. In *BR* (Ch. 4), we observed that the Bayesian approach was sweeping across cognitive psychology. We might also have added that its influence in computational neuroscience is at least as significant (Friston 2005). Although our Bayesian analyses of higher-level reasoning do not directly imply Bayesian implementations at the algorithmic level, it is intriguing that influential theorists (Doya et al. 2007; Friston 2005; Ma et al. 2006) view Bayesian inference as providing the driving computational principle for neural information processing. Such models, using population codes (Ma et al. 2006), which avoid treating the brain as representing probabilities directly on a numerical scale, can model simple perceptual decision tasks (Gold & Shadlen 2000). Such convergence raises the possibility that Bayesian rational analyses of reasoning may one day find rather direct neural implementations.

De Neys specifically appeals to the implementation level in commenting on *BR*. He draws attention to imaging studies of reasoning that suggest a role for the anterior cingulate cortex in detecting conflict and inhibiting responses. As we have seen (sect. R3.4), such a role is entirely consistent with Bayesian approaches. Indeed, more broadly, imaging work on human reasoning, pioneered by Goel (e.g., Goel 2007), is at an exploratory stage, and currently provides few constraints on theory. Moreover, as we have seen, where cognitive

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

neuroscientists concentrate on *what* computations the brain performs rather than *where*, the emerging answer is Bayesian.

R4.5. Optimality and rational analysis

A range of commentators (e.g., **Brighton & Olsson**, **Danks & Eberhardt**, **Evans**, and **Schroyens**) argue that the methodology of rational analysis faces conceptual problems. Our general response to these concerns is pragmatic. As with any methodology, we see rational analysis, using probabilistic methods or otherwise, as primarily to be judged by its results. Anderson's path-breaking work (1990; 1991a), and the huge literature on Bayesian models across the brain and cognitive sciences, of which *BR* is a part, is therefore, in our view, the best argument for the value of the approach. Parallels with closely related work in behavioral ecology and rational choice explanation in economics give further weight to the view that a "rational" style of explanation can yield considerable insights. But, like any style of explanation, rational analysis has its limits. Just as, in biology, some behaviors or structures are products of "history" rather than adaptation (Carroll 2005), and some economic behaviors are the product of cognitive limitations (e.g., Ariely et al. 2003; Thaler 2005), so in the brain and cognitive sciences, we should expect some phenomena to arise from specific aspects of algorithms/representations or neural implementation.

We are therefore happy to agree with commentators who suggest that there are cognitive phenomena for which purely rational considerations provide an incomplete, or indeed incorrect, explanation (e.g., **Brighton & Olsson**, **Evans**). We also agree that rational analysis is challenged where there are many, perhaps very different, near-optimal rational solutions (Brighton & Olsson). In such situations, rational analysis provides, at best, a range of options – but it does not provide an explanation of why one has been chosen. Nonetheless, these issues often cause few problems in practice, as the results in *BR* and in the wider program of rational explanation illustrate.

We agree, moreover, with concerns that finding *exactly* the optimal solution may be over-restrictive (**Brighton & Olsson**, **Evans**). Consider the case of perceptual organization, where the cognitive system must decide between multiple interpretations of a stimulus (Gregory 1970; von Helmholtz 1910/1925). Accounts based on Bayesian probability and on the closely related idea of maximizing simplicity (Chater 1996; Hochberg & McAlister 1953; Leeuwenberg & Boselie 1988) adopt the perspective of rational analysis, but they do so *comparatively*. That is, the perceptual system is presumed to choose interpretation A, rather than interpretation B, if A is more likely than B (or, in simplicity-based formulations, if it provides a *simpler* encoding of the sensory input). Neither the likelihood nor the simplicity principles in perceptual organization are presumed to imply that the perceptual system can optimize likelihood/simplicity – and indeed, in the general case, this is provably impossible (see Chater 1996, for discussion). Indeed, we suspect that rational analysis will, in many cases, primarily be concerned with providing a measure of the relative "goodness" of different cognitive processes or behaviors; and it is explanatory to the degree to which the "good" mechanisms are more prevalent than the "bad." The parallel with evolutionary explanation seems to be exact here: Inclusive fitness provides a crucial explanatory measure in explaining the evolution of biological structures, but the explanatory "bite" is comparative (i.e., in a certain environment, a flipper yields greater fitness than a leg). There is no assumption that biological evolution, in any context, reaches a state of completely optimized perfection; indeed, quite the reverse (Jacob 1977). Thus, Evans' emphasis on satisficing rather than optimizing, and Brighton & Olsson's focus on relative rationality, seem to us entirely consistent with *BR*.

Note, too, that in modeling many aspects of cognition, a full-scale rational analysis (specifying a task, environment, and computational limitations) may not be required. For example, conditional inference can be modeled in Bayesian terms, assuming only a probabilistic interpretation of the premises, and the requirement of maintaining consistent degrees of belief. The success of the probabilistic, rather than a logical, interpretation of the premises can be assessed by comparing the predictions of both approaches to data on human reasoning, as well general philosophical principles.

Brighton & Olsson also raise a different concern: that the specific sets of probabilistic assumptions (such as the independence assumptions embodied in naïve Bayes) may sometimes be justified *not* by rational analysis, but instead in the light of their general, formal properties, combined with empirical success in solving some externally defined task (e.g., estimating the relative sizes of German cities, Gigerenzer & Goldstein 1996). For example, a model such as naïve Bayes, they note, may be effective because it has few parameters and hence avoids over-fitting. We suggest, however, that this is not a separate type of explanation of inferential success, distinct from Bayesian rational analysis. Instead, the justification for preferring simple models can, itself, be provided in terms of Bayesian reasoning, and closely related formalisms, including minimum description length (Chater & Oaksford 2008b; MacKay 2003; Rissanen 1989; Vitányi & Li 2000).

R4.6. Need rational explanation be causal?

Brighton & Olsson, together with **Danks & Eberhardt**, raise the fundamental concern that rational explanation does not provide a *causal* explanation of behavior. We agree. Rational explanation is teleological (Fodor 1968) – it explains by reference to purpose, rather than cause.

In particular, rational explanation does not require that the rational analysis is itself represented in the mind of the agent, and does not, therefore, imply that behavior is governed by any such representation. Aerodynamics may provide an optimality-based explanation of the shape of the bird's wing; but aerodynamic calculations by the bird (or any other agent) are not causally responsible for the wing's shape.

Similarly, delineating the circumstances in which algorithms such as naïve Bayes (**Brighton & Olsson**; Domingos & Pazzani 1997), Take the Best (Gigerenzer & Goldstein 1996; Martignon & Hoffrage 1999), or unit-weighted regression (Dawes 1979) are reliable may require highly sophisticated rational explanation. Yet a

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

cognitive system that employs such models may know nothing of such rational explanations – and indeed, these rational assumptions typically play no causal role in determining the behavior. Thus, in behavioral ecology, for example, the strategies animals use in foraging, mate selection, and so on, are typically explained using optimality explanations; but animals are not assumed to carry out optimality calculations to validate their behavioral strategies.

Danks & Eberhardt suggest that there is a "requirement for a teleological explanation that the normative principle must have played a causal role – ontogenetic, phylogenetic, or both – in the behavior's existence or persistence. 'Origin stories' are required for teleological explanation." We find this claim puzzling: normative principles, and rational explanations in general, are abstract – they are not part of the causal realm. Thus, a Bayesian rational analysis can no more *cause* a particular piece of behavior or reasoning, than the principles of arithmetic *cause* a calculator to display a particular number. Teleological explanations are distinctively non-causal, and necessarily so.

In this section, we have considered concerns about the general project of rational analysis. We now turn to consider specific issues relating to the rational models and empirical data presented in *BR*.

R5. Reconsidering models and data

Even if the broad sweep of arguments from the preceding sections is endorsed, there remain doubts about the details of the particular models described in *BR* and their ability to account for human reasoning data. Indeed, in the commentaries, issues of detail emerge most often between researchers who otherwise are in broad agreement. It is in this light that we consider the comments of **Liu**, **Oberauer**, **Over & Hadjichristidis**, and **Wagenmakers**. We also consider here **Halford**'s comments on syllogistic reasoning, drawn from a different framework.

