4.5 Bayesian Rationality in the Psychology of Reasoning

Nick Chater and Mike Oaksford

Summary

Bayesian models have become widespread throughout the cognitive and brain sciences, as providing natural ways to understand how the brain deals with uncertainty in perception, belief updating, decision making, and motor control. In this chapter, we ask how far the Bayesian approach helps psychologists understand perhaps the most direct expression of human rationality: explicit verbal reasoning. Historically, verbal reasoning has been modeled in the framework of mathematical logic; the "new paradigm" in the psychology of reasoning developed over the past 25 years has suggested that human reasoning and argumentation may often be better understood using probabilistic Bayesian methods.

1. The Psychology of Verbal Reasoning: A Brief History

Following in the footsteps of 20th-century analytic philosophy, the starting point for the psychology of verbal reasoning in the 1960s was the assumption that the standards of formal, binary logic should judge human reasoning.[1] Accordingly, researchers focused on how people reason with the so-called logical terms of language, most notably those that have natural counterparts in first-order logic: *and*, *or*, *not*, *if–then*, *all*, *some*. In a typical experiment, people would be given one or more premises (e.g., *Birds fly*, *Tweety is a bird*) and asked to generate, or evaluate, possible conclusions that may follow (e.g., *Tweety flies*). The assessment of specific patterns of logical reasoning was seen as central to a much wider project within cognitive science and artificial intelligence: viewing cognition as a whole as operating through logical reasoning over world knowledge encoded in an internal, logical, language of thought—a program encapsulated in the phrase "cognition is proof theory" (Fodor & Pylyshyn, 1988).

Since the mid-1990s, however, the psychology of verbal reasoning has begun to adopt what has become known as the "new paradigm" (Over, 2009), in which Bayesian probability theory, rather than logic, has been taken as the standard against which human performance should be compared (Oaksford & Chater, 1994). Moreover, the shift to a Bayesian viewpoint has also been associated with two connected changes: a focus on knowledge-rich real-world inferences, rather than narrow logical puzzles, and viewing verbal reasoning not as an isolated, internal cognitive process but as having an inherently social function, in argument and persuasion.

We begin with a brief sketch of logic-based accounts of reasoning and its difficulties before turning to probabilistic accounts of deduction, induction, abduction, and argumentation, concluding with brief reflections on how Bayesian approaches to verbal reasoning may connect with the wider program of Bayesian cognitive science.[2]

2. Logic-Based Approaches to Reasoning

In line with early research in the psychology of reasoning, let us take logic as our starting point and consider the modus ponens (MP) inference (see sentence display 1 below).

In classical logic, this inference is of course *valid*: the truth of the premises guarantees the truth of the conclusion, and it can readily be converted into an experiment: participants are instructed to assume the premises are true (whether they believe them or not) and to draw inferences from those premises alone. Notice that, from a Bayesian point of view, such reasoning may be highly unnatural: our subjective probabilities will, after all, typically be influenced by the entirety of our background

(1)	If I turn the hot tap, I get hot water	If p then q	(A)
	I turn the hot tap	p	(B)
Therefore:	I get hot water	q	(MP: A, B)

knowledge—we shall return to the implications of this observation below.

According to logic-based psychological approaches, verbal reasoning follows logical lines: either directly, through implementing a psychological version of the logician's proof theory (a putative mental logic; Rips, 1994; see chapter 3.3 by Khemlani, in this handbook), or indirectly, through implementing a psychological version of the logician's model theory (mental model theory [MMT], Johnson-Laird, 1983; see chapter 2.3 by Johnson-Laird, this handbook, and, for a recent critique of MMT, Oaksford, Over, & Cruz, 2019). But the appropriateness of any logic-based approach is thrown into question by the following observation. Suppose a friend says *I didn't get hot water this morning*; by the logical law of modus tollens (MT), we can infer *You didn't turn the hot tap*. But this is precisely the opposite of the correct conclusion: our friend is telling us that she didn't get hot water *despite turning the hot tap* (otherwise, the nonappearance of hot water would be too uninteresting to mention). So, in real discourse, we confidently infer that the hot tap was turned, in precise contradiction to the apparent recommendations of logic.

To explain how we reason in such cases requires recognizing, presumably using general principles of conversation (e.g., Grice, 1975; Sperber & Wilson, 1986), that the speaker is aiming to be relevant and informative. Moreover, what counts as relevant and informative depends not just on the given premises but on background knowledge (e.g., that a lack of hot water is sufficiently bothersome to be worth mentioning).

Note, too, that this example illustrates how real-world inferences routinely violate a fundamental law of classical treatments of the conditional. According to Strengthening of the Antecedent, *If p then q* implies *If p and r then q* for any *r*. Yet *If I turn the hot tap, I get hot water* does not imply *If I turn the hot tap and the heating is broken, I get hot water*. Indeed, quite the opposite! Unlike logical reasoning, everyday reasoning is *nonmonotonic*: conclusions can be overturned if more information is added (e.g., Oaksford & Chater, 1991).

