Reasoning

Sangeet S. Khemlani

NAVY CENTER FOR APPLIED RESEARCH IN ARTIFICIAL INTELLIGENCE, NAVAL RESEARCH LABORATORY

Abstract

Reasoning concerns the cognitive processes by which people draw conclusions from the salient, meaningful pieces of information that they comprehend or observe. Reasoning processes are challenging to investigate because both their initiation and their final product (the inference), can be nonverbal and unconscious. This chapter summarizes recent developments in the science of reasoning. It briefly reviews the differences between “core” patterns of inference, i.e., deduction, induction, and abduction: deductions are inferences that are true in every case that the premises are true. Inductions concern all other sorts of reasoning. And abductions are special types of inductions that yield explanatory hypotheses. The chapter then addresses three fundamental debates that engage contemporary reasoning researchers. The first addresses how to separate rational from irrational deductions. The second concerns the relation between deduction and induction. And the third focuses on how people create explanations. The chapter concludes by addressing ways of making progress to a general, unified account of higher-level reasoning.

Key Terms: inference, deduction, induction, abduction, logic, probability, models

Introduction

Good reasoning helps your survival, and bad reasoning prevents it. To make it easier to investigate reasoning behavior, theorists often lump inferences into different abstract domains (such as reasoning about morals and ethics, space and time, quantity and number, cause and effect). The compartmentalization often quarantines inferences from their contexts in daily life, but reasoning is as ordinary a mental activity as breathing is a physiological one, and it affects the daily decisions humans make. A mother “susses out” her child’s guilt from his stammered responses, and decides on a suitable punishment. A traveler figures out when to take a flight, how to get to and from the airport, and how much it will cost. A driver diagnoses a strange sound in his car as a mechanical problem. These may seem like prosaic inferences to make, but they concern morality, spatiotemporal relations, quantity, and causality. And, even in these narrow contexts, mistakes in reasoning can exact a heavy price: they can affect your relationships, your finances, and your safety.

Reasoning describes the processes that occur between the point when reasoners attend to salient, meaningful information (linguistic or perceptual) and when they draw one or more conclusions based on that information. The processes are challenging to study because both their initiation and their product can be nonverbal and unconscious. Scientific investigations of reasoning began over a century ago, and over a few decades, they coalesced into a belief that was embodied in Inhelder and Piaget's (1958, p. 305) claim that "[human] reasoning is nothing more than the propositional calculus itself." The idea was that mature human reasoning is equivalent to symbolic logic, and so logic formed the basis of the first psychological accounts of reasoning (e.g., Braine, 1978; Johnson-Laird, 1975; Osherson, 1974-1976). Logic does not explain mistakes in reasoning, and so proponents of a form of mental logic argue that erroneous inferences are rare and the result of simple malfunctions in an otherwise capable logical machine (Cohen, 1981; Henle, 1978). Despite the prevailing theoretical consensus, the mid-20th century was a period of quiet confusion. An early pioneer in the field was the British psychologist, Peter Wason, who recognized that human reasoning diverged from logical competence. For one thing, some reasoning tasks revealed biased strategies in otherwise intelligent individuals (Wason, 1960). For another, reasoning seemed to differ from person to person (Wason & Brooks, 1979) and from problem to problem (e.g., Chapman & Chapman, 1959; Ceraso & Provitera, 1971). And, a seminal breakthrough by Wason and his colleagues showed that the contents of a logical problem – i.e., the meanings of the words and their relations to one another – matter just as much as their formal structure (Wason & Johnson-Laird, 1972; Wason & Shapiro, 1971). Logic, of course, deals primarily with formal structures, and it cannot explain human errors or strategies, so it could not account for any of these phenomena.

The quiet confusion that marked the initial decades of the psychology of reasoning gave way to an upheaval of the foundations on which the field was built. Modern researchers are near unanimous in their belief that orthodox logic is an inadequate basis for rational inference (pace Inhelder & Piaget, 1954; see, e.g., Johnson-Laird, 2010; Oaksford & Chater, 1991), but the current era is one of controversy, flux, and shifting paradigms. Broad new frameworks of human rationality exist. One characterizes rationality as optimal reasoning under uncertainty, which is best formalized using the language of probabilistic inference (Oaksford & Chater, 2007, 2009; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). Another argues that the only way to explain the mental processes that underlie reasoning is by understanding how people build mental simulations of the world – mental models – in order to reason (Johnson-Laird, 2006; Johnson-Laird, Khemlani, & Goodwin, 2015). Table 1 summarizes the differences between psychological accounts based on recent frameworks (mental logic, probabilistic logic, and mental models). The frameworks motivated the investigation of a wide variety of reasoning behaviors, such as spatiotemporal reasoning (e.g., Ragni & Knauff, 2013), reasoning about cause and effect (e.g., Waldmann, in press), and argumentation (e.g., Hahn & Oaksford, 2007; Mercier & Sperber, 2011). But the revised scientific view came at the cost of contentious debate: to overturn the view that people reason based on logic, there needs to be a replacement for logic, and psychologists disagree vehemently on what that replacement should be.

Insert Table 1 about here.

Nevertheless, there is reason for enthusiasm. The new frameworks provide varying perspectives on what the mind computes when it reasons and how it carries those computations out. Researchers increasingly rely on methodologies such as mathematical and computational modeling, eyetracking, neuroimaging, large sample studies, and process-tracing to develop and refine novel theoretical proposals. Debates about reasoning helped to motivate broad, architectural descriptions of higher-order thinking (e.g., Johnson-Laird, Khemlani, & Goodwin, 2015; Stanovich, West, & Toplak, 2016; Tenenbaum, Kemp, Griffiths, & Goodman, 2011), and there exists hope in the field that the culmination of these new theories and methodologies will explain long-standing puzzles of human rationality.

Perhaps that hope should be balanced by some pessimism, too. A fundamental problem that besets the community of reasoning researchers – and perhaps the experimental sciences more generally (see Greenwald, 2012) – is that it is nearly impossible to eliminate a theory of reasoning. Consider one very small corner of the field: the psychology of syllogisms. Syllogisms are simple arguments that involve two quantified premises and a conclusion, e.g.:

1. On some days, James doesn't read the newspaper.
   Every day James drinks coffee, he reads the newspaper.

People can spontaneously draw a *valid* conclusion, i.e., one that must be true if the premises are true, from the premises, e.g.:

Therefore, James doesn't drink coffee on some days.

The first published experiment on human reasoning was conducted by Störring more than a hundred years ago, and it concerned syllogisms (Störring, 1908). Störring discovered that his (four!) participants took longer and used a greater variety of strategies for certain types of syllogisms (see Politzer, 2004, p. 213-216). Hence, their data were reliable, predictable, systematic, …and perhaps worth investigating.

In general, syllogisms are tractable to study because they consist of (only) sixty-four problems in their classical form. The early hope was that if reasoning researchers could concur on how humans solve these sixty-four problems, they could build outward to advance a more general theory of human reasoning. Thus far, there exist twelve different theories of the psychology of syllogisms, though each one fails to provide a complete account of syllogistic reasoning phenomena (see Khemlani & Johnson-Laird, 2012). No empirical result is compelling enough to convince adherents of any theory that their proposals should be abandoned. The problem is not isolated to reasoning about syllogisms: accounts of reasoning behaviors flourished in the last few decades across a variety of domains (e.g., causal, counterfactual, and moral reasoning), but few proposals are ever excised from discussion.

New theories are healthy for a burgeoning field because they promote criticism, creativity, and debate. But genuine progress demands eventual consensus, and the existence of twelve competing scientific theories of a narrow corner of reasoning devoted to sixty-four problems portends a looming disaster, since every new theory makes it increasingly difficult to resolve broader arguments about the nature of reasoning. Let us, then, focus on the outstanding debates in the field with the goal of resolving them.

This chapter highlights current controversies in the psychology of reasoning. The goal is not to be disputatious or to adjudicate the various debates; instead, it is to recognize that investigators of reasoning must soon resolve each controversy. Alas, the chapter stays silent on many recent and exciting trends in the investigation of human reasoning. For instance, little is said about analogical, numerical, or causal reasoning (e.g., Holyoak, 2012), or how animals and children learn to reason (e.g., Mody & Carey, 2016; Pepperberg et al., 2013), or how neural circuitry gives rise to higher-order inference (Goel, 2009; Prado, Over, & Booth, 2011). A general survey of recent discoveries in the investigation of human reasoning processes may prove meandering, and so the overview is restricted to three separate debates: first, *what counts as a rational deduction?* Second, *what is the relation between deductive and inductive reasoning?* And third, *how do people create explanations?* Each of these questions corresponds to one of three core patterns of reasoning: deduction, induction, and abduction. The chapter starts by considering what is “core” about the three patterns.