R5.1. Conditional inference

Liu, **Oberauer**, and **Over & Hadjichristidis**, who have also advocated a probabilistic approach (in particular, to conditional inference), have concerns about our specific model. We addressed, in section R3.6, Over & Hadjichristidis's argument that we are not Bayesian enough, and that we should employ Jeffrey's rule to deal with uncertainty in the categorical premise of conditional inference. We pointed out that we too explicitly adopted Jeffrey's rule in *BR*. They also cite some unpublished results apparently showing that people have an imperfect understanding of Jeffrey's rule. These results are intriguing and suggest that more extensive empirical testing of this rule is required.[3]

Oberauer argues that our models of conditional inference and data selection may lead to absurdity. He argues that if the marginals, $P(p)$ and $P(q)$, remain fixed, which he describes as "axiomatic" in our theory,[4] then if one increases the probability that someone gets a headache, given they take drug X, then those who don't take X will get fewer headaches. This apparent absurdity stems from a conflation in Oberauer's description between the factual and the epistemic/doxastic: Changing this conditional *degree of belief* does not mean that these people *actually* achieve these benefits. In ignorance of the real conditional probability, but knowing the values of the marginals, I *should* revise my *degree of belief* that not taking this drug leads to fewer headaches. Yet this will only be appropriate when the marginals are known – which is clearly inappropriate in Oberauer's example.

Oberauer also perceives an inconsistency between our adoption of The Equation – $P(\textit{if } p \textit{ then } q) = P(q|p)$ – and our use of a contingency table to represent the conditional hypothesis in data selection. However, by The Equation there is only sufficient information in the premises of a conditional inference to draw MP by Bayesian (or Jeffrey) conditionalization (at least a point value). The remaining inferences can only be drawn on the assumption that people use the marginals to calculate the relevant conditional probabilities, for example, $P(\neg q|\neg p)$ for Denying the Antecedent (DA). Once $P(q|p)$ and the marginals are fixed, the contingency table is determined. Knowing the meaning of a statement is often equated with knowing the inferences that a statement licenses (Dowty et al. 1981). According to The Equation, the conditional only licenses "probabilized" MP. Probabilistically, to draw further inferences requires more information to be drawn from world knowledge. Hence, there is no inconsistency. Moreover, in the selection task, people are presented with an array of possible evidence types that makes the marginals relevant in the same way as presenting more than just MP in the conditional inference task. The degree of belief that is modified by selecting data is in the conditional *and the marginals*, which constitute the dependence and independence models. Thus, Oberauer's concerns can be readily addressed.

Oberauer also suggests that contingency tables are consistent with a probabilistic contrast approach, that is, the measure of the strength of an argument, for example, MP, is $P(q|p) - P(q|\neg p)$. It is for this reason that we believe that argument strength may indeed be two-dimensional (Oaksford & Hahn 2007). The conditional probability alone can mean that a good argument leads to no increase in the degree of belief in the conclusion, for example, for MP when $P(q|p) = P(q) = 1$. The probabilistic contrast (and other measures; see, e.g., **Nelson**, **Poletiek**, and Oaksford & Hahn 2007) captures the change in the probability of the conclusion brought about by an argument. Oberauer suggests that there is no evidence for people's use of the probabilistic contrast. Yet Over et al. (2007) found significant sensitivity to $P(q|\neg p)$, consistent with some use of the probabilistic contrast or a related measure of change, and the evidence is currently equivocal.

Oberauer also raises two concerns over evidence for our model of conditional inference. First, fitting a model with two free parameters to four data points "is no convincing accomplishment." Even so, as **Hahn** observes, the move to detailed model fitting of quantitative data represents significant progress in the psychology of reasoning (for early examples, see Krauth [1982] and Klauer [1999]). Moreover, in *BR* (pp. 146–49) we fitted the model to the 32 data points produced in Oaksford et al.'s (2000) Experiment 1 using only nine parameters, collapsing far more degrees of freedom than the model fitting reported in Oberauer (2006). Although Oberauer (2006) found poorer fits for our model than alternative theories,

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Oaksford and Chater (2008) found that the revised model presented in *BR* may provide better fits to Oberauer's data. Second, Oberauer argues that the most relevant empirical evidence comes from studies where probabilities were directly manipulated, of which he mentions two, Oaksford et al. (2000) and Oberauer et al. (2004). Moreover, he argues that their results are equivocal. However, several other studies have manipulated probabilities in conditional inference and found evidence in line with a probabilistic account (George 1997; Liu 2003; Liu et al. 1996; Stevenson & Over 1995). Oberauer also leaves aside the many studies on data selection showing probabilistic effects (see *BR*, Ch. 6).

Liu's arguments about second-order conditionalization point, we think, to an important factor that we have yet to consider in reasoning, that is, the effects of context. **Liu** has found that people often endorse the conclusion that, for example, *Tweety flies* on being told that *Tweety is a bird* in the absence of the conditional premise (reduced problems). This occurs because they fill in this information from world knowledge. However, **Liu** also found that endorsements increase when the conditional premise is added (complete problems). In *BR*, we argued that this occurs because people take the conditional premise as evidence that the conditional probability is higher (an inference that may arise from conversational pragmatics). **Liu** argues that our account implies that manipulations affecting reduced problems should also affect complete problems and provides evidence against this. Yet context, both cognitive and physical, may explain these differences in a way similar to recent studies of decision-making (Stewart et al. 2006). For example, suppose one is told about two swanneries, both containing the same number of swans. In one, 90% of swans are black (P(black|swan) = .9); in the other, 90% of swans are white (P(white|swan) = .9). On being told that *Tweety is a swan*, presumably one would only endorse *Tweety is white* at .5. This is because conversational pragmatics and world knowledge indicate that Tweety is in one of the just mentioned swanneries, but the dialogue up to this point does not indicate which one.[5] However, the addition of the conditional premise *if a bird is a swan it is white* immediately indicates which swannery is being talked about, that is, the one in which P(white|swan) is high, and now endorsements should increase to .9. Clearly, although manipulations of the relative number of swans in each swannery might affect the reduced problem, they should not affect the complete problem. So if the swannery in which most swans are black were one tenth of the size of the other swannery, then, given natural sampling assumptions, endorsements for the reduced problem should increase to .83, but endorsements of the complete problem should remain the same.

R5.2. Data selection

Wagenmakers raises a variety of concerns about our optimal data selection model. First, why do we concede that people should select the standard "logical" A card and 7 card choices, *if* the rule only applies to the four cards? In *BR* (p. 210), we argue that people rarely use conditionals to describe just four objects – they assume that the cards are drawn from a larger population. Consequently, we quite explicitly do not make the counterintuitive prediction that Wagenmakers ascribes to us. Second, Wagenmakers wonders why – when all cards carry some information – do participants not select all the cards, if they are maximizing information gain? We assume that the pragmatics of the task suggests to participants that they should select some cards, but not others (*BR*, pp. 200–201). Third, Wagenmakers suggests that incentivized individuals with more time might make the logical response. Work on individual differences (e.g., Stanovich & West 2000) is consistent with the view that logical competence is learned, either directly (e.g., studying logic or math) or indirectly (e.g., learning to program or learning conventional, non-Bayesian statistics); such logical competence is a prerequisite for "logical" responses, and covaries with IQ as measured in University populations. Wagenmakers also remarks that, as Bayesians, we should avoid null hypothesis testing in statistically assessing our models. This choice is purely pragmatic: it conforms to the current demands of most journals.

R5.3. Syllogisms and development

Halford argues that mental models theory and a relational complexity measure fit the data as well as the probability heuristics model (PHM), conceding, however, that only PHM generalizes to *most* and *few*. Copeland (2006) has also recently shown that PHM provides better fits than mental models and mental logic for extended syllogisms involving three quantified premises. Halford also suggests that basing confidence in the conclusion on the least probable premise, as in our *max*-heuristic, is counterintuitive. He proposes that confidence should instead be based on relational complexity, which covaries with the least probable premise. But perhaps Halford's intuition goes the wrong way: the least probable premise is the most informative; and surely the more information you are given, the stronger the conclusions you can draw?

De Neys and **Straubinger, Cokely, & Stevens** (**Straubinger et al.**) both argue that there are important classes of evidence that we do not address. De Neys argues that attention to latency data and imaging studies provides a greater role for logic, a claim we disputed earlier. Note, also, that the algorithmic theory in PHM has been applied to latency data and accounts for the data, as well as mental models (Copeland & Radvansky 2004). Straubinger et al. are concerned that we ignore developmental data. In particular, they view the findings on the development of working memory as providing a particular challenge to a Bayesian approach. They do, however, acknowledge that in different areas (e.g., causal reasoning), Bayesian ideas are being successfully applied to developmental data (Navarro et al. 2006; Sobel et al. 2004). Straubinger et al.'s emphasis on working memory provides good reason to believe that our particular approach to deductive reasoning may extend to development. Copeland and Radvansky (2004) explicitly related working-memory limitations to PHM, finding that it provided as good an explanation as mental models theory of the relationship between working-memory capacity and reasoning performance. This result provides some indication that, at least for syllogistic reasoning, developmental trajectories explicable by mental models may be similarly amenable

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

to explanation in terms of probability heuristics. Our approach also provides a natural way in which experience, leading to the learning of environmental statistics, might influence reasoning development. Exploring these possibilities must await future research.