3. Deductive Reasoning with Propositions

Patterns of reasoning about the logical connectives, *and*, *or*, *not*, and, in particular, *if–then*, have been intensively studied experimentally and theoretically. One approach is probability logic (see chapter 4.4 by Pfeifer, this handbook), which seeks to generalize standard logic by requiring that reasoners respect not merely logical consistency but also probabilistic coherence: degrees of belief associated with propositions should be consistent with probability theory (Coletti & Scozzafava, 2002; Pfeifer & Kleiter, 2009). Moreover, researchers in the new paradigm have explored a variety of related formal ideas ranging from so-called p-validity, Bayesian and Jeffrey conditionalization, and dynamic inference.

We begin by illustrating how probabilistic coherence applies to a generalization of the MP inference illustrated in (1), above:

(2)		
	If p then q	$\Pr(q \mid p) = a$
	p	$\Pr(p) = b$
Therefore:	q	$\Pr(q) = [ab, ab + (1 - b)]$

Now, we no longer suppose that the premises are definitely true. Instead, each premise is assigned a probability (a and b, respectively), reflecting their degree of belief. Probabilistic coherence demands that the degree of belief in the conclusion must lie in the interval between (and including) the values ab and $ab + (1 - b)$.[3]

What exactly does it mean to talk about the probability of a conditional Pr(if p then q)? The simplest approach is to equate this with the conditional probability $\Pr(q \mid p)$, an identification known as "the Equation" (Edgington, 1995). The conditional probability itself will depend on arbitrary background knowledge, using what is known as the Ramsey test, proposed by philosopher and mathematician Frank Ramsey (1929/1990). We add the antecedent (e.g., *The hot tap was turned*) of the conditional to our entire set of beliefs, adjust them to take account of this new information, and read off the probability of the conclusion (*There is hot water*) from this new set of beliefs. Notice, in particular, that from this point of view, reasoning depends not just on the given premises but on world knowledge in general.

Is the Equation appropriate from a psychological point of view? It has been confirmed in a large number of experiments (e.g., Evans, Handley, & Over, 2003; Oberauer & Wilhelm, 2003; Politzer, Over, & Baratgin, 2010), and experimental participants reject a crucial step[4] in philosophical arguments (Lewis, 1976) that appear to undermine the Equation by suggesting that it leads to an unacceptable triviality result (Douven & Verbrugge, 2013). Notice, too, that a conventional material-implication interpretation of the conditional (that *If p then q* is equivalent to *Not-p or q*) makes very different predictions, which are strongly at odds with participants' judgments.

How does probability coherence fare in explaining empirical data about how people reason with conditionals? Varying a and b, in the MP inference above, is seen reliably to yield probabilities of the conclusion that fall within the predicted coherence interval at above-chance

levels (Cruz, Baratgin, Oaksford, & Over, 2015; Evans, Thompson, & Over, 2015; Pfeifer & Kleiter, 2009; Politzer & Baratgin, 2016; Singmann, Klauer, & Over, 2014). This is not true for all inference types, however, including modus tollens (if *p* then *q*; not-*q*; therefore not-*p*). Cruz et al. (2015) argue that this may stem from the complexity of the inference and show that probabilistic coherence is respected above chance for eight simple one-premise arguments (e.g., *p* or *q*; therefore if not-*p* then *q*).

Just as logical consistency has a parallel in probabilistic coherence, so logical validity has a corresponding probabilistic analogue, p-validity (Adams, 1998; Coletti & Scozzafava, 2002; Pfeifer & Kleiter, 2009). A p-valid argument is one for which it is not probabilistically coherent that the uncertainty of the conclusion (1 – Pr(*conclusion*)) exceeds the sum of the uncertainties of the premises ($\sum_{i=1}^{n}(1 - \Pr(premise_i))$).

For example, let us reconsider (2). Suppose, for example, $a = .9$ and $b = .5$, so that the sum of the premise uncertainties is $.1 + .5 = .6$. According to probabilistic coherence, the probability of the conclusion is in the interval [.45, .95], with corresponding uncertainties, therefore, ranging from .55 down to .05. So it is not probabilistically coherent that the uncertainty of the conclusion (in the interval [.05, .55]) exceeds the sum of the premise uncertainties, .6. Hence, the inference is p-valid (and, indeed, this will be true for MP quite generally).

P-validity mitigates all the more puzzling paradoxes that arise from the "material-conditional" (or what logicians prefer to call "material-implication") interpretation of *If p then q* used in standard first-order logic: that, as noted above, is equivalent to *Not-p or q*. According to this interpretation, numerous apparent paradoxes arise (see chapter 5.2 by Rott and chapter 6.1 by Starr, both in this handbook). For example, the premise *There is no life on Mars* appears to imply *If there is life on Mars, there are kangaroos on the moon*. Such inferences are not, though, p-valid (and, indeed, people do not endorse them; see Pfeifer & Kleiter, 2011).

Despite its theoretical interest, however, p-validity does not appear directly to conform to human reasoning judgments. Indeed, from a psychological point of view, reasoning seems oriented to determining how one should change one's beliefs in the light of new evidence and background knowledge (Harman, 1986), rather than determining the validity of arguments, whether validity is viewed from a logical or a probabilistic standpoint.