Core inference: Deduction, induction, and abduction

Reasoning is a mental process that draws conclusions from the information available in a set of observations or premises. Aristotle recognized two different types of inference: *deduction*, which he examined through syllogistic reasoning, and *induction*, which he described as an inference “from the particular to the universal” (*Topics*, 105a13-19). Since the advent of symbolic logic in the mid-19th century, the difference is more concrete, and the Aristotelian emphasis on particular and universal assertions no longer applies. What distinguishes the two is that deduction concerns conclusions that are *valid*, i.e., those that must be true “in every case in which all its premises are true” (Jeffrey, 1981, p. 1). Induction concerns arguments whose conclusions *need not be true* when the premises are true. Inductions often describe inferences that are reasonable, typical, or plausible. The two can be distinguished by the way they treat *semantic information* (Bar-Hillel & Carnap, 1954; Johnson-Laird, 1983), which describes the number of possibilities that a set of assertions eliminates. An assertion that eliminates many possibilities, e.g.,

The butler (and nobody else) committed the murder.

is more informative – and less probable (see Adams, 1998) – than one that allows for additional possibilities, e.g.,

The butler, the chef, or the chauffer committed the murder.

Inductive conclusions *increase* semantic information, i.e., they eliminate more possibilities than the premises allow, whereas deductive conclusions maintain semantic information, or even reduce it. Consider the inductive conclusion in this inference:

2. The housing market crashes.
The derivatives market crashes.
Therefore, the stock market will also crash. (induction)

The two premises do not necessarily imply that the stock market will crash, but the conclusion eliminates the possibility that the housing and derivatives markets crash in isolation, so it is more informative than the premises. In contrast, this inference:

3. If the housing market crashes, then the stock market will also crash.
The housing market crashes.
Therefore, the stock market will also crash. (deduction)

is a deduction: its conclusion is if its premises are true. The conclusion explicitly articulates a consequence that is implicit within the two premises, which is another way of saying that it maintains the semantic information in the premises.

Many sets of valid deductions have the same sentential structure. For instance, the deduction above is an instance of the following structure:

4. If A then B.
A.
Therefore, B.

which is a pattern of inference in sentential logic known as "modus ponens". Modus ponens hinges on the meanings of the logical connective *if…then…*. Other logical connectives are *and*, *or,* and *not*. These logical connectives can be used to combine sentences whose meanings are elusive, e.g.,

Either Charvaka is right or else if Jainism is wrong then Buddhism is right.

A reasoner does not need to know the central claims of the Charvaka, Jainists, or Buddhists to draw conclusions from the statement above. (Consider what you might conclude if you learned that, in fact, Charvaka is wrong.) In contrast, inductive inferences resist analyses based on their logical structure alone: reasoners who draw the inductive conclusion in (2) do so based on their background knowledge of housing markets, derivatives markets, and stock markets, and the possible interrelations between them (but cf. Collins & Michalski, 1989).

Abduction is a special kind of induction that yields a hypothesis to explain the premises. This is an example of an abductive inference:

If the housing market crashes, then the stock market crashes.
The housing market crashes.
Therefore, mortgage defaults caused the crashes. (abduction)

The inference is inductive, because the truth of the premises does not guarantee the truth of the conclusion (it is possible that the housing market crashed in the absence of mortgage defaults). But it is also abductive because it yields a causal explanation for the housing market crash, namely that it was caused by mortgage defaults. Abductions, like inductions in general, are difficult to analyze based on their structure alone – but recent theorists have proposed structural preferences in abductive inference, such as a tendency to prefer explanations with simpler causal structures over more complex ones (Lombrozo, 2016; but cf. Zemla et al., in press).

Deductive, inductive, and abductive inferences are a convenient taxonomy with which to organize different patterns of reasoning, though reasoners make all three sorts of inference in daily life, often in tandem with one another. Consider this line of reasoning:

5. If the housing market crashes, then the stock market will crash.
The housing market crashes.
Therefore, the stock market will crash. (deduction)
And so, unemployment will rise. (induction)
And perhaps consumer debt caused the housing market to crash.
(abduction)

The reasoner *deduces* that the stock market must crash, *induces* the effect of the crash on the economy, and attempts to *explain* the downturn. Like other forms of thinking, psychologists can only analyze inferences through indirect means, and researchers have no tools with which to definitively characterize any particular inference “in the wild”. Hence, it is impossible to argue that any pattern of reasoning is more prevalent than any other (pace, e.g., Dewar & Xu, 2010; Oaksford & Chater, 2009; Singmann, Klauer, & Over, 2014). Reasoners often draw a combination of deductive, inductive, and abductive conclusions from given information, and inferences can depend on both the grammatical structure and the content of the information in the premises.

The taxonomy above is useful in characterizing the information contained in the conclusions that reasoners draw. It does not, however, reveal whether those conclusions were rational or not. Was it rational to infer that the stock market crashes in (5)? Was it similarly rational to infer that consumer debt caused the market crash? Rational inferences circa 1950 was uncontroversial – it referred to the kind of rationality sanctioned by symbolic logic. But logical rationality fails as an account of human rationality, and cognitive scientists have searched for alternative ways to characterize rational thought and to identify faulty reasoning. The next section explores the debate over what makes deductions rational.

What counts as a rational deduction?

Does it matter that some of our inferences are faulty? One line of argument holds that if reasoning mechanisms contain fundamental flaws, it would have been impossible to overcome the gauntlet of natural selection. Cohen (1981, p. 317) argued that reasoning mistakes are the “malfunction[s] of an information-processing mechanism,” and reasoners “have to be attributed a competence for reasoning validly, [which] provides the backcloth against which we can study defects in their actual performance.” In other words, mistakes are mere kinks in an otherwise error-free system. A consequence of this view is that mistakes should pose few impediments to survival. They may be infrequent in everyday inference, just as optical illusions – while informative to vision scientists – do not undermine our ability to cope.

The view flies in the face of intuition, because acute errors in reasoning spark controversies in every major social and political debate. Poor reasoning can yield dangerous physical, social, and financial consequences. As the Scottish journalist Charles Mackay recounted, tulips were the most expensive objects in the world during the Dutch tulip mania (Kindleberger, 1978; Mackay, 1869), a bizarre outcome of the erroneous inductive inference that the value of tulips would continue to increase into the foreseeable future. The inference bears close resemblance to the fallacious belief that housing prices will continue to increase (Case, Shiller, & Thompson, 2012). And just as Dutch commerce suffered in the 17th century when the tulip bubble collapsed, the housing bubble and subsequent global financial crisis in the early years of the present century plunged many countries into economic crises. It may be challenging, in the face of such dramatic examples of irrational inferential exuberance, to argue against the idea that human reasoning contains flaws.

A major debate over rationality addresses, not the existence of systematic errors in reasoning, but rather their psychological antecedents (see, e.g., Oaksford & Hall, 2016; Khemlani & Johnson-Laird, in press). Psychologists debate what counts as a mistake in reasoning, as well as whether people are generally optimal reasoners or not (Marcus & Davis, 2013). Until recently, logical validity seemed to be the only metric of human rationality. Logic concerns inferences that can be made with certainty, and in this section, I explain how logic gave way to thinking about reasoning as inherently uncertain.

Logic and its limitations

Theories of deduction no longer posit that reasoning and logic are identical – nevertheless, logic is central to inferences in mathematics, science, and engineering. It eschews the imprecision and vagueness of natural language, and as a result, its principles form the bedrock of computability theory (Boolos & Jeffrey, 1989; Davis, 2000; Turing, 1937). Provided that you can translate a statement in natural language into a logical expression, logic provides a way of deriving new expressions, all of which are also true. For example, you might translate the following statements:

Juan eats an apple or he eats biscotti or he eats both.
If Juan eats an apple, then he does not eat biscotti.

into logical expressions using sentential logic, a type of logic that concerns inferences from categorical sentences (often symbolized as capital letters, e.g., *A, B, C*) that are combined through operators (e.g., '*&*', '*v*', '→', and '¬', which are logical analogs of "and", "or", "if…then", and "not", respectively). Sentential logic, like most logics, has two parts – a model theory and a proof theory. Model theory defines the meanings of the symbols by the truth conditions that render them true, often illustrated through *truth tables*, while proof theory describes a set of rules that operate over the symbols independent of what makes them true or false. Table 2 provides an overview of some of the compound sentences and their corresponding truth tables. In sentential logic, proof theory and model theory coincide: any conclusion that can be derived through syntactic transformations (proofs) can also be derived through semantic analysis (models). The English sentences above, for instance, might be translated into the following formulas, respectively:

$A \vee B$
$A \rightarrow \neg B$

where *A* stands for "Juan eats an apple," *B* stands for "Juan eats biscotti," and "¬" denotes logical negation. Proof theory can transform the symbols above into new formulas using *rules of inference*. For example, one rule of inference (called "disjunction elimination") states that if the following symbols are given or derived:

A v B
¬A

then a derivation from them is:

B

which logicians take to roughly correspond to the following sensible inference:

Juan eats an apple or he eats biscotti or he eats both.
He doesn't eat an apple.
Therefore, he eats biscotti.