R6. The Bayesian turn

BR is part of a larger movement across the brain and cognitive sciences – a movement which sees cognition as centrally concerned with uncertainty; and views Bayesian probability as the appropriate machinery for dealing with uncertainty. Probabilistic ideas have become central to theories of elementary neural function (Doya et al. 2007), motor control (Körding & Wolpert 2004), perception (Knill & Richards 1996), language processing (Manning & Schütze 1999), and high-level cognition (Chater & Oaksford 2008a; Chater et al. 2006). They also cut across Marr's (1982) computational (Anderson 1990; Pearl 2000), algorithmic (Jacobs et al. 1991), and implementational (Doya et al. 2007) levels of explanation. In arguing that commonsense reasoning should be understood in terms of probability, we are merely recasting Laplace's (1814/1951) classic dictum concerning the nature of probability theory: "The theory of probabilities is at bottom nothing but common sense reduced to calculus."

NOTES

1. Although Braine and O'Brien (1991) explicitly reject the use of relevance logic (Anderson & Belnap 1975), this does provide an interesting possible route for developing these ideas. In particular, interpretations of the semantics of relevance logics as a ternary relation between possible worlds, or from an information-theoretic perspective, as a ternary relation between a source, a receiver, and a channel (Restall 1996), may provide interesting connections with nonmonotonic reasoning.

2. By contrast, we know of just one paper in the psychology of reasoning discussing Dempster-Shafer belief functions, namely, George (1997).

3. Its normative status has also been questioned for many years (see, e.g., Field 1978).

4. This is despite the fact that they were not fixed in Oaksford and Chater (1994).

5. Of course, different assumptions would yield different results. For example, if the previous dialogue had been talking about the swannery, where most swans are black, just before introducing Tweety, the assumption may be that Tweety comes from that swannery and so *Tweety is white* might only be endorsed at .1.

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Adams, E. W. (1975) *The logic of conditionals: An application of probability to deductive logic*. Reidel. [arMO, NP]

(1998) *A primer of probability logic*. CLSI Publications. [arMO]

Ali, N., Schlottmann, A., Shaw, A., Chater, N. & Oaksford, M. (in press) Causal discounting and conditional reasoning in children. In: *Cognition and conditionals: Probability and logic in human thought*, ed. M. Oaksford & N. Chater. Oxford University Press. [rMO]

Anderson, A. & Belnap, N. D. (1975) *Entailment: The logic of relevance and necessity, vol. 1*. Princeton University Press. [rMO]

Anderson, J. R. (1983) *The architecture of cognition*. Harvard University Press. [rMO]

(1990) *The adaptive character of thought*. Erlbaum. [TLG, UH, ELK, CRMM, arMO, WS]

(1991a) Is human cognition adaptive? *Behavioral and Brain Sciences* 14:471–84; discussion 485–517. [HB, arMO, NS]

(1991b) The adaptive nature of human categorization. *Psychological Review* 98:409–29. [aMO]

Anderson, J. R. & Matessa, M. (1998) The rational analysis of categorization and the ACT-R architecture. In: *Rational models of cognition*, ed. M. Oaksford & N. Chater, pp.197–217. Oxford University Press. [aMO]

Anderson, J. R. & Milson, R. (1989) Human memory: An adaptive perspective. *Psychological Review* 96:703–19. [aMO]

Anderson, J. R. & Schooler, L. J. (1991) Reflections of the environment in memory. *Psychological Science* 2:396–408. [aMO]

Antoniou, G. (1997) *Nonmonotonic reasoning*. MIT Press. [KS]

Ariely, D., Loewenstein, G. & Prelec, D. (2003) Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics* 118:73–105. [rMO]

Aristotle (1980) *Nicomachean ethics*, trans. W. D. Ross. Clarendon Press. [aMO]

Ball, L. J., Philips, P., Wade, C. N. & Quayle, J. D. (2006) Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology* 53:77–86. [WDN]

Baltes, P. B., Staudinger, U. M. & Lindenberger, U. (1999) Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology* 50:471–507. [NS]

Baron, J. (1981) An analysis of confirmation bias. Paper presented at the 22nd Annual Meeting of the Psychonomic Society. 6–8 November, 1981, Philadelphia, PA. [aMO]

(1985) *Rationality and intelligence*. Cambridge University Press. [JDN, aMO]

(2004) Normative models of judgment and decision making. In: *Blackwell handbook of judgment and decision making*, ed. D. J. Koehler & N. Harvey, pp. 19–36. Blackwell. [JDN]

Barwise, J. & Cooper, R. (1981) Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–219. [aMO]

Becker, G .S. (1976) *The economic approach to human behaviour*. University of Chicago Press. [ELK]

Benferhat, S., Bonnefon, J. F. & Da Silva Neves, R. (2005) An overview of possibilistic handling of default reasoning, with experimental studies. *Synthese* 146:53–70. [GP]

Bennett, J. (2003) *A philosophical guide to conditionals*. Oxford University Press. [arMO]

Bernoulli, J. (1713/2005) *Ars conjectandi* [*The art of conjecture*], trans. E. D. Sylla. Johns Hopkins University Press. [aMO]

Biazzo, V., Gilio, A., Lukasiewicz, T. & Sanfilippo, G. (2005) Probabilistic logic under coherence: Complexity and algorithms. *Annals of Mathematics and Artificial Intelligence* 45(1–2):35–81. [NP]

Blakemore, C., Adler, K. & Pointon, M. (1990) *Vision: Coding and efficiency*. Cambridge University Press. [rMO]

Bookstaber, R. & Langsam, J. (1985) On the optimality of coarse behavior rules. *Journal of Theoretical Biology* 116:161–93. [HB]

Boole, G. (1854/1958) *An investigation of the laws of thought*. Macmillan/Dover. (Reprinted by Dover, 1958). [aMO]

Bovens, L. & Hartmann, S. (2003) *Bayesian epistemology*. Clarendon Press. [aMO]

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

(1990) The natural logic approach to reasoning. In: *Reasoning, necessity, and logic: Developmental perspectives*, ed. W. F. Overton. Erlbaum. [DPO]

Braine, M. D. S. & O'Brien, D. P. (1991) A theory of *if*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review* 98:182–203. [DPO, rMO]

Braine, M. D. S. & O'Brien, D. P., eds. (1998) *Mental logic*. Erlbaum. [DPO]

Brighton, H. & Gigerenzer, G. (2008) Bayesian brains and cognitive mechanisms: Harmony or dissonance? In: *The probabilistic mind: Prospects for Bayesian cognitive science*, ed. N. Chater & M. Oaksford, pp. 189–208, Oxford University Press. [HB, NS]

Bullinaria, J. A. & Chater, N. (1995) Connectionist modelling: Implications for neuropsychology. *Language and Cognitive Processes* 10:227–64. [UH]

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

Byrne, R. M. J., Espino, O. & Santamaria, C. (1999) Counterexamples and the suppression of inferences. *Journal of Memory and Language* 40:347–73. [I-mL]

Carnap, R. (1950) *The logical foundations of probability*. University of Chicago Press. [rMO, FHP]

Carroll, S. (2005) *Endless forms most beautiful*. W. W. Norton. [rMO]

Chater, N. (1995) Neural networks: The new statistical models of mind. In: *Connectionist models of memory and language*, ed. J. P. Levy, D. Bairaktaris,