So far, we have considered the situation in which both premises are potentially uncertain. But what happens if one premise is learned to be true for certain? For example, suppose I am certain that you turned the hot tap, because I saw you do it or because I know you to be completely truthful. Then, $\Pr(p) = 1$ in (8), and the probability of the conclusion can directly be read off from the conditional probability of the conclusion, *q*, given *p*: $\Pr(q \mid p)$. This is an example of what is known as Bayesian conditionalization, and it yields a precise probability rather than an interval. This approach (combined with background assumptions concerning the probability of *q* given not-*p*) successfully captured early empirical data (Oaksford & Chater, 2007; Oaksford, Chater, & Larkin, 2000) on how people reason with MP, MT, and two logically invalid patterns of reasoning with conditionals (Denying the Antecedent [DA], where not-*p* is taken to imply not-*q*, and Affirming the Consequent [AC], where *q* is taken to imply *p*).

But suppose we are not completely certain that the hot water tap is turned on (i.e., what if $\Pr(p) < 1$)? A natural assumption from prior knowledge is that most taps, most of the time, are not being turned: tap-turning events are rare. So given evidence that the tap was turned may considerably raise an initial estimate, say, $\Pr_0(p) = .01$, to $\Pr_1(p) = .9$. In such cases, Jeffrey conditionalization (Jeffrey, 2004) provides a more general updating procedure (see chapter 4.1 by Hájek & Staffel, this handbook):

$$(10)\quad \Pr_1(q) = \Pr_0(q \mid p)\Pr_1(p) + \Pr_0(q \mid \neg p)(1 - \Pr_1(p)).$$

Here, then, the probability of the conclusion $\Pr_1(q)$ will depend on prior assumptions about $\Pr_0(q \mid \neg p)$, so that belief updating depends on background knowledge even for modus ponens. Zhao and Osherson (2010; see also Hadjichristidis, Sloman, & Over, 2014) asked people to estimate the relevant probabilities and found that the probability of the conclusion aligns well with Jeffrey's rule.

We have so far assumed that changing the probability of one premise does not modify the probabilities of others. But this will not, of course, always be true, and in such cases, known as *dynamic inference*, reasoning is less straightforward (Adams, 1998; Gilio & Over, 2012; Oaksford & Chater, 2007, 2013).

Recall the example we outlined earlier, where we are informed *You didn't get hot water this morning*, but learning this will drastically change our belief in the conditional premise *If you turn the hot tap, you get hot water*. If it were not revised, we would simply conclude *You didn't turn the hot tap*. But the appropriate conclusion is that the probability of the conditional should drastically be revised downward: while heating systems normally work, learning *You didn't get hot water this morning* leads to the strong suspicion that your heating system is malfunctioning.

Such cases, where other premises modify, perhaps drastically, the conditional probability of the conditional premise, are said to violate invariance or rigidity between the initial (Pr_0) and the revised distribution (Pr_1): such violations imply that $Pr_1(q \mid p) \neq Pr_0(q \mid p)$. They are sometimes, but not always, observed (Zhao & Osherson, 2010), and allowing violations for MT (as in our example), as well as for AC and DA, improves fits to empirical data on conditional reasoning (Oaksford & Chater, 2007, 2013).

4. Deductive Reasoning with Quantifiers

So far, we have considered how people reason about combinations of, and relations between, propositions *p*, *q*, and so on. Researchers have also studied how people reason with the quantifiers *all*, *some*, *some-not*, and *none*, as well as with the so-called generalized quantifiers, *most* and *few*. For example, consider the syllogism

(3) All beekeepers (*B*) are artists (*A*)
Some chemists (*C*) are beekeepers
Therefore: Some chemists are artists

How can quantified inference be viewed in probabilistic terms? The Probability Heuristics Model (PHM; Chater & Oaksford, 1999) takes a direct approach, viewing quantified assertions as expressing probabilistic constraints: *All*: $\Pr(y \mid x) = 1$, *Some*: $\Pr(y \mid x) > 0$, *Some-not*: $\Pr(y \mid x) < 1$, *None*: $\Pr(y \mid x) = 0$, *Few*: $0 < \Pr(y \mid x) < \Delta$, *Most*: $1 - \Delta < \Pr(y \mid x) < 1$, for small Δ.

How these probabilities combine depends on the structural relationships between the premises—what is known in syllogistic logic as "figure" (i.e., the relative positions of the end terms, *A* and *C*, with respect to the middle term that links them, *B*). Each syllogistic figure can be represented by a dependency graph. For example, (3) has the following structure: *Chemists* → *beekeepers* → *artists*. The premises constrain the probabilities associated with the links in the graph. Here, for example, *Some chemists are beekeepers*, which applies to the first link in the graph and embodies the constraint that $\Pr(beekeeper \mid chemist) > 0$. Similarly, turning to the second link, *All beekeepers are artists*, corresponds to $\Pr(artist \mid beekeeper) = 1$. It is then possible to apply the notion of p-validity and to prove whether the conclusion *Some chemists are artists* follows (which in this case it does).