And model theory shows that whenever both *A v B* and ¬*A* true, *B* must be true too, and so the inference is valid. Hence, both proof theory and model theory concur in what can be inferred from those logical formulas. In computer science, researchers develop techniques to automate the process of searching for proofs (Bibel, 2013).

Insert Table 2 about here.

In psychology, logic can serve both normative and descriptive functions. Its role is normative whenever theorists claim that to be rational, reasoners must infer only logically valid inferences. Its role is descriptive when theorists argue that the process of reasoning depends on representing assertions in English as logical expressions, applying rules of inference over those expressions, and building up mental proofs. Many early accounts of reasoning proposed this notion (Braine, 1978; Johnson-Laird, 1975; Osherson, 1974–1976; Rips, 1994) and recent treatments maintain it (Baggio, Lambalgen, & Hagoort, 2015; Monti, Parsons, & Osherson, 2009; O'Brien, 2014; Stenning & van Lambalgen, 2016).

Despite the efforts to characterize human reasoning as fundamentally logical, the prevailing view in psychology is that logic is a flawed yardstick of human rationality (pace Piaget). Three problems vex logical accounts of reasoning: first, there exists no algorithm to recover the logical form of a natural language expression (Johnson-Laird, 2010). The contents within an assertion matter just as much as their structure, and reasoners use background knowledge in comprehending discourse and interpreting sentences. Consider the following inference in logic (and its English translation in parentheses):

6. A → B (If A then B)
   ¬B (Not B)

$\therefore \neg A$ (Therefore, not A)

The inference, known as “modus tollens”, is valid based on its abstract form. But certain contents of *A* and *B* can render the inference counterintuitive (see Johnson-Laird & Byrne, 2002). Consider this inference:

7. If Marnie visited Portugal then he didn’t visit Lisbon.
He visited Lisbon.
Therefore, he didn’t visit Portugal.

Reasoners know that Lisbon is in Portugal, and so the inference, despite its logical validity, seems incorrect. Of course, one can represent the geographic relation between Lisbon and Portugal using some logical formula as an addendum, e.g.,

If Marnie visited Lisbon, then he visited Portugal.

But, incorporating that additional premise does not prevent the counterintuitive inference in (7). Indeed, it allows reasoners to derive a contradiction, e.g.,

Marnie visited Portugal and he didn’t visit Portugal.

Logic is *monotonic* in that any set of premises, even a contradictory one, yields an infinitude of deductions, and nothing requires conclusions to be withdrawn. Human reasoning, in contrast, is *non-monotonic*: new information can overturn old assumptions (Oaksford & Chater, 1991), and contradictions cause reasoners to reject assumptions or else explain inconsistencies (Johnson-Laird, 2006).

A second reason logic fails as an account of human reasoning is that reasoners systematically avoid making many valid, but vapid, inferences. Orthodox logic provides no guide as to whether some are more reasonable than others. Consider these redundant inferences:

8. Ellsworth is in Ohio.
Therefore, he’s in Ohio and he’s in Ohio.
Therefore, he’s in Ohio and he’s in Ohio and he’s in Ohio.
…and so on, ad infinitum.

These deductions are silly in daily life, and no reasonable psychological theory should expect reasoners to produce them (Johnson-Laird et al., 2015). Accounts that rely on logic tend to ignore the problem by explaining only how reasoners evaluate given inferences, and not how they generate them (Braine & O’Brien, 1998; Rips, 2002).

A final difficulty for logic concerns the word “if.” Consider its use in a conditional statement mentioned earlier:

If the housing market crashes, then the stock market crashes.

How do people interpret such statements? Many researchers argue that the answer requires a radical shift away from logic: in logic, the connective that bears the closest resemblance to “if” in the sentence above is the “material conditional” (see

Table 2; and Nickerson, 2015, for a review). Material conditionals are truth functional, i.e., they are true in every situation except when *A* is true and *C* is false (see Table 2). As a result, they yield counterintuitive norms, e.g., they permit the following inference:

9. James is hungry.
   Therefore, if he is happy, then he is hungry.

If people reason based on material conditionals, then the deduction in (9) is valid – no matter what the *if*-clause is! That is because a true *then*-clause renders the conditional conclusion true, even when the *if*-clause is false. The inference may strike the reader as a rebuke to common sense: why should James's hunger imply any dependency between his happiness and his appetite? Orthodox logic calls for the validity of this so-called "paradox" of the material conditional, but many psychologists argue for its invalidity. Indeed, conditionals and their associated inferences may be the topic that most vexes students of reasoning.

One alternative idea has risen to prominence in the last decade: conditionals are not logical, deterministic, or truth functional – but rather *probabilistic*. The idea (which originates from Adams, 1975, 1998) has sweeping implications, and its proponents argue that a probabilistic conditional calls for a probabilistic view of reasoning and rationality more generally. A primary rationale for the probabilistic view is that reasoners rarely deal with certain information, and so the formal framework of thinking needs to take into account uncertainty. I turn to examine the central claims of this new probabilistic paradigm.

Probability and uncertainty

Recent theorists argue that human rationality is fundamentally probabilistic (e.g., Evans & Over, 2013; Fugard, Pfeifer, Mayerhofer, & Kleiter, 2011; Oaksford & Chater, 2007, 2009; Politzer, Over, & Baratgin, 2010). The view is based in the idea that conditionals, quantified statements, assertions about causes and effects, decisions, and perceptual input all convey information about degrees of belief, and that reasoning about beliefs is inherently uncertain (see Elqayam & Over, 2013, for an overview). Early probabilistic accounts of reasoning proposed that subjective probabilities reflect degrees of belief (Adams, 1998; de Finetti, 1995; Ramsey, 1990; Tversky & Kahneman, 1983). Hence, a conditional such as:

If it rains, then the ground is muddy.

means something akin to:

Probably, if it rains then the ground is muddy.

A probabilistic conditional tolerates exceptions, i.e., it can be true even in situations in which it rains and the ground is not muddy. And it can be modeled with mathematical precision using the conditional probability, P(*muddy* | *rain*). The conditional probability assigns a numerical value to the belief that it is muddy under the supposition that it rains. Accordingly, the suppositional theory of conditionals advocated by Evans and Over (2004) argues that reasoners establish their subjective belief in a conditional statement by applying the "Ramsey test" (Ramsey, 1929/1990): they first suppose that the *if*-clause is true, and then they assess the likelihood of the *then*-clause through mental simulation. A corollary of the Ramsey test is that reasoners should equate their belief in a conditional, P(*if A then C*) with an assessment of a conditional probability, P(*C* | *A*). For example, their answers to the following two questions should be nearly identical:

10. What is the probability that if it rains, then the ground is muddy?

P(*if A then C*)

Given that it rains, what is the probability that the ground is muddy?

P(*C* | *A*)

This equivalence, colloquially known as The Equation (Edgington, 1995), has a striking consequence. In the probability calculus, P(*it's muddy* | *it doesn't rain*) has no bearing on P(*muddy* | *it rains*), and so, if people equate P(*if it rains then it's muddy*) with P(*it's muddy* | *it rains*), then they should judge that P(*it's muddy* | *it doesn't rain*) is undefined, irrelevant, or indeterminate. It follows that the truth table of a basic conditional is *defective* and not a function of the truth of the *if*- and *then*-clauses (see Table 2; and also de Finetti, 1936/1995; Ramsey, 1929/1990). The defective interpretation of a conditional is a consequence of the Ramsey test and the Equation, and the three assumptions provide a formal framework for reasoning about conditionals and other sentences non-monotonically (Oaksford & Chater, 2013).

A final assumption of the probabilistic paradigm is that *probabilistic validity* (p-validity) supplants logical validity (Adams, 1998; Evans & Over, 2013; Oaksford & Chater, 2007). A deduction is probabilistically valid whenever its conclusion's probability exceeds or is equal to the probability of its premises. When the probability of its conclusion is lower than that of its premises, the inference is invalid, though it may be a plausible induction. This final assumption is powerful enough to dispense with the paradoxes of material implication. Consider how the probabilistic approach handles the paradox in (9):

9'. James is hungry = P(*hungry)*
Therefore, if he is happy, then he is hungry = P(*hungry* | *happy)*

Except in the unlikely case that James being happy has no effect on the probability that he's hungry, the probability of the conclusion is likely to be less probable than the premise: the former describes situations that are a proper subset of the latter. Since the conclusion's probability is lower than that of the premise, it is not probabilistically valid. And so, by rejecting material implication in favor of a defective interpretation of conditionals, and by relying on probabilistic instead of logical validity, the probabilistic paradigm posits a viable solution to explain why humans reject the “paradoxes” of material implication.