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

J. A. Bullinaria & P. Cairns, pp. 207–27. University College London Press. [UH]
(1996) Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review* 103:566–81. [rMO]
Chater, N., Crocker, M. & Pickering, M. (1998) The rational analysis of inquiry: The case of parsing: In: *Rational models of cognition*, ed. M. Oaksford & N. Chater, pp. 441–68. Oxford University Press. [aMO]
Chater, N. & Manning, C. D. (2006) Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* 10:335–44. [TLG, aMO]
Chater, N. & Oaksford, M. (1999a) Ten years of the rational analysis of cognition. *Trends in Cognitive Science* 3:57–65. [TLG]
(1999b) The probability heuristics model of syllogistic reasoning. *Cognitive Psychology* 38:191–258. [arMO]
(2006) Mental mechanisms: Speculations on human causal learning and reasoning. In: *Information sampling and adaptive cognition*, ed. K. Fiedler & P. Juslin, pp. 210–38. Cambridge University Press. [rMO]
eds. (2008a) *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press. [UH, rMO]
(2008b) The probabilistic mind: Where next? In: *The probabilistic mind: Prospects for Bayesian cognitive science*, ed. N. Chater & M. Oaksford, pp. 501–14. Oxford University Press. [rMO]
Chater, N., Oaksford, M., Heit, E. & Hahn, U. (in press) Inductive logic and empirical psychology. In: *The handbook of philosophical logic, vol. 10*, ed. S. Hartmann & J. Woods. Springer. [rMO]
Chater, N., Oaksford, M., Nakisa, R. & Redington, M. (2003) Fast, frugal and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes* 90:63–86. [JDN, rMO, WS]
Chater, N., Tenenbaum, J. B. & Yuille, A. eds. (2006) Probabilistic models of cognition: Where next? *Trends in Cognitive Sciences* 10:335–44. [Special Issue.] [UH, rMO]
Chater, N. & Vitányi, P. (2002) Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences* 7:19–22. [rMO]
Cheng, P. W. & Holyoak, K. J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology* 17:391–416. [aMO]
Chomsky, N. (1957) *Syntactic structures*. Mouton. [aMO]
(1965) *Aspects of the theory of syntax*. MIT Press. [aMO]
Chou, T. H. (2007) The mechanism of suppression effects in conditional reasoning. Unpublished Doctoral dissertation, National Chung-Cheng University, Department of Psychology, Chia-Yi, Taiwan. [I-mL]
Cohen, L. J. (1981) Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences* 4:317–70. [JDN, aMO]
Cokely, E. T., Kelley, C. M. & Gilchrist, A. L. (2006) Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychonomic Bulletin and Review* 13:991–97. [NS]
Coletti, G. & Scozzafava, R. (2002) Probabilistic logic in a coherent setting. Kluwer. [NP]
Conlin, J. A., Gathercole, S. E. & Adams, J. W. (2005) Children's working memory: Investigating performance limitations in complex span tasks. *Journal of Experimental Child Psychology* 90:303–17. [NS]
Cooper, G. F. (1990) The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence* 42:393–405. [HB]
Copeland, D. E. (2006) Theories of categorical reasoning and extended syllogisms. *Thinking and Reasoning* 12:379–412. [rMO]
Copeland, D. E. & Radvansky, G. A. (2004) Working memory and syllogistic reasoning. *Quarterly Journal of Experimental Psychology* 57A:1437–57. [arMO]
Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187–276. [arMO]
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. [JDN]
(2000) Evolutionary psychology and the emotions. In: *Handbook of emotions*, 2nd edition, ed. M. Lewis & J. M. Haviland-Jones, pp. 91–115. Guilford. [aMO]
Courville, A. C., Daw, N. D. & Touretzky, D. S. (2006) Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences* 10:294–300. [aMO]
Damer, E. T. (2005) *Attacking faulty reasoning: A practical guide to fallacy-free arguments, 5th edition*. Thomson Wadsworth. [WS]
Danks, D. (2008) Rational analyses, instrumentalism, and implementations. In: *The probabilistic mind: Prospects for Bayesian cognitive science*, ed. N. Chater & M. Oaksford. Oxford University Press. [HB]
Daston, L. (1988) *Classical probability in the enlightenment*. Princeton University Press. [aMO]
Davidson, D. (1984) *Inquiries into truth and interpretation*. Oxford University Press. [aMO]
Dawes, R. M. (1979) The robust beauty of improper linear models in decision making. *American Psychologist* 34:571–82. [rMO]
de Finetti, B. (1974/1975) *Theory of probability, vols. 1 & 2*. Wiley. [NP]
Dempster, A. P. (1968) A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B* 30:205–47. [rMO]
Dennett, D. (1987) *The intentional stance*. MIT Press. [WS]
De Neys, W. & Franssens, S. (2007) The nature of belief inhibition during thinking: How reasoning impairs memory. *Proceedings of the Annual Meeting of the Cognitive Science Society* 29:8. [WDN]
De Neys, W. & Glumicic, T. (2008) Conflict monitoring in dual process theories of thinking. *Cognition* 106:1248–99. [WDN]
De Neys, W., Vartanian, O. & Goel, V. (2008) Smarter than we think: When our brains detect that we are biased. *Psychological Science* 19:483–89. [WDN, rMO]
Dick, F., Leech, R., Moses, P. & Saccuman, M. C. (2006) The interplay of learning and development in shaping neural organization. *Developmental Science* 9:14–17. [NS]
Domingos, P. & Pazzani, M. (1997) On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* 29:103–30. [HB, rMO]
Dowty, D. R., Wall, R. E. & Peters, S. (1981) *Introduction to Montague semantics*. Springer. [rMO]
Doya, K., Ishii, S., Rao, R. P. N. & Pouget, A., eds. (2007) *The Bayesian brain: Probabilistic approaches to neural coding*. MIT Press. [rMO]
Dubois, D. & Prade, H. (1988) *Possibility theory*. Plenum. [GP]
Earman, J. (1992) *Bayes or bust?* MIT Press. [arMO]
Edgington, D. (1995) On conditionals. *Mind* 104:235–329. [UH, arMO]
Elqayam, S. (2008) Normative rationality and the is-ought fallacy. In: *Proceedings of the Second European Cognitive Science Conference*, ed. S. Vosiniadou, D. Kayser & A. Protopaps, pp. 249–99. Psychology Press. [JStBTE]
Ericsson, K. A., Charness, N., Hoffmann, R. R. & Feltovich, P. J. (2006) *Cambridge handbook of expertise and expert performance*. Cambridge University Press. [NS]
Ericsson, K. A. & Kintsch, W. (1995) Long-term working memory. *Psychological Review* 102:211–45. [NS]
Ericsson, K. A. & Lehmann, A. C. (1996) Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Reviews in Psychology* 47:273–305. [NS]
Evans, J. St. B. T. (1972) Reasoning with negatives. *British Journal of Psychology* 63:213–19. [aMO]
(1977) Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology* 29A:297–306. [I-mL]
(1989) *Bias in human reasoning: Causes and consequences*. Erlbaum. [rMO]
(2002) Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin* 128:978–96. [JStBTE]
(2007) *Hypothetical thinking: Dual processes in reasoning and judgement*. Psychology Press. [JStBTE, rMO]
(2008) Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology* 59:255–78. [JStBTE]
Evans, J. St. B. T. & Frankish, K., eds. (in press) *In two minds: Dual processes and beyond*. Oxford University Press. [rMO]
Evans, J. St. B. T. & Handley, S. J. (1999) The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology* 52A:739–69. [KO, aMO]
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. [aMO]
Evans, J. St .B. T., Handley, S. J. & Over, D. E. (2003) Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29321–55. [I-mL, aMO]
Evans, J. St. B. T. & Lynch, J. S. (1973) Matching bias in the selection task. *British Journal of Psychology* 64:391–97. [aMO]
Evans, J. St. B. T., Newstead, S. E. & Byrne, R. J. (1993) *Human reasoning*. Erlbaum. [aMO]
Evans, J. St. B. T. & Over, D. E. (1996a) *Rationality and reasoning*. Psychology Press. [JStBTE, DEO, arMO]
(1996b) Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review* 103:356–63. [aMO]
(2004) *If*. Oxford University Press. [JStBTE, UH, arMO, NP]
Evans, J. St. B. T., Stevenson, R. J., Over, D. E., Handley, S. J. & Sloman, S. A. (2004) *ESRC end of grant report*, No. R00239074. The Economic and Social Research Council. [DEO]
Falk, R. & Wilkening, F. (1998) Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology* 34:1340–57. [NS]
Field, H. (1978) A note on Jeffrey conditionalization. *Philosophy of Science* 45:361–67. [rMO]
Fodor, J. A. (1968) *Psychological explanation*. Random House. [rMO]
(1983) *The modularity of mind*. MIT Press. [UH, aMO]