PHM does not predict reasoning behavior directly from p-validity, however (indeed, as we noted above, it seems that validity, whether probabilistic or logical, is not computed by reasoners, who are concerned instead with belief updating). Instead, PHM applies simple heuristics that reflect p-validity indirectly. The *min*-heuristic is defined over an "informativeness" ordering of the various quantified statements, such that *all* > *most* > *few* > *some* > *none* > *some-not* (based on Shannon surprisal: $I(s) = 1/\Pr(s)$; Shannon & Weaver, 1949). This ordering is justified by assuming *rarity*: most predicates apply only to small subsets of entities (i.e., there are far more nonartists than artists). The *min*-heuristic is that the quantifier in the conclusion should match the *least* informative quantifier in either of the premises. Thus, in (3), the conclusion should use the quantifier *some*. The *max*-heuristic is that the degree of confidence in the conclusion should depend on the expected informativeness of the quantifier in the *most* informative premises (here, *all*). Expected informativeness is ordered: *all* > *most* > *few* > *some* > *some-not* ≈ *none*, where the expectation is taken over the conclusions of the p-valid syllogisms for which that quantifier is the most informative premise (so p-validity enters indirectly into the analysis at this point).

PHM fits prior experiments using logical quantifiers that have been modeled using mental logic (Rips, 1994), with fewer parameters (Chater & Oaksford, 1999). Moreover, extending the approach to syllogisms with *most* and *few* led to data that corroborate the PHM account; these data remain outside the scope of any current logic-based account (Chater & Oaksford, 1999). Mental model theory (Johnson-Laird, 1983) provides an equally good account of experimental results that link memory span to syllogistic reasoning (Copeland & Radvansky, 2004), but PHM seems better able to capture data for syllogisms with many premises (Copeland, 2006). The patterns in syllogistic reasoning are complex and remain contested; one model comparison suggested that no current model of syllogistic reasoning is fully satisfactory (Khemlani & Johnson-Laird, 2012).

One promising recent approach is the Probabilistic Representation Model (PRM; Hattori, 2016), which combines aspects of PHM and mental model theory. PRM proposes that people construct, and reason with, a *probabilistic prototype model*, capturing the eight joint probabilities ranging over the terms, *A*, *B*, and *C*, and their negations (e.g., $\Pr(A, \neg B, C)$). PRM captures these eight probabilities using a simple probabilistic model. The premises may also imply that some of the joint probabilities will be zero. PRM then draws a random sample from the distribution, which will thereby automatically be consistent with the premises. For example, PRM might sample a person who is an *artist* and a *chemist* but not a *beekeeper*. A sequential application of the *min*-heuristic then determines the conclusion that is consistent with the *sample*. As in PHM, conclusion type is determined by the *min*-heuristic. In (3), *Some C are A* is a possible

conclusion, if the sample contains at least one person who is both a *chemist* and an *artist*. If not, PRM checks for a conclusion with the next most informative quantifier and so on iteratively. If no conclusion is consistent with the sample, PRM generates "no valid conclusion." As in PHM, confidence in a conclusion is determined by the *max*-heuristic.

PRM models the empirical data slightly better than either PHM or mental model theory; it is also interesting that the best-fit sample size is six to seven items, roughly consistent with typical assumptions about working-memory capacity. PRM is also theoretically appealing, in that it assumes that reason operates through sampling from a probabilistic model, in a similar way to sampling models in other areas of cognition (e.g., Sanborn & Chater, 2016; Stewart, Chater, & Brown, 2006; Vul, Goodman, Griffiths, & Tenenbaum, 2014).

5. Inductive Reasoning

The psychology of verbal reasoning initially focused on deduction. Yet, from a Bayesian standpoint, reasoning is typically uncertain and knowledge-rich. Thus, Bayesian models are particularly well placed to deal with *inductive* verbal reasoning, where conclusions may extend, often very substantially, beyond the logical consequences of the premises. Inductive verbal reasoning also provides a bridge to the application of Bayesian models in perception, categorization, learning, and so on, which are widespread throughout the cognitive and brain sciences.

One of the best-studied areas of inductive verbal reasoning concerns property induction, such as

(4) Blackbirds have property *X*

Therefore: Penguins/worms have property *X*

Such an induction cannot be a matter of pure logic. After all, any such inference will be logically invalid, as it is perfectly possible that the premise is true and the conclusion false. But also the difference between inferences concerning *penguins* and *worms* does not depend on logical terms but on the nature of the properties themselves. So, for example, the fact that penguins generally seem more similar to blackbirds than either is to worms helps explain why people often generalize properties from blackbirds to penguins more readily (e.g., concerning their anatomy and physiology).

Yet the picture is more complex—and crucially depends on the nature of the property and relevant background knowledge. So, for example, if the property is *being contaminated by a rare poison*, then people may suspect that worms and blackbirds may share this property and that blackbirds may ingest the poison by eating worms or through living in the same environment, while penguins live in a completely different ecosystem (e.g., Kemp & Tenenbaum, 2009).