The four assumptions above form the pillars of the probabilistic paradigm of reasoning (see Table 1), and they work to counter many of the issues that vex those who advocate a form of mental logic. For instance, the probability calculus allows assertions to vary in their certainty, and additional evidence can lower the probability of a conclusion: hence, unlike orthodox logic, the probability calculus needs no additional machinery to implement non-monotonic reasoning. In recent years, researchers extended the paradigm beyond its initial scope of reasoning about conditionals and quantified assertions to various novel domains, such as reasoning about cause and effect (e.g., Ali, Chater, & Oaksford, 2011; Bonnefon & Sloman, 2013), reasoning about what is permissive and impermissible (e.g., Elqayam, Thompson, Wilkinson, Evans, & Over, 2015) and everyday informal argumentation (e.g., Corner & Hahn, 2009; Hahn & Oaksford, 2007; Harris, Hsu, & Madsen, 2012). The probabilistic approach to reasoning remains a fruitful and dominant perspective on what humans compute when they reason, and it serves as a way to reconceptualize the notion of rationality using the language of probability theory.

Yet, evidence in support of the probabilistic paradigm is mixed. For example, when reasoners have to judge the truth of various conditionals, their behavior supports a defective truth table (Evans, Ellis, & Newstead, 1996; Oberauer & Wilhelm, 2003; Politzer et al., 2010) – but when they have to describe what is possible given a conditional statement, they describe the possibilities that correspond to a material implication (Barrouillet, Gauffroy, & Leças, 2008; Barrouillet, Grosset, & Leças, 2000). Reasoners also report that certain conditionals can be falsified (Johnson-Laird & Tagart, 1969; Oaksford & Stenning, 1992), a result that conflicts with the idea that they tolerate exceptions. One of the most striking predictions of the probabilistic paradigm is The Equation: some studies validate it (e.g., Evans et al., 2013; Geiger & Oberauer, 2010; Handley, Evans, & Thompson, 2006; Oberauer & Wilhelm, 2003; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007), while others conflict with it (Barrouillet & Gauffroy, 2015; Girotto & Johnson-Laird, 2004; Schroyens, Schaeken, & Dieussaert, 2008).

Perhaps a more striking disconnect between reasoning behavior and the probabilistic paradigm is that people appear to interpret sentences deterministically, and not probabilistically, by default. For instance, Goodwin’s (2014) studies show that unmarked, basic conditionals, such as *if A then C*, generally admit no exceptions, whereas conditionals marked as probabilistic, such as *if A then probably C*, allow for violations (see Figure 1). Such a difference should not occur if conditional reasoning is inherently probabilistic. In a similar fashion, proponents of the probabilistic framework argue that causation and causal conditionals are probabilistic (e.g., Ali et al., 2011; Cheng, 2000) and can be formalized using Bayesian networks (e.g., Glymour, 2001; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). They propose that assertions such as *runoff causes contamination* are probabilistic statements that denote that contamination is more likely when runoff is present: P(*contamination* | *runoff*) > P(*contamination* | *no runoff*). But, reasoners can use single observations to establish causal relations (e.g., Schlottman & Shanks, 1992; White, 1999; Ahn & Kalish, 2000; Sloman, 2005) and refute them (Frosch & Johnson-Laird, 2011). They also recognize the distinction between causal and enabling conditions (Khemlani, Barbey, & Johnson-Laird, 2014; Wolff, 2007). The difference between the two is evident in the following two assertions:

> Pulling the trigger *caused* the gun to fire.
> Loading the chamber with bullets *enabled* the gun to fire.

because the causal verbs "caused" and "enabled" are not interchangeable. These results run counter to probabilistic interpretations of causation (see also Pearl, 2009).

Insert Figure 1 about here.

Additional open questions remain about the probabilistic paradigm of rationality. First, how does the paradigm prevent vapid inferences, such as *A, therefore A and A and A* (see also example 8 above)? These inferences are both logically and probabilistically valid, since the conclusions are just as probable as the premises. Second, what do reasoners represent, and how do they process those representations, when they reason? Probabilistic theories of reasoning often describe at the "computational level" of analysis (see Marr, 1982), which describes *what* reasoners compute but not *how* they compute it, and so few of its proponents model online measures of reasoners' inferential processes such as response times and eye-tracking (but cf. Chater & Oaksford, 1999). Third, how can probabilism apply to spatiotemporal and kinematic reasoning domains (e.g., Hegarty, 2004; Khemlani, Mackiewicz, Bucciarelli, & Johnson-Laird, 2013; Ragni & Knauff, 2013; Knauff, 2013)? These domains reflect structural relations amongst entities (e.g., the spoon *is next to* the fork), and it is difficult to see how probabilities enter into these structures. Finally, why do people make systematically erroneous inferences? All valid deductions are also p-valid (but not vice versa; see Evans, 2012), and so a systematic failure to draw a valid inference is a failure of p-validity, too. Humans appear to be predictably irrational on any measure of rationality.

Despite these questions, advocates of the probabilistic paradigm are unanimous in their proposal that uncertainty plays a crucial role in rational thinking – and the preponderance of data corroborates this claim. As they argue, progress in understanding rationality requires an account of why people reason in degrees of belief, and why some experimental tasks systematically elicit uncertain judgments. As the paradigm continues to develop, new theoretical insights, patterns of behavior, and computational models may resolve the open issues highlighted above. An older paradigm, however, engages each of the issues directly. It advocates that reasoning depends on a more rudimentary notion of uncertainty: possibilities.

Models of possibilities

A model of the world – an architect's blueprint, for instance – represents a possibility. Models of possibilities are intrinsically uncertain, because they mirror only some properties of the things they represent: architectural models typically do not have working plumbing and electrical systems that correspond to those in the buildings they beget, and so they are compatible with different physical instantiations. Models were introduced to psychology by the Scotsman Kenneth Craik, who argued that people build "a 'small-scale model' of external reality and of its own possible actions" and consider alternatives to draw conclusions about the past, understand the present, and anticipate the future (Craik, 1943, p. 61). Craik died prematurely, and so his idea lay dormant until psychologists discovered its importance in vision (Marr, 1982), imagination (Shepard & Metzler, 1971), conceptual knowledge (Gentner & Stevens, 1983), and reasoning (Johnson-Laird, 1975, p. 50; 1983).

Mental model theory – the "model" theory, for short – applies to reasoning of many sorts, including reasoning based on quantifiers, such *all* and *some*, and on sentential connectives, such as *if*, *or*, and *and* (Goodwin, 2014; Johnson-Laird & Byrne, 1991), reasoning about cause and effect (Goldvarg & Johnson-Laird, 2001; Johnson-Laird & Khemlani, in press) and reasoning about probabilities (Johnson-Laird et al., 1999; Khemlani, Lotstein, & Johnson-Laird, 2015). Three main principles underlie the theory. First, each model is an iconic representation – i.e., its structure corresponds to the structure of whatever it represents (see Peirce, 1931-1958) – of a distinct set of possibilities. They capture what is common to all the different ways in which the possibility might occur (Barwise, 1993), and individuals use the meanings of assertions and their own background knowledge to construct them. To represent temporal sequences of events, people can construct static, spatial models that arrange events along a linear dimension (Schaeken, Johnson-Laird, d'Ydewalle, 1996), or else they can kinematic models that unfold in time the way the events do (Johnson-Laird, 1983; Khemlani et al., 2013). But, models can also include abstract tokens, e.g., the symbol for negation (Khemlani, Orenes, & Johnson-Laird, 2012).

Second, models demand cognitive resources, and the more models an inference requires, the more difficult it will be. Reasoners tend to rely on their initial model for most inferences, but they can revise their model to check initial conclusions. Hence, the theory supports two primary reasoning processes: a fast process that builds and scans models without the use of working memory, and a slower, memory-greedy process that revises and rebuilds models and searches for alternative possibilities consistent with the premises (Johnson-Laird, 1983, Chapter 6). The model theory predicts that reasoners should spontaneously use counterexamples to refute invalid deductions.

Third, mental models abide by a "principle of truth" in that they represent only what is true in a possibility, not what is false. Consider the following disjunction:

11. Ann visited Beijing or she visited Sydney.

The mental models of the assertion refer to a set of possibilities that can be depicted in the following diagram:

Beijing	
	Sydney
Beijing	Sydney

The diagram uses tokens in the form of words to stand in place for the mental simulations that reasoners construct, e.g., a simulation of Ann visiting Beijing. The first row above represents the possibility that Ann visited Beijing, but it does not explicitly represent the information that she didn't visit Sydney. The second model captures the opposite scenario, and the third captures the scenario in which she visited both places. In many cases, the mental model of the disjunction suffices. But the incomplete representation leads reasoners to systematically err on problems that require them to think about falsity (Johnson-Laird, Lotstein, & Byrne, 2012; Khemlani & Johnson-Laird, 2009). Reasoners can reduce their errors by fleshing their model out, e.g., by appending tokens that use negation to represent what is false in the model:

Beijing	¬ Sydney
¬ Beijing	Sydney
Beijing	Sydney

This fully explicit model represents both what is true and what is false.