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Fodor, J. A. & Pylyshyn, Z. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition* 28:183–204. [UH]
Ford, M. (2004) System LS: A three tiered nonmonotonic reasoning system. *Computational Intelligence* 20:89–108. [GP]
Franks, B. (1995) On explanation in the cognitive sciences: Competence, idealization, and the failure of the classical cascade. *British Journal of the Philosophy of Science* 46(4):475–502. [WS]
(1999) Idealizations, competence and explanation: A response to Patterson. *British Journal of the Philosophy of Science* 50:735–46. [WS]
Friedman, M. (1953) *Essays in positive economics*. University of Chicago Press. [ELK]
Friedman, N., Getoor, L., Koller, D. & Pfeffer, A. (1999) Learning probabilistic relational models. In: *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI)*, ed. T. Dean, pp. 1300–309. Morgan Kaufmann. [TLG]
Friedman, N. & Halpern, J. Y. (1995) Plausibility measures: A user's guide. In: *Proceedings of the eleventh conference on uncertainty in AI*, ed. P. Besnard & S. Hanks, pp. 175–84. Morgan Kaufmann. [GP]
Friston, K. (2005) A theory of cortical responses. *Philosophical Transactions of the Royal Society B* 360:815–36. [rMO]
Galotti, K. M., Komatsu, L. K. & Voelz, S. (1997) Children's differential performance on deductive and inductive syllogisms. *Developmental Psychology* 33:70–78. [NS]
Gärdenfors, P. (1986) Belief revisions and the Ramsey test for conditionals. *Philosophical Review* 95:81–93. [rMO]
George, C. (1997) Reasoning from uncertain premises. *Thinking and Reasoning* 3:161–90. [rMO]
Geurts, B. (2003) Reasoning with quantifiers. *Cognition* 86:223–51. [aMO]
Gigerenzer, G. (1996) On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review* 103:592–96. [CRMM]
(2004) Fast and frugal heuristics: The tools of bounded rationality. In: *Blackwell handbook of judgment and decision making*, ed. D. J. Koehler & N. Harvey, pp. 62–88. Blackwell. [JStBTE]
Gigerenzer, G. & Goldstein, D. (1996) Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103:650–69. [arMO]
Gigerenzer, G. & Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102(4):684–704. [JDN, aMO]
Gigerenzer, G., Swijinck, Z., Porter, T., Daston, L., Beatty, J. & Kruger, L. (1989) *The empire of chance*. Cambridge University Press. [aMO]
Gigerenzer, G., Todd, P. & the ABC Research Group. (1999) *Simple heuristics that make us smart*. Oxford University Press. [HB, arMO]
Gilio, A. (2002) Probabilistic reasoning under coherence in System P. *Annals of Mathematics and Artificial Intelligence* 34:5–34. [NP]
Gilovich, T., Griffin, D. & Kahneman, D. (2002) *Heuristics and biases: The psychology of intuitive judgement*. Cambridge University Press. [JStBTE, CRMM]
Glymour, C. (2001) *The mind's arrow*. MIT Press. [rMO]
Goel, V. (2007) The anatomy of deduction. *Trends in Cognitive Science* 11:435–41. [rMO]
Gold, J. I. & Shadlen, M. N. (2000) Representation of a perceptual decision in developing oculomotor commands. *Nature* 404:390–94. [rMO]
Good, I. J. (1950) *Probability and the weighing of evidence*. Griffin. [JDN]
(1975) Explicativity, corroboration, and the relative odds of hypotheses. *Synthese* 30:39–73. [JDN]
Green. D. W. & Over, D. E. (1997) Causal inference, contingency tables and the selection task. *Current Psychology of Cognition* 16:459–87. [arMO]
(2000) Decision theoretical effects in testing a causal conditional. *Current Psychology of Cognition* 19:51–68. [arMO]
Gregory, R. L. (1970) *The intelligent eye*. Weidenfeld & Nicolson. [rMO]
Grice, H. P. (1975) Logic and conversation. In: *The logic of grammar*, ed. D. Davidson & G. Harman, pp. 64–75. Dickenson. [rMO]
Griffin, D. R. (1992) *Animal minds*. University of Chicago Press. [ELK]
Griffiths, T. L. & Ghahramani, Z. (2006) Infinite latent feature models and the Indian buffet process. In: *Advances in neural information processing systems, vol. 18*, ed. Y. Weiss, B. Scholkopf & J. Plaut, pp. 475–82. MIT Press. [TLG]
Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. (2007) Topics in semantic representation. *Psychological Review* 114:211–44. [rMO]
Griffiths, T. L. & Tenenbaum, J. B. (2005) Structure and strength in causal induction. *Cognitive Psychology* 51:354–84. [TLG, arMO]
Hacking, I. (1975) *The emergence of probability*. Cambridge University Press. [aMO]
(1990) *The taming of chance*. Cambridge University Press. [aMO]
Hadjichristidis, C., Sloman, S. A. & Over, D. E. (in preparation) Revision of beliefs from probabilistic categorical arguments. [DEO]
Hahn, U. & Oaksford, M. (2007) The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review* 114:704–32. [rMO]
Halford, G. S., Cowan, N. & Andrews, G. (2007) Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences* 11(6):236–42. [GSH]
Halpern, J. Y. (1997) Defining relative likelihood in partially-ordered preferential structures. *Journal of Artificial Intelligence Research* 7:1–24. [GP]
Harman, G. (1986) *Change in view: Principles of reasoning*. MIT Press. [rMO]
Hattori, M. (2002) A quantitative model of optimal data selection in Wason's selection task. *Quarterly Journal of Experimental Psychology* 55A:1241–72. [aMO]
Hawks, J., Wang, E. T., Cochran, G. M., Harpending, H. C. & Moyzis, R. K. (2007) Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences USA* 104(52):20753–58. [JDN]
Hawthorn, J. (2008) Inductive logic. In: *Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/entries/logic-inductive/. [rMO]
Hawthorne, J. & Makinson, D. (2007) The quantitative/qualitative watershed for rules of uncertain inference. *Studia Logica* 86:247–97. [NP]
Henle, M. (1978) Foreword. In: *Human reasoning*, ed. R. Revlin & R. E. Mayer. Winston. [aMO]
Hilton, D. J., Kemmelmeier, M. & Bonnefon, J-F. (2005) Putting Ifs to work: Goal-based relevance in conditional directives. *Journal of Experimental Psychology: General* 134:388–405. [rMO]
Hochberg, J. & McAlister, E. (1953) A quantitative approach to figure "goodness." *Journal of Experimental Psychology* 46:361–64. [rMO]
Hoffman, D. (in press) The interface theory of perception: Natural selection drives true perception to swift extinction. In: *Object categorization: Computer and human vision perspectives*, ed. S. Dickinson, M. Tarr, A. Leonardis & B. Schiele. Cambridge University Press. [JDN]
Hoffrage, U. (2000) Why the analyses of cognitive processes matter. *Behavioral and Brain Sciences* 23:679–80. [WDN]
Horwich, P. (1982) *Probability and evidence*. Cambridge University Press. [aMO]
Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B. & Tzourio-Mazoyer, N. (2000) Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience* 12:721–28. [WDN, rMO]
Howe, M. L. & Rabinowitz, F. M. (1996) Reasoning from memory: A lifespan inquiry into the necessity of remembering when reasoning about class inclusion. *Journal of Experimental Child Psychology* 61:1–42. [NS]
Howson, C. & Urbach, P. (1993) *Scientific reasoning: The Bayesian approach*, 2nd edition. Open Court. [DEO, aMO]
Hurley, S. & Nudds, M., eds. (2006) *Rational animals?* Oxford University Press. [rMO, ELK]
Inhelder, B. & Piaget, J. (1955) *De la logique de l'enfant à la logique de l'adolescent*. Presses Universitaires de France. (English version: *The growth of logical thinking from childhood to adolescence*. Routledge, 1958). [aMO]
Jacob F. (1977) Evolution and tinkering. *Science* 196:1161–66. [rMO]
Jacobs, R. A., Jordan, M. I., Nowlan, S. & Hinton, G. E. (1991) Adaptive mixtures of local experts. *Neural Computation* 3:1–12. [rMO]
Jeffrey, R. C. (1967) *Formal logic: Its scope and limits*, 2nd edition. McGraw-Hill. [rMO]
(1983) *The logic of decision*, 2nd edition. University of Chicago Press. [DEO, arMO]
Johnson-Laird, P. N. (1983) *Mental models*. Cambridge University Press. [arMO]
(1992) *Syllogs* (computer program). Available at: http://webscript.princeton.edu/~mentmod/models.php. [rMO]
Johnson-Laird, P. N. & Byrne, R. M. J. (1991) *Deduction*. Erlbaum. [aMO]
(2002) Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review* 109:646–78. [aMO]
Juhl, C. F. (1993) Bayesianism and reliable scientific inquiry. *Philosophy of Science* 60:302–19. [DD]
Kahneman, D., Slovic, P. & Tversky, A., eds. (1982) *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. [CRMM, aMO]
Kahneman, D. & Tversky, A. (2000) *Choices, values, and frames*. Cambridge University Press. [CRMM]
Kakade, S. & Dayan, P. (2002) Acquisition and extinction in autoshaping. *Psychological Review* 109:533–44. [aMO]
Kant, E. (1787/1961) *Critique of pure reason, First edition, second impression*, trans. N. K. Smith. Macmillan. [aMO]
Kearns, M., Mansour, Y., Ng, A. Y. & Ron, D. (1997) An experimental and theoretical comparison of model selection methods. *Machine Learning* 27:7–50. [HB]
Kemp, C. & Tenenbaum, J. B. (2008) The discovery of structural form. *Proceedings of the National Academy of Sciences USA* 105:10687–92. [rMO]
Kirby, K. N. (1994) Probabilities and utilities of fictional outcomes in Wason's four card selection task. *Cognition* 51:1–28. [aMO]
Klauer, K. C. (1999) On the normative justification for information gain in Wason's selection task. *Psychological Review* 106:215–22. [arMO]

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Klauer, K. C., Stahl, C. & Erdfelder, E. (2007) The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory and Cognition* 33:680–703. [UH]