A further natural question concerns the impact of multiple premises. Consider the following example:

(5) Blackbirds have property *X*

Ostriches have property *X*

Therefore: Penguins have property *X*

Here, the diversity of examples seems persuasive (e.g., Heit & Feeney, 2005). If blackbirds and *sparrows* have property *X*, we may suspect it is limited to garden birds; if blackbirds and *ostriches* do, we may suspect *X* is very widespread across birds as a whole, most likely including penguins. Getting such predictions right requires capturing complex knowledge, which may, for example, be modeled by hierarchical Bayesian models (Kemp & Tenenbaum, 2009).

Abductive reasoning involves "inference to the best explanation" given a set of sensory, linguistic, or scientific data (Harman, 1965). But how do we decide which explanation is best? One criterion, with a long intellectual history dating back to William of Ockham and beyond, is that we should prefer explanations that capture as much as possible with the fewest assumptions. Experimental work with adults and children indicates that people do indeed prefer explanatory simplicity (Lombrozo, 2016) but that they are also drawn to explanations that are as integrated as possible (Lombrozo & Vasilyeva, 2017). For example, people are more convinced by an explanation of weight loss and fatigue that does not explain each symptom independently (loss of appetite coincidentally combined with insomnia) but rather points to a single underlying cause (e.g., depression). Explanations are also preferred to the degree that they cover a diverse range of phenomena (Kim & Keil, 2003), and this is particularly true when people are prompted to pay attention to evidential diversity (Preston & Epley, 2005).

How are considerations such as simplicity and explanatory breadth connected to Bayesian explanations of reasoning? One direct connection is that a preference for simple explanations (where simplicity is judged by the length of description in a coding language) is equivalent to choosing the most probable explanation (according to an appropriate prior) (e.g., Chater, 1996). Moreover, Bayesian philosophy of science has attempted to recast a variety of explanatory virtues as corollaries of a Bayesian analysis (e.g., Bovens & Hartmann, 2003; Howson & Urbach, 1989), although whether this project

is successful has been challenged (Douven & Schupbach, 2015; but see Wojtowicz & DeDeo, 2020).

So far, we have focused on cases where the premises from which reasoning proceeds are taken as given. But people are also active investigators into the world. This raises the question of how people should select information, or conduct experiments, that may be the basis for further reasoning. Initial work in the psychology of reasoning (e.g., Wason, 1968) presupposed that enquiry should seek only to falsify generalizations, on the principle that only falsification can give certainty (in the spirit of Popper's [1934/1959] falsificationist scientific methodology). But from a Bayesian point of view, both disconfirming and confirming evidence can lead us to update our beliefs. Thus, under some circumstances, attempts to confirm the hypothesis may be particularly fruitful.

Suppose we wish, for example, to investigate the somewhat fanciful hypothesis that eating blueberries turns your hair blue temporarily. According to the logic-based falsificationist search strategy, the sole objective is to find someone who has eaten blueberries and whose hair is nonblue and hence to search among blueberry eaters, on the one hand, and people with nonblue hair, on the other. Yet the latter search is likely to be futile—most people have nonblue hair and have also not recently eaten blueberries. On the other hand, it may well be worth asking the much smaller set of people with blue hair whether they have recently eaten blueberries; finding even a few of these might make one suspect that the rule has some validity.

From a Bayesian point of view, the objective of inquiry will often simply be to maximize the amount of information gained from sampling new evidence; of course, before gathering evidence, we cannot know how informative it will be. But Bayesian optimal data selection (Lindley, 1956) recommends that we maximize the *expected* amount of information gained (where the expectation is taken in the light of our probabilistic assumptions about the world). This is the starting point for a rational Bayesian explanation of what was previously viewed as "confirmation bias" in data acquisition tasks, including most notably Wason's (1968) card selection task (Oaksford & Chater, 1994). According to the Bayesian framework, the optimal strategy should depend on probabilistic background assumptions (e.g., the probability of having blue hair, in the example above), and these do, indeed, seem to moderate performance in data selection tasks.

From the standpoint of the logical paradigm, it is natural to see a profound division between deductive reasoning (guided by logical rules) and inductive reasoning (where conclusions extend beyond the logical consequences of the premises). But, as we have seen, both types of process can be modeled in a uniform way in a Bayesian framework. Indeed, it seems natural to see both types of reasoning as inherently uncertain and dependent on rich background knowledge, rather than purely on the verbal premises that are given in the problem. The Bayesian approach, therefore, seems to sit uneasily with the proposal that the mind combines a slow, logical reasoning system with fast associative processes for inductive reasoning (e.g., Heit & Rotello, 2010; Klauer, Beller, & Hütter, 2010; Rips, 2001; Stanovich, 2011). Interestingly, recent applications of signal detection theory to assess whether deductive and inductive reasoning are supported by separate processes or a single system appear to favor the latter (Stephens, Dunn, & Hayes, 2018).