When people build models, they take into account the meanings of words and their relations to one another. Contrast (11) above with (12) below:

12. Ann is in Beijing or she is in Sydney.

Background knowledge prevents reasoners from building the scenario in which Ann is in Beijing and Sydney at the same time, and so the mental model of (12) omits this possibility:

Beijing

Modulation refers to the process of incorporating background knowledge into the construction of a model, and it also operates by establishing temporal and spatial relations between the events. For instance, consider the exclusive disjunction in (13):

13. Ann studied for the test or else she failed it.

The model of the scenario has a parallel structure to the model of (12), but reasoners also know that studying for a test must precede the test itself, and that failing a test happens during (or after) a test. Hence, the full model of the scenario establishes a temporal sequence of events (where time moves from left to right), e.g.,

Studied	Took-test	
	Took-test	Failed

In general, reasoners draw inferences from models by scanning them. If a putative conclusion holds in all models, it is *necessary*; if it holds in most models, it is *probable*; and if it holds in at least one model, it is *possible*. By incorporating background knowledge into a model, the theory explains how reasoners make so-called "bridging" inferences (Clark, 1975; Gernsbacher & Kaschak, 2003). Hence, one reasonable conclusion from the model of (13) is that Ann took the test. But, a single model can support multiple conclusions, and so a valid inference is that if she studied for the test, she didn't fail it. And, conclusions can be "modal", i.e., they can concern what's possible, and so the theory predicts that some people will infer that it is possible that she failed the test (if she didn't study). The premise leaves uncertain whether or not she passed, and so reasoning about uncertainty is central to model-based reasoning (see Khemlani et al., under review).

The model theory makes three predictions unique to theories of reasoning, and they have been borne out by recent studies. First, reasoners should spontaneously use counterexamples when they reason (e.g., Johnson-Laird & Hasson, 2003; Kroger, Nystrom, Cohen, & Johnson-Laird, 2008), and reasoners' reliance on counterexamples should reveal that they treat assertions, such as conditionals and causal statements, deterministically (e.g., Goodwin, 2014; Frosch & Johnson-Laird, 2011). Second, reasoners should fall prey to "illusory" inferences in which mental models suggest a conclusion that contradicts the correct response. Illusions have been discovered in all major domains of reasoning (see Johnson-Laird, Khemlani, & Goodwin, 2015; Khemlani & Johnson-Laird, under review). Third, valid inferences that require one model should be easier than those that require multiple models (e.g., Khemlani, Lotstein, Trafton, & Johnson-Laird, 2015; Knauff, 2013; Ragni & Knauff, 2013).

The model theory differs from logic because it posits that people build sets of possibilities. Logic, instead, concerns truth conditions. Consider this inference, which shows how possibilities diverge from truth conditions:

14. Ann visited Beijing or she visited Sydney, but not both.
 Therefore, Ann visited Beijing or she visited Sydney, or both.

The conclusion in (14) is valid in sentential logic, because the conclusion is true in any set of premises in which the premise is true. But it invalid in the model theory, because the models of the conclusion do not correspond to the models of the premise, i.e., the models of the conclusion permit a possibility (in which Ann visits both cities) that the premise explicitly denies.

How does the model theory handle the seemingly paradoxical inferences that come from the material conditional, e.g., (9)?

9. James is hungry.
 Therefore, if he is happy, then he is hungry.

As in the previous inference, the model theory does not warrant the counterintuitive conclusion in (9) because the model of the conclusion concerns possibilities that the model of the premise does not make explicit. In particular, the fully explicit models of the conditional conclusion are:

happy	hungry
¬ happy	hungry
¬ happy	¬ hungry

As the models show, nothing in the premise that James is hungry supports the possibility in models of the conclusion in which he is not hungry.

So, the model theory explains why people reject "paradoxical" inferences. But, as Orenes and Johnson-Laird (2012) show, theory goes a step further: it predicts that in certain scenarios in which the conditional assertions are modulated, people should accept paradoxical inferences. Consider (15) below, which has a structure that parallels (9) above:

15. Lucia didn't wear the bracelet.
 Therefore, if Lucia wore jewelry then she didn't wear the bracelet.

The model of the premise in (15) is:

¬ bracelet

which is compatible with Lucia either wearing jewelry or not wearing jewelry:

jewelry	¬ bracelet
¬ jewelry	¬ bracelet

And, reasoners know that a bracelet is a type of jewelry, so it is impossible to wear a bracelet without also wearing jewelry. Hence, the fully explicit models of the conditional in (15) are:

jewelry	¬ bracelet
¬ jewelry	¬ bracelet

The premise is consistent with both cases in which she did not wear the bracelet, and so, the conclusion follows. Participants accept inferences akin to (15) on 60% of trials but accept inferences akin to (9) on only 24% of trials (Orenes & Johnson-Laird, 2012, Experiment 1).

Neither logic-based theories nor accounts based on probabilities explain why people do not draw vapid inference, such as the conjunction of a premise with itself. But, the model theory does: there is no mechanism in the model theory to introduce possibilities beyond what is provided by background knowledge or the meanings of the premises. Once a reasoner builds a set of models, he or she can reason only from that finite representation, and so the theory explains why people do not draw the infinitude of valid (but useless) conclusions that any arbitrary set of premises allows.

The model theory paints a clear picture of human rationality: good reasoning requires people to draw relevant and parsimonious inferences after they have considered all possible models of the premises. They need to take into account whether the meanings of the terms and verbs in the premises prohibit certain possibilities or introduce certain relations. A reasoner's failure to take meaning into account can lead to errors and counterintuitive responses, such as paradoxical inferences. Hence, people are equipped with the mechanisms for rational inference, but they often err in practice when they fail to consider what is false, when they fail to search for counterexamples, and when they do not possess the relevant background knowledge to solve a problem accurately.

One central limitation of the theory is that it is difficult to derive quantitative predictions from it. That is because it posits that people build iconic mental simulations, and iconicity differs depending on the reasoning domain. An iconic representation of a quantified assertion, e.g., "Most of the dishes are tasty", concerns sets of entities and their properties, whereas an iconic representation of a spatial assertion, e.g., "The dog is on top of the bed," demands a three-dimensional spatial layout. Hence, researchers build formal computational implementations of the model theory by writing computer programs that simulate its tenets (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). More recent implementations can, in fact, yield quantitative predictions (see, e.g., Khemlani et al., 2015b; Khemlani & Johnson-Laird, 2013). In contrast, proponents of probabilism often eschew notions of what people represent in favor a theory that makes quantitative predictions explicit. Another limitation of the theory is that it does not explain how reasoners learn and induce background knowledge from evidence, whereas probabilistic inference paints a clear picture of learning as an application of Bayes's rule, which explains how to revise beliefs in light of evidence.

Summary

Three overarching models of rationality in deduction exist: the view that logical validity serves as a foundation for rational inference; the view that rationality depends on taking uncertainty into account by modeling it through the probability calculus; and the view that rationality depends on the possibilities to which sentences refer. What remains controversial is the degree to which people's representations are inherently probabilistic and fuzzy, or whether that fuzziness comes from deterministic representations that are processed probabilistically. Any fundamental framework of human rationality needs to explain why some reasoners err systematically and some reasoners get the right answer, why reasoners reject paradoxical inferences and avoid vapid ones, and how to incorporate structured background knowledge into deductive reasoning processes. The next section concerns how people make inductive inferences from that background knowledge.

What's the relation between deductive and inductive reasoning?

Irrationality in deductive reasoning can be easy to characterize. For example, this inference is a conspicuous mistake:

16. A or else B.
    Not A.
    Therefore, not B.

It's not possible for the conclusion to be true given the truth of the premises – the three assertions are inconsistent with another. Researchers debate which definition of validity is most adequate (e.g., Evans & Over, 2013; Johnson-Laird & Byrne, 1991; Khemlani et al., under review; Oaksford & Chater, 2007; Rips, 1994; Singmann et al., 2014), but, provided that an appropriate definition is chosen, violations of it are often transparent. And, many inferences, such as the one in (16), are both invalid and p-invalid. But, as Hume observed, there exists no independent rational way to justify induction (though several scholars have offered proposals, e.g., Carnap, 1962; Skyrms, 1986). Consider the following two inductions:

17a. Horses have property X.
Therefore, cows have property X.
b. Horses have property X.
Therefore, iguanas have property X.

In each case, the truth of the premise does not determine the truth of the conclusion. Knowing nothing about what *property X* means, (17a) might appear a stronger and more plausible argument than (17b), since horses are more similar to cows than they are to iguanas. Seminal work by Rips (1975) established that similarity does indeed affect the propensity to make inductive inferences, and psychologists have cataloged many other relevant aspects of categories and their properties that appear to promote induction (Table 3 provides a partial listing; for more detailed reviews, see Hayes, Heit, & Swendsen, 2010; Heit, 2000). Researchers also developed new computational models of induction whose aim is to account for the aforementioned behaviors (see Table 4).

Insert Table 3 about here.