Klayman, J. & Ha, Y.-W. (1987) Confirmation, disconfirmation, and information. *Psychological Review* 94:211–28. [JDN]

Kneale, W. & Kneale, M. (1962) *The development of logic*. Oxford University Press. [DPO]

Knill, D. & Richards, W., eds. (1996) *Perception as Bayesian inference*. Cambridge University Press. [arMO]

Koehler, D. J. & Harvey, N. (2004) *Blackwell handbook of judgement and decision making*. Blackwell. [JStBTE]

Körding, K. P. & Wolpert, D. (2004) Bayesian integration in sensorimotor learning. *Nature* 427:244–47. [rMO]

Kowalski, R. (1979) Algorithm = Logic + Control. *Communications of the Association for Computing Machinery* 22:424–36. [rMO]

Kraus, S., Lehmann, D. & Magidor, M. (1990) Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44:167–207. [NP]

Krauth, J. (1982) Formulation and experimental verification of models in propositional reasoning. *Quarterly Journal of Experimental Psychology* 34:285–98. [rMO]

Kreps, D. M. (1990) *A course in microeconomic theory*. Princeton University Press. [ELK]

Krueger, J. I. & Funder, D. C. (2004) Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences* 27(3):313–76. [CRMM]

Kuhn, T. (1962) *The structure of scientific revolutions*. University of Chicago Press. [aMO]

Kuncheva, L. I. (2006) On the optimality of Naïve Bayes with dependent binary features. *Pattern Recognition Letters* 27:830–37. [HB]

Kyllonen, P. C. & Christal, R. E. (1990) Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 14:389–433. [NS]

Lakatos, I. (1970) Falsification and the methodology of scientific research programmes. In: *Criticism and the growth of knowledge*, ed. I. Lakatos & A. Musgrave, pp. 91–196. Cambridge University Press. [aMO]

Laplace, P. S. (1951) *A philosophical essay on probabilities*, trans. F. W. Truscott & F. L. Emory. Dover. (Original work published 1814). [rMO]

Lauritzen, S. & Spiegelhalter, D. (1988) Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society B* 50:157–224. [rMO]

Leeuwenberg, E. & Boselie, F. (1988) Against the likelihood principle in visual form perception. *Psychological Review* 95:485–91. [rMO]

Levi, I. (1988) The demons of decision. *Monist* 70:193–211. [DD]

(2002) Money pumps and diachronic books. *Philosophy of Science* 69:S235–47. [DD]

Lindley, D. V. (1956) On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* 27:986–1005. [JDN, aMO]

(2006) *Understanding uncertainty*. Wiley. [NP]

Lipsey, R. G. & Lancaster, K. (1956) The general theory of second best. *Review of Economic Studies* 24:11–32. [HB]

Liu, I.-M. (2003) Conditional reasoning and conditionalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:694–709. [I-mL, rMO]

Liu, I.-M. & Chou, T. H. (2008) A developmental analysis of conditional reasoning. Unpublished manuscript, National Chung-Cheng University, Chia-Yi, Taiwan. [I-mL]

Liu, I. M., Lo, K. C. & Wu, J. T. (1996) A probabilistic interpretation of "If-Then". *The Quarterly Journal of Experimental Psychology* 49A:828–44. [I-mL, rMO]

Lukasiewicz, T. (2005) Weak nonmonotonic probabilistic logics. *Artificial Intelligence* 168:119–61. [NP]

Ma, W. J., Beck, J., Latham, P. & Pouget, A. (2006) Bayesian inference with probabilistic population codes. *Nature Neuroscience* 9:1432–38. [rMO]

MacKay, D. J. C. (2003) *Information theory, inference, and learning algorithms*. Cambridge University Press. [rMO]

Maher, P. (1992) Diachronic rationality. *Philosophy of Science* 59:120–41. [DD]

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

(1991) Social roles and utilities in reasoning with deontic conditionals. *Cognition* 39:85–105. [aMO]

Manktelow, K. I., Sutherland, E. J. & Over, D. E. (1995) Probabilistic factors in deontic reasoning. *Thinking and Reasoning* 1:201–20. [aMO]

Manning, C. & Schütze, H. (1999) *Foundations of statistical natural language processing*. MIT Press. [rMO]

Markovits, H. & Barrouillet, P. (2002) The development of conditional reasoning: A mental model account. *Developmental Review* 22:5–36. [KO, NS]

Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information*. Freeman. [ELK, CRMM, JDN, rMO]

Martignon, L. & Blackmond-Laskey, K. (1999) Bayesian benchmarks for fast and frugal heuristics. In: *Simple heuristics that make us smart*, ed. G. Gigerenzer, P. M. Todd & the ABC Research Group, pp. 169–88. Oxford University Press. [rMO]

McCarthy, J. & Hayes, P. J. (1969) Some philosophical problems from the standpoint of artificial intelligence. In: *Machine intelligence, vol. 4*, ed. B. Meltzer & D. Michie. Edinburgh University Press. [aMO]

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

McKenzie, C. R. M. (2003) Rational models as theories – not standards – of behavior. *Trends in Cognitive Sciences* 7(9):403–406. [JDN]

(2004) Framing effects in inference tasks – and why they are normatively defensible. *Memory and Cognition* 32:874–85. [CRMM]

McKenzie, C. R. M. & Amin, M. B. (2002) When wrong predictions provide more support than right ones. *Psychonomic Bulletin and Review* 9:821–28. [CRMM]

McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J. & Skrable, R. P. (2001) Do conditional statements target rare events? *Organizational Behavior and Human Decision Processes* 85:291–309. [CRMM, arMO]

McKenzie, C. R. M. & Mikkelsen, L. A. (2000) The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review* 7:360–66. [CRMM, arMO]

(2007) A Bayesian view of covariation assessment. *Cognitive Psychology* 54:33–61. [CRMM, arMO]

McKenzie, C. R. M. & Nelson, J. D. (2003) What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review* 10:596–602. [CRMM]

Milch, B., Marthi, B. & Russell, S. (2004) BLOG: Relational modeling with unknown objects. In: *ICML 2004 workshop on statistical relational learning and its connections to other fields, Banff, Alberta, Canada*, ed. T. Dietterich, L. Getoor & K. Murphy, pp. 67–73. Available at: www.cs.umd.edu/projects/srl2004/srl2004_complete.pdf [TLG]

Milne, P. (1995) A Bayesian defence of Popperian science? *Analysis* 55:213–15. [rMO]

(1996) log[P(h|eb)/P(h|b)] is the one true measure of confirmation. *Philosophy of Science* 63:21–26. [rMO]

Moore, G. E. (1903) *Principia ethica*. Cambridge University Press. [rMO]

Murphree, W. A. (1991) *Numerically exceptive logic: A reduction of the classical syllogism*. Peter Lang. [NP]

Navarro, D. J., Griffiths, T. L., Steyvers, M. & Lee, M. D. (2006) Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology* 50:101–22. [rMO, NS]

Nelson, J. D. (2005) Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review* 112(4):979–99. [JDN, arMO]

(2008) Towards a rational theory of human information acquisition. In: *The probabilistic mind: Prospects for rational models of cognition*, ed. M. Oaksford & N. Chater, pp. 143–63. Oxford University Press. [JDN]

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W. & Sejnowski, T. J. (submitted) Experience matters: Information acquisition optimizes probability gain. [JDN]

Newell, A., Shaw, J. C. & Simon, H. A. (1958) Chess-playing programs and the problem of complexity. *IBM Journal of Research and Development* 2:320–25. [aMO]

Newell, A. & Simon, H. A. (1972) *Human problem solving*. Prentice-Hall. [aMO]

Newstead, S. E., Handley, S. J. & Buck, E. (1999) Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory and Cognition* 27:344–54. [aMO]

Nickerson, R. S. (1996) Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning* 2:1–32. [JDN, aMO]

Novick, L. R. & Cheng, P. W. (2004) Assessing interactive causal influence. *Psychological Review* 111:455–85. [aMO]

Oaksford, M. (2004a) Conditional inference and constraint satisfaction: Reconciling probabilistic and mental models approaches? Paper presented at the 5th International Conference on Thinking, July 22–24, 2004. University of Leuven, Leuven, Belgium. [aMO]

(2004b) Reasoning. In: *Cognitive psychology*, ed. N. Braisby & A. Gellatly, pp. 418–55. Oxford University Press. [aMO]

Oaksford, M. & Brown, G. D. A., eds. (1994) *Neurodynamics and psychology*. Academic Press. [UH]