6. Social Reasoning: Argumentation

The psychology of verbal reasoning has traditionally been pursued as if it were only concerned with the cognitive processes of isolated individuals. Yet verbal reasoning, like verbal behavior of all kinds, appears naturally to have a social role: as involved in communication, persuasion, and argumentation (Hahn & Oaksford, 2007; Mercier & Sperber, 2011, 2017). Indeed, understanding verbal reasoning in its communicative context may be crucial to understanding even the most elementary reasoning and to interpreting the significance of the statements over which reasoning occurs (Hilton, 1995). Moreover, the process of pragmatic enrichment may itself involve highly complex reasoning processes (Grice, 1975; Levinson, 2000; Sperber & Wilson, 1986), which are often of far greater sophistication than the patterns of reasoning (MP, MT, etc.) that are ostensibly being probed in experimental tasks. Recall that on hearing *I didn't get hot water this morning*, we immediately infer that the speaker turned the hot tap in the right direction, ran it for sufficiently long, and then discovered the unexpected lack of hot water. We assume, too, that she obtained cold, or perhaps lukewarm, water from the tap, rather than no water all—otherwise, some stronger statement (e.g., *There was no water at all this morning*) would have been given. Speaker and hearer have to do significant inferential work to align on an interpretation—and this process can misfire. It could be, for example, that the speaker is referring to the water at the gym, while the listener assumes they are referring to the water at home. The wider point is that verbal reasoning, as traditionally understood, presupposes that a great deal of "hidden" inferential work has already been carried out (Stenning & Cox, 2006).

Beyond questions of agreeing on the interpretation of what has been said, communication in general, and argumentation in particular, is invariably knowledge rich: whether new arguments or evidence will convince a person to change their beliefs can, in principle, depend on the entirety of their current knowledge. Argumentation is also local and interactive: interlocutors make a series of specific argumentative moves, which may be either accepted or else repelled by further moves, which may themselves be countered, and so on.

How far can a Bayesian approach extend from verbal reasoning to model argument more broadly? This has been a major focus of research for at least 15 years, focusing in particular on the question of whether so-called fallacies of informal argument, which are nonetheless widely deployed in real conversations, have a rational Bayesian foundation that may legitimately lead an audience to change their degrees of belief (Hahn & Oaksford, 2007; Oaksford & Hahn, 2004).

The range of fallacies that have been considered is broad, including arguments from ignorance, ad hominem arguments, and many more. Here, we focus on key examples: slippery-slope and circular arguments.

Consider two slippery-slope arguments (SSAs), both of which have been used in popular debate but differing in strength—from the patently ridiculous (6) to the more credible (7):

(6) If we allow gay marriage, then people will want to marry their pets.

(7) If voluntary euthanasia is legalized, then involuntary euthanasia will be.

From a Bayesian perspective, SSAs can be viewed as founded on a cost–benefit analysis (Corner, Hahn, & Oaksford, 2011; Hahn & Oaksford, 2007): do not take the relevant action (p: allow gay marriage) unless its benefit ($U(p)$) exceeds the expected cost of the consequence (q: allowing interspecies marriage) to which it might conceivably lead ($\Pr(q \mid p)U(q)$). Thus, SSAs are a species of warning (Bonnefon, 2009). A key difference between (6) and (7) is that the probabilities of the consequences seem very different: the chance of significantly raising the probability of interspecies marriage is infinitesimally small, whereas the chance of the extension of euthanasia to at least some nonvoluntary cases may be judged nontrivial (irrespective of how we evaluate either of these consequences).

Interestingly, SSAs can depend crucially on the degree to which our categories are flexible (Hahn & Oaksford, 2007). One reason that the jump from gay marriage to interspecies marriage seems so ludicrous is that these cases are judged to be highly dissimilar, and moreover, the category of allowable marriages is itself quite rigid (it is not a matter of casual interpretation but is rooted in the law). On the other hand, different types of euthanasia may seem far more similar—and the category boundary (e.g., between quite what counts as voluntary or involuntary) may seem anything but rigid. More broadly, the similarity between p and q affects people's willingness both to assign them to the same category and to endorse the corresponding SSA (Corner et al., 2011; see also Rai & Holyoak, 2014, in the context of moral hypocrisy).

The "fallacy" of circularity (Hahn, 2011) is often viewed as problematic even though it is logically valid. To take the extreme case, *If p then p* is not objectionable because it succumbs to logical counterexamples—indeed, it is tautologically true. Instead, circular reasoning can seem argumentatively useless—it presupposes precisely what is to be proved. But how the presupposition operates can significantly affect how fallacious they appear to be, from the patently unconvincing (8) to the rather more compelling (9):