In the last decade, however, scientists returned to the question of whether inductive and deductive inferences arise from distinct mental operations, or whether the two forms of inference reflect different properties of a unitary process of reasoning. The issue is central to advancing new theories of inductive inference: if deduction and induction come from one unitary process, then it is possible to apply the same computational modeling framework (see Table 1) to each set of problems. If the two types of inference are distinct, the frameworks developed for characterizing deductive inference cannot be used to characterize induction. A study by Rips (2001) sparked the debate: he posited that if deduction and induction rely on the same processes, the instructional manipulations designed to elicit one kind of reasoning over another should have no effect on reasoners' evaluations of an argument. To elicit deductive reasoning, one group of participants in his study was instructed to judge whether a conclusion necessarily followed from a set of premises. To elicit inductive reasoning, another group was instructed to judge the strength of the conclusion (i.e., how plausible and convincing it was) from the same set of premises. The instructional manipulation uncovered a critical interaction. Consider (18a) and (18b) below:

18a. If car X10 runs into a brick wall, it will speed up.
Car X10 runs into a brick wall.
Therefore, Car X10 will speed up.
b. Car X10 runs into a brick wall.
Therefore, Car X10 will stop.

The conclusion in (18a) is valid but inconsistent with background knowledge. The conclusion in (18b) is invalid but consistent with background knowledge. Participants accordingly evaluated (18b), but not (18a) as inductively strong, while they evaluated (18a) as deductively valid. Rips took the divergent behavior between assessing an argument’s strength compared to its validity as evidence against a unitary view of inductive inference, and he argued that induction and deduction reflect different ways of evaluating arguments. In his view, the former incorporates content into the evaluation, while the latter “[takes] a more abstract approach” and “generalizes over specific content” (Rips, 2001, p. 133).

Insert Table 4 about here.

Heit and Rotello (2010) reprised Rips’ argument and applied it to reasoning about categories. Consider (19a) and (19b) below:

19a. Mammals have property X.
Therefore, cows have property X.
b. Horses have property X.
Therefore, cows have property X.

The authors characterized (19a) as deductively valid and (19b) as inductively strong (since horses and cows are similar to one another; see Table 3), though they varied the similarity between premise and conclusion categories for inductive strong arguments. Critics may wonder whether the argument in (19a) is in fact deductively valid. Generic statements such as “mammals give live birth” do admit exceptions, and they are often interpreted as referring to properties about categories instead of quantifications over individuals (Carlson & Pelletier, 1995; Prasada, Khemlani, Leslie, & Glucksberg, 2013). But, reasoners interpret novel generics such as “mammals have property X” as referring to nearly all members of a category (Cimpian, Brandone, & Gelman, 2010). Heit and Rotello also adopted Rips’ (2001) technique of varying the instructions to elicit deductive or inductive inferences. They posited that if the cognitive processes that underlie deduction and induction are distinct, then those processes should vary in their sensitivity to deductive validity. To measure sensitivity, the authors applied a metric from signal detection theory, *d’*, which specifies the difference between a hit rate (i.e., evaluating an argument as valid when it was indeed valid) and a false alarm rate (i.e., incorrectly evaluating an argument as valid when it was invalid) to their data. They found that reasoners who were instructed to reason deductively were more sensitive (their *d’* value was higher) than those instructed to reason inductively. And they echoed Rips’ (2001) conclusion: deduction and induction can sometimes arise from different cognitive processes.

In response to Rips’ (2001) and Heit and Rotello’s (2010) research, Lassiter and Goodman (2015) developed a unitary theory capable of explaining the differences in sensitivity as a function of instructions. Their theory builds on the idea (originally due to Oaksford & Hahn, 2007) that the epistemic modal words used in the instructions, i.e., “necessary” and “plausible”, can be mapped onto thresholded scales for comparison purposes (Kennedy, 2007). The locations of the thresholds may be imprecise and unstable, so Goodman and Lassiter interpreted thresholds as referring to a probability distribution instead of a fixed value. They chose a power law distribution for their thresholds under the assumption that the noise inherent in the thresholds should vary less in situations of extremely high confidence or extremely low confidence. Their model predicts a difference in sensitivity analogous to what Rotello and Heit (2010) discovered. In addition, it predicts that extensive instructional manipulations are not necessary to yield the difference – it should suffice to vary only words "necessary" and "plausible" from problem to problem. They reported data that cashed out their predictions, and argued that their theory serves as a counterexample to the claim that only fundamental differences in inductive and deductive reasoning processes explain differences in sensitivity.

The amount of overlap between the cognitive processes that underlie induction and deduction remains unknown. However, it may prove difficult to maintain the view that the processing of semantic content is pertinent to inductive inferences alone, as contents affect purely deductive inferences in systematic ways. In particular, contents help establish a priori truth values. As Steinberg (1970; 1975) showed, people assess certain statements (such as 20a-c) as redundant (i.e., vacuously true):

20a. The apple is a fruit.
b. The automobile is a vehicle.
c. The husband is a man.

They assess other statements as vacuously false when those statements are nonsensical (e.g., "the chair is a sheep") or else contradictory (e.g., "the infant is an adult"). Recent work shows that reasoners make similar distinctions when engaging in deductive inference. Quelhas and Johnson-Laird (2016) report studies in which they gave participants premises such as those in (21a):

21a. José ate seafood or he ate shrimp.
José ate shrimp.

Most participants (71%) concluded that José ate seafood. The deduction is sensible but impossible to make on the basis of the abstract form of the sentences alone. Neither an inclusive nor an exclusive disjunction permits the inference. Reasoners need to take into account the meaning of the words "shrimp" and "seafood", and in particular, they need to use those meanings to block the consideration of any possibility in which José has shrimp but not seafood (see Khemlani et al., under review; and example (7) above). Hence, contents enter into both deductive and inductive inferences, and do not serve as a means to distinguish between the two (pace Rips, 2001).

Inductive reasoning is often studied through the lens of category and property induction, but people draw inductive inferences beyond reasoning about categories and properties. In particular, two understudied forms of inductive inference appear central to human thinking: probabilistic reasoning about unique events and reasoning about defaults. The chapter addresses each in turn.

Probabilistic reasoning

Probabilistic inferences are often inductions, and both numerate and innumerate cultures reason about probabilities. This statement, for instance:

> 22. SURGEON GENERAL'S WARNING: Tobacco Use Increases The Risk Of Infertility, Stillbirth, and Low Birth Weight

invites the probabilistic induction that if you are a pregnant female smoker, you are likely (but not guaranteed) to suffer from the maladies above. Until the advent of the probability calculus in the late 17th and early 18th centuries (see Hacking, 2006), the dominant view of probabilistic thinking came from Aristotle, who thought that a probable event concerned “things that happened for the most part” (Aristotle, Rhetoric, Book I, 1357a35; Barnes, 1984; Franklin, 2001). The calculus turned qualitative inductions into quantitative ones, and contemporary reasoners have little difficulty drawing quantitative conclusions. For example, what would you guess is the numerical probability that Iran will restart its nuclear weapons program? Some might give a low estimate (less than 10%, say), while others consider it likely (more than 90%). And uncertain reasoners may provide a range to their estimates, e.g., between 30% and 50%. People reason inductively whenever they make such estimates. Some researchers wonder whether probability estimates of unique events are sensible to make (Cosmides & Tooby, 1996). Theorists who assume that probabilities must be based on the frequencies of events argue that probabilities of unique events are unsystematic and unprincipled, and the calculus itself may be irrelevant to events that cannot be interpreted as members of a set of similar events (Gigerenzer, 1994). But, pioneering work by Tversky and Kahneman (1983) suggests that reasoners’ estimates of the probabilities of unique events reflect either their implicit use of heuristics or else their explicit consideration of relevant evidence (see also Tversky & Koehler, 1994). Tversky and Kahneman show that reasoners violate the norms of the probability calculus systematically, e.g., they estimate the probability of a conjunction, P(*A&B*), to be higher than the probability of its individual conjuncts, i.e., P(*A*) and P(*B*). Many researchers have subsequently proposed accounts of this “conjunction fallacy” (Barbey & Sloman, 2007; Fantino, Kulik, Stolarz-Fantino, & Wright, 1997; Wallsten, Budescu, Erev, & Diederich, 1997; Wallsten, Budescu, & Tsao, 1997).

Following Tversky and Kahneman, much of the work on probabilistic reasoning concerned how reasoners estimate the probability of sentential connectives such as conditionals and disjunctions. For example, one dominant view is that reasoners interpret the probability of conditionals, e.g., P(*If A then B*) as equivalent to the conditional probability of *B* given that *A* is true, P(B | A). Recall from the discussion on deductive reasoning that this relation is often referred to as The Equation, and it is a central assumption of many probabilistic theories of reasoning (see Table 1). The extent to which naïve reasoners make use of The Equation remains unknown. In probability theory, a conditional probability, P(*A* | *B*), can be computed from the ratio of P(*A&B*) to P(*B*). But Zhao, Shah, and Osherson (2009) showed that reasoners do not tend to carry out that procedure in estimating real future events. Some authors propose instead that people rely on the aforementioned "Ramsey test" in which they add *B* to their stock of knowledge and then estimate the conditional probability from their estimate of A (e.g., Evans, 2007; Gilio & Over, 2012).