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

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

(1994) A rational analysis of the selection task as optimal data selection. *Psychological Review* 101:608–31. [JDN, arMO, UH]
(1996) Rational explanation of the selection task. *Psychological Review* 103:381–91. [aMO]
(1998a) *Rationality in an uncertain world*. Psychology Press. [aMO, WS]
eds. (1998b) *Rational models of cognition*. Oxford University Press. [UH, arMO]
(2001) The probabilistic approach to human reasoning. *Trends in Cognitive Sciences* 5:349–57. [E-JW]
(2002) Common sense reasoning, logic and human rationality. In: *Common sense, reasoning and rationality*, ed. R. Elio, pp. 174–214. Oxford University Press. [rMO]
(2003a) Conditional probability and the cognitive science of conditional reasoning. *Mind and Language* 18:359–79. [aMO, DPO]
(2003b) Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin and Review* 10:289–318. [aMO, E-JW]
(2007) *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press. [NA, HB, DD, WDN, JStBTE, TLG, UH, GSH, ELK, I-mL, CRMM, JDN, KO, arMO, DPO, DEO, NP, FHP, GP. KS, NS, WS, E-JW]
(2008) Probability logic and the *Modus Ponens–Modus Tollens* asymmetry in conditional inference. In: *The probabilistic mind: Prospects for Bayesian cognitive science*, ed. N. Chater & M. Oaksford, pp. 97–120. Oxford University Press. [arMO]
(in press) Conditionals and constraint satisfaction: Reconciling mental models and probabilistic approaches? In: *Cognition and conditionals: Probability and logic in human thought*, ed. M. Oaksford & N. Chater. Oxford University Press. [rMO]
Oaksford, M., Chater, N. & Grainger, B. (1999) Probabilistic effects in data selection. *Thinking and Reasoning* 5:193–244. [arMO]
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*, 23:441–58. [aMO]
Oaksford, M., Chater, N. & Larkin, J. (2000) Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26:883–89. [KO, arMO]
Oaksford, M. & Hahn, U. (2007) Induction, deduction and argument strength in human reasoning and argumentation. In: *Inductive reasoning*, ed. A. Feeney & E. Heit. pp. 269–301. Cambridge University Press. [rMO]
Oaksford, M. & Moussakowski, M. (2004) Negations and natural sampling in data selection: Ecological vs. heuristic explanations of matching bias. *Memory and Cognition* 32:570–81. [arMO]
Oaksford, M., Roberts, L. & Chater, N. (2002) Relative informativeness of quantifiers used in syllogistic reasoning. *Memory and Cognition* 30:138–49. [aMO]
Oaksford, M. & Stenning, K. (1992) Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory and Cognition* 18:835–54. [aMO]
Oaksford, M. & Wakefield, M. (2003) Data selection and natural sampling: Probabilities do matter. *Memory and Cognition* 31:143–54. [arMO]
Oberauer, K. (2006) Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology* 53:238–83. [UH, KO, rMO]
Oberauer, K., Weidenfeld, A. & Fischer, K. (2007) What makes us believe a conditional? The roles of covariation and causality. *Thinking and Reasoning* 13:340–69. [KO]
Oberauer, K., Weidenfeld, A. & Hörnig, R. (2004) Logical reasoning and probabilities: A comprehensive test of Oaksford and Chater (2001) *Psychonomic Bulletin and Review* 11:521–27. [KO, arMO]
Oberauer, K. & Wilhelm, O. (2003) The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:680–93. [I-mL, aMO]
Oberauer, K., Wilhelm, O. & Diaz, R. R. (1999) Bayesian rationality for the Wason selection task? A test of optimal data selection theory. *Thinking and Reasoning* 5:115–44. [aMO, E-JW]
O'Brien, D. P. (1993) Mental logic and irrationality: We can put a man on the moon, so why can't we solve those logical reasoning problems? In: *Rationality: Psychological and philosophical perspectives*, ed. K. I. Manktelow & D. E. Over. Routledge. [DPO]
(2004) Mental-logic theory: What it proposes, and reasons to take this proposal seriously. In: *The nature of reasoning*, ed. J. Leighton & R. J. Sternberg, pp. 205–33. Cambridge University Press. [DPO]
O'Brien, D. P. & Manfrinati, A. (in press) The mental logic theory of conditional propositions. In: *The psychology of conditionals*, ed. M. Oaksford & N. Chater. Oxford University Press. [DPO]
Osherson, D. N., Stob, M. & Weinstein, S. (1988) Mechanical learners pay a price for Bayesianism. *Journal of Symbolic Logic* 53:1245–51. [DD]
Over, D. E. & Evans, J. St. B. T (1994) Hits and misses: Kirby on the selection task. *Cognition* 52:235–43. [aMO]
Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J. & Sloman, S. A. (2007) The probability of causal conditionals. *Cognitive Psychology* 54:62–97. [I-mL, DEO, rMO]
Over, D. E. & Jessop, A. (1998) Rational analysis of causal conditionals and the selection task. In: *Rational models of cognition*, ed. M. Oaksford & N. Chater, pp. 399–414. Oxford University Press. [aMO]
Pearl, J. (1988) *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann. [arMO]
(2000) *Causality: Models, reasoning and inference*. Cambridge University Press. [arMO]
Perham, N. & Oaksford, M. (2005) Deontic reasoning with emotional content: Evolutionary psychology or decision theory? *Cognitive Science* 29:681–718. [aMO]
Peters, S. & Westerståhl, D. (2006) *Quantifiers in language and logic*. Oxford University Press. [NP]
Peterson, P. L. (2000) *Intermediate quantifiers. Logic, linguistics, and Aristotelian semantics*. Ashgate. [NP]
Pfeifer, N. & Kleiter, G. D. (2005a) Coherence and nonmonotonicity in human reasoning. *Synthese* 146(1–2):93–109. [GP, NP]
(2005b) Towards a mental probability logic. *Psychologica Belgica* 45(1):71–99. Updated version available at: http://www.users.sbg.ac.at/~pfeifern/problog.htm [NP]
(2006) Inference in conditional probability logic. *Kybernetika* 42:391–404. [NP]
(in press a) Framing human inference by coherence based probability logic. *Journal of Applied Logics*. [NP]
(in press b) The conditional in mental probability logic. In: *Cognition and conditionals: Probability and logic in human thoughts*, ed. M. Oaksford & N. Chater. Oxford University Press. [NP]
Piaget, J. (1957) *Logic and psychology*. Basic Books. [GSH]
Piaget, J. & Inhelder, B. (1975) *The origin of the idea of chance in children*. W. W. Norton. [NS]
Pitt, M. A., Myung, I. J. & Zhang, S. (2002) Toward a method of selecting among computational models of cognition. *Psychological Review* 109:472–91. [HB]
Poletiek, F. H. (1996) Paradoxes of falsification. *Quarterly Journal of Experimental Psychology* 49A:447–62. [FHP]
(2001) *Hypothesis-testing behavior*. Psychology Press. [FHP]
Poletiek, F. H. & Berndsen, M. (2000) Hypothesis testing as risk behavior with regard to beliefs. *Journal of Behavioral Decision Making* 13:107–23. [FHP]
Politzer, G. & Braine, M. D. (1991) Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition* 38:103–08. [aMO]
Popper, K. R. (1935/1959) *The logic of scientific discovery*. Basic Books. [aMO]
(1963/1978) *Conjectures and refutations*, 4th edition. Routledge and Kegan Paul. [FHP]
Putnam, H. (1974) The "corroboration" of theories. In: *The philosophy of Karl Popper, vol. 2*, ed. A. Schilpp. Open Court. [aMO]
Pylyshyn, Z., ed. (1987) *The robot's dilemma: The frame problem in artificial intelligence*. Ablex. [aMO]
Quine, W. V. O. (1953) *From a logical point of view*. Harvard University Press. [aMO]
Raby, C. R., Alexis, D. M., Dickinson, A. & Clayton, N. S. (2007) Planning for the future by Western scrub-jays. *Nature* 445:919–21. [ELK]
Ramsey, F. P. (1931/1990a) General propositions and causality. In: *Philosophical papers*, ed. D. H. Mellor, pp. 145–63. Cambridge University Press. [I-mL]
(1931/1990b) *The foundations of mathematics and other logical essays*. Routledge and Kegan Paul. [aMO]
Rao, R. P. N., Olshausen, B. A. & Lewicki, M. S., eds. (2002) *Probabilistic models of the brain: Perception and neural function*. MIT Press. [rMO]
Reichenbach, H. (1938) *Experience and prediction*. University of Chicago Press. [TLG]
Reiter, R. (1980) A logic for default reasoning, *Artificial Intelligence* 13:81–132. [aMO]
Restall, G. (1996) Information flow and relevant logics. In: *Logic, language and computation*, ed. J. Seligman, & D. Westerståhl, pp. 463–78 CSLI Publications. [rMO]
Rips, L. J. (1983) Cognitive processes in propositional reasoning. *Psychological Review* 90:38–71. [aMO]
(1994) *The psychology of proof*. MIT Press. [arMO]
(2002) Reasoning imperialism. In: *Common sense, reasoning, and rationality*, ed. R. Elio, pp. 215–35. Oxford University Press. [WS]
Rissanen, J. J. (1989) *Stochastic complexity and statistical inquiry*. World Scientific. [rMO]
Rosch, E. (1975) Cognitive representation of semantic categories. *Journal of experimental psychology: General* 104:192–233. [aMO]
Rosser, J. B. (1939) An informal exposition of proofs of Gödel's theorem and Church's theorem. *Journal of Symbolic Logic* 4:53–60. [DPO]