(8) God exists because the Bible says so and the Bible is the word of God

(9) Electrons exist because they make 3-cm tracks in a cloud chamber

A good argument, of any kind, should have the potential to change our beliefs (Harman, 1986). (9) is a type of inference to the best explanation: if we suppose the existence of electrons, that would explain (via a good deal of collateral knowledge about physical laws, the operation of cloud chambers, etc.) the 3-cm tracks; we can't think of any other compelling explanation for the observation of these tracks, so we infer that electrons do indeed exist. Thus, (9) concludes that electrons exist by assuming a theory that requires the existence of electrons and tracing its consequences, and so some nontrivial work has been done (although this work is "hidden" in the details of rival scientific theories and their predictions, which is not, of course, visible in the verbal statement of the argument). Indeed, (9) illustrates what appears to be a benign circularity, inevitable in abductive argument, whether concerning science or not: to explain an observation, we postulate an explanation that is then bolstered (or disconfirmed) to the degree that the explanation captures the observation of interest. (8) is, by contrast, much less compelling: there is no bolstering of the explanation by the observation—in essence, because the prediction is not a distinctive prediction. So, while rival physical theories will struggle to explain the observation of cloud chamber tracks, it is all too easy for any number of nontheological theories to explain why a

religious text will claim divine authorship (these types of considerations can be captured in a hierarchical Bayesian model; Hahn, 2011; Hahn & Oaksford, 2007). Thus, the perceived strength of a circular argument depends not on its structure but on the credibility and number of possible alternative explanations (Hahn & Oaksford, 2007, experiments 1 and 2).

More broadly, from a Bayesian point of view, the study of verbal reasoning and the study of argumentation can be addressed using the same methods and are continuous, rather than corresponding to distinct domains. Indeed, despite the historical priority of the study of individual reasoning, it may be that communication and argumentation are more fundamental and that individual reasoning can only be understood in a social, communicative context (Hahn & Oaksford, 2007; Hilton, 1995; Mercier & Sperber, 2011, 2017; Oaksford & Chater, 2020).

7. Conclusion

Bayesian theories are now widespread in the psychology of verbal reasoning and argument, from what have traditionally been viewed as deductive reasoning tasks, to the study of induction, abduction, and argumentation. In each case, reasoning is presumed to be probabilistic and to depend not merely on the stated premises but typically on the totality of a person's beliefs. The purpose of reasoning of all kinds in daily life is, we have suggested, the updating of belief, rather than, for example, the evaluation of the validity of links between specific premises and conclusions (although this may be crucial in important special cases, such as checking mathematical proofs, legal arguments, and the predictions of scientific theories). Traditional logic-based models of reasoning, by contrast, typically focus on validity and reasoning from the information given. The complexities of semantic and pragmatic interpretation required to even start such reasoning, the relevance of unstated background knowledge, and the social and argumentative context have traditionally been viewed as rather difficult-to-eliminate confounding factors. But from a Bayesian point of view, these are not confounds but the very topic of interest. Indeed, more broadly, Bayesian approaches to verbal reasoning hold the promise of integration with the wider program of modeling world knowledge, social interaction, communication, and the basic cognitive mechanisms of perception, categorization, memory, and motor control within a single theoretical framework (Chater & Oaksford, 2008; Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

Acknowledgments

N.C. was supported by the ESRC Network for Integrated Behavioural Science (grant ES/P008976/1).

Notes

1. As we have pointed out before (e.g., Oaksford & Chater, 2010), this starting point, especially for the conditional, contrasts starkly with the philosophy of logic and language (Nute, 1984), which abandoned the truth-functional view with the advent of the possible-worlds semantics for the conditional (Stalnaker, 1968).

2. We refer the reader to Oaksford and Chater (2020) for a more extensive review of the issues raised here.

3. An intuitive explanation of the origin of this interval is that $\Pr(q \mid \text{not-}p)$ can range between 0 and 1, the former yielding the lower bound and the latter the upper bound.

4. Douven and Verbrugge (2013) argue that the triviality arguments rely on the assumption that the conditional-forming operator is independent of belief states (van Fraassen, 1976). They present evidence that this is not the case because people do not judge the conditional probability of a conditional, that is, $\Pr(\text{if } p \text{ then } q \mid r)$, to be equal to the conditional probability $\Pr(q \mid p \,\&\, r)$.

References

Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: CSLI Publications.

Bonnefon, J.-F. (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review*, *116*, 888–907.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford, England: Oxford University Press.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*, 566–581.

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

Chater, N., & Oaksford, M. (Eds.). (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford, England: Oxford University Press.

Coletti, G., & Scozzafava, R. (2002). *Probabilistic logic in a coherent setting*. Dordrecht, Netherlands: Kluwer.

Copeland, D. E. (2006). Theories of categorical reasoning and extended syllogisms. *Thinking & Reasoning*, *12*, 379–412.

Copeland, D. E., & Radvansky, G. A. (2004). Working memory and syllogistic reasoning. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *57A*, 1437–1457.

Corner, A., Hahn, U., & Oaksford, M. (2011). The psychological mechanism of the slippery slope argument. *Journal of Memory and Language*, *64*, 133–152.

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology*, *6*, 192.

Douven, I., & Schupbach, J. N. (2015). The role of explanatory considerations in updating. *Cognition*, *142*, 299–311.

Douven, I., & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science*, *37*, 711–730.

Edgington, D. (1995). On conditionals. *Mind*, *104*, 235–329.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 321–335.