If the Equation holds in daily life, there remains a profound mystery: where to the numbers in estimates of the conditional probability of a unique event come from? A recent dual-process account shows how humans make probabilistic inductions about conditional probabilities (Khemlani et al., 2015a). It posits that reasoners simulate evidence in the form of mental models to build a primitive analog magnitude representation that represents uncertainty. They can then map the representation to an intuitive scale to yield informal estimates of probabilities, e.g., "highly probable", or else they can deliberate to convert the representation into a numerical probability, e.g., 95%. The theory explains systematic violations of the probability calculus such as the conjunction and disjunction fallacies discovered by Kahneman and Tversky, and it supports a Bayesian interpretation of probabilities, which states that reasoners interpret subjective probabilities as degrees of belief. But, it takes a further step in proposing that degrees of belief and estimates of numerical probabilities come from analog magnitude representations of the sort found in animals, children, and adults. Subsequent theories of how people compute probabilities need to explain the representations that underlie them.

Inductive reasoning occurs even in the absence of estimates of probabilities. One sort of induction concerns default reasoning, i.e., reasoning about properties, events, and states of affairs that hold in the absence of contravening information. I conclude the discussion on induction by examining default inference.

Default reasoning

If you learn nothing else about an arbitrary dog other than that he is named Fido, you are apt to conclude that Fido has four legs, absent any information to the contrary. This form of inductive reasoning is called *default inference*, because you would give up your conclusion if, say, you found out that Fido was injured. Default reasoning is particularly prevalent in situations of uncertainty. Reiter (1978) observed that "the effect of a default rule is to implicitly fill in some [gaps in knowledge] by a form of plausible reasoning...Default reasoning may well be the rule, rather than the exception, in reasoning about the world since normally we must act in the presence of incomplete knowledge." Selman and Kautz (1989) echoed Reiter's sentiment and added that "…default reasoning allows an agent to come to a decision and act in the face of incomplete information. It provides a way of cutting off the possibly endless amount of reasoning and observation that an agent might perform." Default reasoning affords monumental efficiency gains in computation, and indeed, many theoretical accounts of default reasoning are due to computer scientists such as the authors above (see also Khardon & Roth, 1995; Thielscher & Schaub, 1995; Gilio, 2012). The theories conflict on what they consider a valid default inference (Doyle & Wellman, 1991), and many systems of default inference are built into object-oriented programming languages.

Given the conflicts, the dearth of experiments on how people carry out default inference is surprising (but cf. Benferhat, Bonnefon, & da Silva Neves, 2005; Pelletier & Elio, 2005). Default inferences appear to depend on reasoners' background knowledge about the world, and so empirical insights can prove instructive. Pelletier and Elio (2005) argued that experimentation is the only appropriate way to understand default inference. There is merit to their argument, as experimentation can uncover subtleties in reasoning by default. The inference about Fido above may seem compelling, but consider a parallel example: suppose you meet a Canadian named Sarah. How confident are you that she is right-handed? Most reasoners appear less willing to draw the default inference in Sarah's case (i.e., that she is right-handed) than in Fido's case (i.e., that he has four legs). Understanding why some default inferences are felicitous and some are not may provide psychological constraints on formal accounts of default reasoning. For instance, a potential account might posit that reasoners have access to the underlying statistics of the world, i.e., they represent four-legged dogs as more prevalent than right-handed Canadians. Nobody has proposed such an account, but it is an implicit view of unstructured probabilistic models of cognition. Still, while subjective evaluations of prevalence are important, reasoners' conceptual understanding of the world may be even more so, because conceptual representations of categories contain structure beyond information about prevalence. For example, reasoners agree with the following generic assertion:

23. Mosquitoes carry malaria.

even though they recognize that only a small minority of mosquitoes exhibit that behavior (Leslie, Khemlani, & Glucksberg, 2011; Leslie, 2008). In other words, had people operated based on prevalence alone, they should have assessed (23) as false. Generic assertions appear to provide researchers a window onto reasoners' conceptual structures (Brandone, Cimpian, Leslie, & Gelman, 2012; Carlson & Pelletier, 1995; Gelman, 2003), i.e., reasoners appear to agree with generics only when certain relations between the category and the property hold (Prasada et al., 2013). A recent study examined whether people's agreement to generic assertions should govern default reasoning behavior (Khemlani, Leslie, & Glucksberg, 2012). Participants in the study received the following problem:

24. Suppose you're told that Leo is a lion.
    What do you think of the following statement: Leo eats people.

They were asked to judge the statement on a confidence scale that ranged from 3 ("I'm confident it's true") to -3 ("I'm confident it's false"). The crucial manipulation was the connection between the kind (*lions*) and the property (*eating people*). In (24) above, the connection is that eating people is a *striking* property of lions, i.e., it is a behavior that signifies a dangerous predisposition to be avoided, and it renders the corresponding generic assertion ("lions eat people") true (see Leslie et al., 2011). Hence, reasoners should be more likely to make the inference. In contrast, consider (25) below:

25. Suppose you're told that Viv is an athlete.
    What do you think of the following statement: Viv is a student.

In (25), there is less of a semantic connection between the kind (*athlete*) and the property (*being a student*), and so the corresponding generic ("athletes are students") is judged false. Examples (24) and (25) are comparable with regard to their prevalences because the properties (*eating people* and *being a student*) hold for only a minority of the kind, i.e., very few lions eat people, and very few athletes are students. The data in our study were subjected to regression analyses that compared participants' performance to normed evaluations of generic agreement and prevalence estimation; they revealed that generic agreement accounted for more variance than prevalence alone.

In sum, reasoners make default inferences based on more than just statistical information. They pay attention to semantic considerations such as how striking or dangerous a property is, and other semantic relations as well, such as whether a property is characteristic of a kind (Gelman, 2003; Medin & Ortony, 1989; Prasada et al., 2013).

Summary

Reasoners engage in different forms of inductive reasoning: they induce properties of categories, they estimate probabilities of events, and they make default inferences. The mathematics of the probability calculus provides ways of formalizing people's inductive inferences, but people systematically violate simple applications of the calculus (see Sanborn & Chater, 2016, for a recent synthesis). They appear to base their inductions on information about both probabilities (e.g., prevalence information) and structures (e.g., mental representations of kinds). Future theories must explain how probabilities and structural information coexist.

One aspect of inductive reasoning is the ability to construct explanations of observations, both expected and anomalous. Explanations require reasoners to consult their background knowledge, and so they heavily rely on preexisting concepts and representations. The next section examines recent investigations of explanatory reasoning.

How do people create explanations?

A core feature of human rationality is the ability to explain observed behaviors and phenomena (Harman, 1965). Explanations allow reasoners to make sense of the past and anticipate the future (Anderson & Ross, 1980; Craik, 1943; Einhorn & Hogarth, 1986; Gopnik, 2000; Lombrozo & Carey, 2006; Ross, Lepper, Strack, & Steinmetz, 1977), and they are central to the way we communicate our understanding of the world (Johnson-Laird, 1983; Lombrozo, 2007). The need to explain the world has its downside, too: explanatory reasoning is the genesis of superstitions, magical thinking, and conspiracy theories, all of which can be resistant to factual refutation. A compelling explanation can be a powerful way of synthesizing disparate sources of information, whether or not that synthesis is warranted. Hence, the process of constructing an explanation is separate from how explanations are evaluated.

The logician Charles Sanders Peirce coined the phrase “abduction” to describe the process by which reasoners infer explanations as a way of highlighting its differences from deductive and inductive patterns of reasoning. He argued that when reasoners abduce, they form a set of explanatory hypotheses. And he argued that abduction “is the only logical operation which introduces any new idea” (CP 5.172).

Explanatory reasoning poses a challenge to empirical investigations because, while experiments on deduction and induction can systematically remove portions of background knowledge from reasoning problems, explanations seem inextricably tied to that knowledge, and so researchers worry that experimental manipulations of content can heavily bias the kinds of explanations participants produce. Applied domains afford systematic ways of understanding reasoners’ background knowledge, and the earliest research on explanatory reasoning examined domain-specific explanations such as those produced in fault diagnosis (e.g., Besnard & Bastien-Toniazzo, 1999; Rasmussen, 1981; Rouse & Hunt, 1984) and medical decision-making (e.g., Elstein, Shulman, & Sprafka, 1978; Kassirer, 1989). Interest in domain-general explanatory reasoning and its underlying cognitive processes is relatively new (Keil, 2006), and researchers are only beginning to discover how explanations are central to a broad swathe of domains, including inductive reasoning (see Table 3), categorization, conceptual development, and learning (Lombrozo, 2006, 2016).