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press

Sanborn, A. N., Griffiths, T. L. & Navarro, D. J. (2006) A more rational model of categorization. In: *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Erlbaum. [TLG]

Savage, L. J. (1972) *Foundations of statistics*. Dover. [DD]

Schroyens, W. (in press) Logic and/in psychology: The paradoxes of material implication and psychologism in the cognitive science of human reasoning. In: *Cognition and conditionals: Probability and logic in human thoughts*, eds. M. Oaksford & N. Chater. Oxford University Press. [WS]

Schroyens, W. & Schaeken, W. (2003) A critique of Oaksford, Chater and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29:140–49. [UH, aMO]

Schulte, O. (1999) The logic of reliable and efficient inquiry. *Journal of Philosophical Logic* 28:399–438. [DD]

Shafer, G. (1976) *A mathematical theory of evidence*. Princeton University Press. [rMO, GP]

Shannon, C. E. & Weaver, W. (1949) *The mathematical theory of communication*. University of Illinois Press. [rMO]

Shepard, R. N. (1987) Towards a universal law of generalization for psychological science. *Science* 237:1317–23. [TLG]

(1995) Mental universals: Toward a twenty-first century science of mind. In: *The science of the mind: 2001 and beyond*, ed. R. L. Solso & D. W. Massaro, pp. 50–62. Oxford University Press. [TLG]

Sher, S. & McKenzie, C. R. M. (2006) Information leakage from logically equivalent frames. *Cognition* 101:467–94. [CRMM, rMO]

(2008) Framing effects and rationality. In: *The probabilistic mind: Prospects for Bayesian cognitive science*, ed. N. Chater & M. Oaksford, pp. 79–96. Oxford University Press. [CRMM]

Shi, L., Feldman, N. & Griffiths, T. L. (2008) Performing Bayesian inference with exemplar models. In: *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*, ed. B. Love, K. McRae & V. Sloutsky, pp. 745–50. Cognitive Science Society. [TLG]

Shultz, T. R. (2007) The Bayesian revolution approaches psychological development. *Developmental Science* 10:357–64. [NS]

Skov, R. B. & Sherman, S. J. (1986) Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology* 22:93–121. [JDN]

Sloman, S. A. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [rMO]

(2005) *Causal models*. Oxford University Press. [rMO]

Slowiaczek, L. M., Klayman, J., Sherman, S. J. & Skov, R. B. (1992) Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory and Cognition* 20:392–405. [JDN]

Smolensky, P. (1990) Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artificial Intelligence* 46:159–216. [UH]

Sobel, D. M., Tenenbaum, J. B. & Gopnik, A. (2004) Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science* 28:303–33. [rMO, NS]

Sobel, J. H. (2004) Probable modus ponens and modus tollens and updating on uncertain evidence. Unpublished manuscript, Department of Philosophy, University of Toronto, Scarborough. Available at: www.scar.toronto.ca/~sobel/Conf/Disconf.pdf. [rMO]

Sober, E. (2002) Intelligent design and probability reasoning. *International Journal for Philosophy of Religion* 52:65–80. [arMO]

Stanovich, K. E. (1999) *Who is rational? Studies of individual differences in reasoning*. Elrbaum. [JStBTE]

(2008) Individual differences in reasoning and the algorithmic/intentional level distinction in cognitive science. In: *Reasoning: Studies of human inference and its foundations*, ed. L. Rips & J. Adler, pp. 414–36. Cambridge University Press. [rMO]

Stanovich, K. E. & West, R. F. (2000) Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23:645–65. [rMO]

Stenning, K. & van Lambalgen, M. (2005) Semantic interpretation as reasoning in non-monotonic logic: The real meaning of the suppression task. *Cognitive Science* 29(6):919–60. [KS]

(2008a) *Human reasoning and cognitive science*. MIT Press. [KS]

(2008b) Logic in a noisy world. In: *Cognition and conditionals: Probability and logic in human thought*, eds. M. Oaksford & N. Chater. Oxford University Press. [KS]

Sternberg, R. J. (1999) Intelligence as developing expertise. *Contemporary Educational Psychology* 24:359–75. [NS]

Stevenson, R. J. & Over, D. E. (1995) Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology* 48A:613–43. [DEO, rMO]

(2001) Reasoning from uncertain premises. Effects of expertise and conversational context. *Thinking and Reasoning* 7:367–90. [DEO]

Stewart, N., Chater, N. & Brown, G. D. A. (2006) Decision by sampling. *Cognitive Psychology* 53:1–26. [arMO]

Taplin, J. E. (1971) Reasoning with conditional sentences. *Journal of Verbal Learning and Verbal Behavior* 10:219–25. [I-mL]

Taplin, J. E. & Staudenmayer, H. (1973) Interpretation of abstract conditional sentences in deductive reasoning. *Journal of Verbal Learning and Verbal Behavior* 12:530–42. [I-mL]

Teller, P. (1973) Conditionalization and observation. *Synthese* 26:218–38. [DD]

Tenenbaum, J. B. (1999) A Bayesian framework for concept learning. Doctoral dissertation, Brain and Cognitive Sciences Department, MIT. [rMO]

Tenenbaum, J. B. & Griffths, T. L. (2001) Structure learning in human causal induction. In: *Advances in neural information processing systems, vol. 13*, ed. T. K. Keen, T. G. Dietterich & V. Tresp, pp. 59–65. MIT Press. [rMO]

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science* 10:309–18. [TLG]

Tenenbaum, J. B., Kemp, C. & Shafto, P. (2007) Theory based Bayesian models of inductive reasoning. In: *Inducive reasoning*, ed. A. Feeney & E. Heit, pp. 167–204. Oxford University Press. [UH, rMO]

Thaler, R. H. (2005) *Advances in behavioral finance, Vol. II*. Princeton University Press. [rMO]

Todd, P. M. & Gigerenzer, G. (2003) Bounding rationality to the world. *Journal of Economic Psychology* 24:143–65. [HB]

Toussaint, M., Harmeling, S. & Storkey, A. (2006) Probabilistic inference for solving (PO)MDPs. Technical Report EDI-INF-RR-0934, University of Edinburgh. [rMO]

van Fraassen, B. C. (1980) *The scientific image*. Oxford University Press. [ELK]

Verhaeghen, P. & Salthouse, T. A. (1997) Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin* 122:231–49. [NS]

Vermeij, G. J. (2004) *Nature: An economic history*. Princeton University Press. [ELK]

Viswanathan, G. M., Buldyrev, S. V., Havlin, S., da Luz, M. G., Raposo, E. P. & Stanley, H. E. (1999) Optimizing the success of random searches. *Nature* 401:911–14. [JDN]

Vitányi, P. M. B. & Li, M. (2000) Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Transactions on Information Theory* IT-46446–64. [rMO]

von Helmholtz, H.(1910/1925) *Physiological optics. Volume III. The theory of the perception of vision*. Dover. (Translated from 3rd German edition, 1910). [rMO]

Wagner, C. G. (2004) Modus tollens probabilized. *British Journal for Philosophy of Science* 55:747–53. [aMO]

Walton, D. N. (1989) *Informal logic*. Cambridge University Press. [rMO]

Wang, Z.-J. (1999) A two-component model of conditional reasoning. Unpublished Master's Thesis, National Chung-Cheng University, Chia-Yi, Taiwan. [I-mL]

Wells, G. L. & Lindsay, R. C. L. (1980) On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin* 88:776–84. [JDN]

Williamson, J. & Gabbay, D., eds. (2003) Special issue on combining logic and probability. *Journal of Applied Logic* 1:135–308. [rMO]

Yama, H. (2001) Matching versus optimal data selection in the Wason selection task. *Thinking and Reasoning* 7:295–311. [aMO]

Yuille, A. L. & Kersten, D. (2006) Vision as Bayesian Inference: Analysis by Synthesis? *Trends in Cognitive Sciences* 10:301–08. [aMO]

Zadeh, L. A. (1975) Fuzzy logic and approximate reasoning. *Synthese* 30:407–28. [rMO]

Zeelenberg, M., Van Dijk, W. W., Manstead, A. S. R. & Van der Pligt, J. (2000) On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment. *Cognition and Emotion* 14:521–41. [aMO]

https://doi.org/10.1017/S0140525X09000387 Published online by Cambridge University Press