Evans, J. St. B. T., Thompson, V., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, *6*, 398.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Gilio, A., & Over, D. (2012). The psychology of inferring conditionals from disjunctions: A probabilistic study. *Journal of Mathematical Psychology*, *56*, 118–131.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 47–59). New York, NY: Academic Press.

Hadjichristidis, C., Sloman, S. A., & Over, D. E. (2014). Categorical induction from uncertain premises: Jeffrey's doesn't completely rule. *Thinking & Reasoning*, *20*, 405–431.

Hahn, U. (2011). The problem of circularity in evidence, argument, and explanation. *Perspectives on Psychological Science*, *6*, 172–182.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704–732.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review*, *64*, 88–95.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press.

Hattori, M. (2016). Probabilistic representation in syllogistic reasoning: A theory to integrate mental models and heuristics. *Cognition*, *157*, 296–320.

Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin and Review*, *12*, 340–344.

Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 805–812.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, *118*, 248–271.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. LaSalle, IL: Open Court.

Jeffrey, R. C. (2004). *Subjective probability: The real thing*. Cambridge, England: Cambridge University Press.

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

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*, 20–58.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*, 427–457.

Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition*, *31*, 155–165.

Klauer, K., Beller, S., & Hütter, M. (2010). Conditional reasoning in context: A dual-source model of probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 298–323.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*, 297–315.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, *27*(4), 986–1005.

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*, 748–759.

Lombrozo, T., & Vasilyeva, N. (2017). Causal explanation. In M. Waldmann (Ed.), *Oxford handbook of causal reasoning* (pp. 415–432). Oxford, England: Oxford University Press.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–74.

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.

Nute, D. (1984). Conditional logic. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. 2, pp. 387–439). Dordrecht, Netherlands: Reidel.

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

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

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, England: Oxford University Press.

Oaksford, M., & Chater, N. (2010). Causation and conditionals in the cognitive science of human reasoning. *Open Psychology Journal*, *3*, 105–118.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, *19*, 346–379.

Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*, 305–330.

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

Oaksford, M., & Hahn, U. (2004). A Bayesian analysis of the argument from ignorance. *Canadian Journal of Experimental Psychology*, *58*, 75–85.

Oaksford, M., Over, D. E., & Cruz, N. (2019). Paradigms, possibilities and probabilities: Comment on Hinterecker, Knauff, and Johnson-Laird (2016). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 288–297.

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*, 680–693.

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*, 431–438.

Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic*, *7*, 206–217.

Pfeifer, N., & Kleiter, G. D. (2011). Uncertain deductive reasoning. In K. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A Festschrift for Jonathan St. B. T. Evans* (pp. 145–166). Hove, England: Psychology Press.

Politzer, G., & Baratgin, J. (2016). Deductive schemas with uncertain premises using qualitative probability expressions. *Thinking & Reasoning*, *22*, 78–98.

Politzer, G., Over, D. E., & Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning*, *16*, 172–197.

Popper, K. (1959). *The logic of scientific discovery*. Abingdon, England: Routledge. (Original work published 1934)

Preston, J., & Epley, N. (2005). Explanations versus applications: The explanatory power of valuable beliefs. *Psychological Science*, *16*, 826–832.

Rai, T. S., & Holyoak, K. J. (2014). Rational hypocrisy: A Bayesian analysis based on informal argumentation and slippery slopes. *Cognitive Science*, *38*, 1456–1467.

Ramsey, F. P. (1990). General propositions and causality. In H. A. Mellor (Ed.), *Frank Ramsey: Philosophical Papers*. Cambridge, England: Cambridge University Press. (Original work published 1929)

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, *12*, 129–134.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *113*(33), E4764–E4766.

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

Singmann, H., Klauer, K. C., & Over, D. E. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, *5*, 316.

Sperber, D., & Wilson. D. (1986). *Relevance: Communication and cognition*. Oxford, England: Blackwell.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford, England: Blackwell.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford, England: Oxford University Press.

Stenning, K., & Cox, R. (2006). Reconnecting interpretation to reasoning through individual differences. *Quarterly Journal of Experimental Psychology*, *59*, 1454–1483.

Stephens, R. G., Dunn, J. C., & Hayes, B. K. (2018). Are there two processes in reasoning? The dimensionality of inductive and deductive inferences. *Psychological Review*, *125*, 218–244.

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

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

van Fraassen, B. C. (1976). Probabilities of conditionals. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science* (Vol. 1, pp. 261–301). Dordrecht, Netherlands: Reidel.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*, 599–637.

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

Wojtowicz, Z., & DeDeo, S. (2020). From probability to consilience: How explanatory values implement Bayesian reasoning. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.09.013

Zhao, J., & Osherson, D. (2010). Updating beliefs in light of uncertain evidence: Descriptive assessment of Jeffrey's rule. *Thinking & Reasoning*, *16*, 288–307.

This is a section of doi:10.7551/mitpress/11252.001.0001

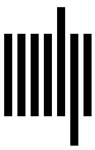
The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455