Some researchers investigate explanations through the lens of mechanisms involved in encoding and retrieving memories, because people retrieve previously inferred explanations from memory (Melhorn, Taatgen, Lebiere, & Krems, 2011), and they spontaneously produce new ones when they encoding categories (Shafto & Coley, 2003). Other researchers investigate how explanations aid in conceptual development (Murphy, 2000; Patalano, Chin-Parker, & Ross, 2006), and cognitive development more broadly (Keil, 2006; Legare, 2012; Wellman, Hickling, & Schult, 1997). But, by far, most research into explanatory reasoning comes from research into causal cognition (Ahn & Kalish, 2000; Alicke, Mandel, Hilton, Gerstenberg, & Lagnado, 2015; Ferbach, Macris, & Sobel, 2012; Johnson-Laird, Girotto, & Legrenzi, 2004; Khemlani, Sussman, & Oppenheimer, 2011; Lombrozo, 2016; Sloman, 2005).

Causality and explanatory reasoning

Explanations needn't be causal (Aristotle, trans. 1989, 1013a). You can explain the logic of compound interest, why the Panama Canal connects two oceans, and why the elements of a painting make it beautiful without appealing to any causal relations. Recent research into Aristotle's non-causal explanations includes Prasada and Dillingham's (2006) explorations of "formal" explanations, i.e., how individuals explain certain properties of an individual by appealing to only to the kind of thing that it is. Consider (26a) and (26b) below:

26a. A lion has a mane because it is a lion.
b. A lion has four legs because it is a lion.*

The former is an example of a felicitous formal explanation while the latter is an infelicitous explanation, and as the examples show, some properties afford formal explanations while others do not (Prasada & Dillingham, 2006, 2009; Prasada et al., 2013). "Teleological" explanations, also referred to as "final" explanations, are similarly non-causal. They help explain a property by appealing to its function, goal, or end result. Children use teleological explanations throughout development, e.g., they endorse statements such as "pens are for writing" and "mountains are for climbing" (Kelemen, 1999; Kelemen & DiYanni, 2005). Adults are more skeptical and introspective about these claims; they use teleological explanations for artifacts more often than for natural kinds (Lombrozo & Carey, 2006).

Nevertheless, the vast majority of daily explanations refer to causal relations. To explain how wine turns into vinegar, why coral reefs are dying, or how a prediction market works, for instance, you need to identify the underlying components in each phenomenon as well as their causal relations to one another.

Explanatory reasoning appears to develop alongside causal reasoning (Wellman & Liu, 2007). Callanan and Oakes (1992) conducted a study in which they asked mothers to keep records of their children's requests for explanations. The children in the study asked numerous questions about causal relations concerning natural and mechanical phenomena (e.g., "Why do stars twinkle?", "How does that wheelchair work?"). More recently, Hickling and Wellman (2001) examined children's conversations and coded them for causal questions and explanations. In both approaches, requests for causal explanations appeared early in development and were produced more frequently than causal propositions. Indeed, "why?" questions were amongst the earliest utterances produced by the children.

Adults frequently generate causal explanations, too (Einhorn & Hogarth, 1986; Hilton & Erb, 1996). When explanations contain both causal and non-causal elements, causal elements tend to influence patterns of judgment over non-causal ones (Murphy & Medin, 1985). Causal explanations appear to facilitate category learning and induction (Rehder & Hastie, 2004). Causes that occur early in a causal chain, and causes that are causally interconnected, are deemed more important (Ahn, Kim, Lassaline, & Dennis, 2000; Khemlani & Johnson-Laird, 2015). Hence, causal structures have a unique and indispensable role in abductive reasoning.

There are two overarching ways in which causes enter into explanatory reasoning. First, reasoners evaluate the causal structure of explanations based on numerous factors, such as how parsimonious the structure is, how complete it is, and how well it coheres with other beliefs. Second, reasoners generate causal structures that suffice as explanations. Let us examine each of these behaviors.

Evaluating explanatory fitness

As Lombrozo (2016) argues, untrained reasoners reliably prefer certain types of explanations over others (see also Keil, 2006; Lombrozo, 2006). In daily life, people often evaluate whether a given explanation is compelling, justified, and worth pursuing (Zemla, Sloman, & Lagnado, in press) particularly to understand complex phenomena and to resolve inconsistencies. From Newton to Peirce, philosophers and scientists argue that scientific explanations should be broad, and recent studies suggest some biases toward preferring simpler explanations (Chater, 1996; Einhorn & Hogarth, 1986; Lagnado, 1994; Lombrozo, 2007). For example, in one study, Lombrozo (2007) gave participants problems concerning diseases and symptoms of aliens on another planet, e.g.:

The alien, Treda, has two symptoms: her minttels are sore and she has developed purple spots. Tritchet's syndrome always causes both sore minttels and purple spots. Morad's disease always causes sore minttels, but the disease never causes purple spots. When an alien has a Humel infection, that alien will always develop purple spots, but the infection will never cause sore minttels. What do you think is the most satisfying explanation for the symptoms that Treda is exhibiting?

Participants preferred the simpler explanation (Tritchet's syndrome) to a complex one (the combination of Morad's disease and a Humel infection) despite knowing that Treda could possess multiple illnesses. Lombrozo argues that this preference reflects a general bias toward more simple explanations, and other data in support of simplicity biases has led some researchers to argue that simplicity is a fundamental cognitive principle (Chater & Vitanyi, 2003).

An idea that runs parallel to simplicity is that good explanations are often coherent, i.e., their causal elements cohere with themselves (internal coherency), with facts about the world (external coherence), and do not contain inconsistencies. Proponents of coherentism hold that good explanations ought to be thought of as sets of beliefs which act as an organized, interdependent unit that does not depend on lower levels of explanation (Amini, 2003; Read & Marcus-Newhall, 1993; Thagard, 2000; Van Overwalle, 1998). Coherent explanations do not make assumptions specific to a particular phenomenon, and so Lombrozo (2016) interprets coherentism as related to simplicity. But, few empirical studies have directly examined the preference for coherency apart from studies by Read and Marcus-Newhall (1993), and so commitments to coherency must be qualified: as Keil (2006) observes, coherency is often violated because knowledge is incomplete and inconsistent (see also Gillespie & Esterly, 2004).

Indeed, preferences for simpler and more coherent explanations may be overridden by other factors. Individuals often prefer complex explanations that are more complete, i.e., ones that satisfy expectations about an underlying causal mechanism. For instance, Johnson-Laird et al. (2004) ran experiments in which they solicited participants' spontaneous explanations as well as their probability ratings for explanations that consisted of a cause versus those that consisted of a cause and an effect when reasoning about conflicting information. For example, reasoners were given the following problem:

> If someone pulled the trigger, then the gun fired. Someone pulled the trigger, but the gun did not fire. Why not?

and rated two putative explanations for the gun not firing:

> 27a. A prudent person unloaded the gun and there were no bullets in the chamber.
> b. There were no bullets in the chamber.

They systematically rated (27a) more probable than (27b). Recent studies replicated the phenomenon, and suggest that reasoners prefer complete explanations to incomplete ones (Legrenzi & Johnson-Laird, 2005) and to non-explanations (Khemlani & Johnson-Laird, 2011). These preferences may be construed as an error, because rating (27a) as more probable than (27b) is an instance of the conjunction fallacy (Tversky & Kahneman, 1983). More recently, Zemla and colleagues (in press) asked reasoners to rate naturalistic explanations for various questions submitted to the online bulletin board system, Reddit (e.g., "Why has the price of higher education skyrocketed in the US…?") and found that people rated complex explanations submitted by users more convincing than simpler explanations. Subsequent experimental studies corroborated this preference.

Good explanations are often relevant and informative (Grice, 1975; Wilson & Sperber, 2004), and they can fail when speakers provide too much information – under the assumption that the listener lacks information that, in fact, she knows – or else when they provide too little – under the assumption that the listener knows information than she, in fact, lacks. Irrelevant explanations can occur as a result of an egocentric bias in which people mistakenly assume that listeners share the same knowledge they do (Hilton & Erb, 1996; Keysar, Barr, & Horton, 1998; Krauss & Glucksberg, 1969; McGill, 1990; Nickerson, 2001). To calibrate an explanation to a listener's knowledge, speakers must overcome a "curse of knowledge" (Birch & Bloom, 2003) in which their detailed knowledge can interfere with their understanding of what their listener does not know. Egocentrism can be overcome by negotiating or inferring common background information between speakers and listeners (Clark, 1996; Levinson, 2000). One promising account of this negotiation process is rational speech-act theory, which formalizes the inference of common background information as form of Bayesian inference over a speaker's knowledge state and a listener's interpretation of the speaker's words (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013).

The cosmologist Max Tegmark wrote that a good explanation "answers more than you ask" (cited in Brockman, Ferguson, & Ford, 2014). He shares the view of many scientists and philosophers who note that scientific explanations should be broad (Kuhn, 1977; Whewell, 1840). A reasonable psychological prediction is that people should prefer explanations with broad scope as well (see Thagard, 1992), and they often do. In the aforementioned studies by Read and Marcus-Newhall (1993), participants learned a few facts about an arbitrary woman, e.g., that she has nausea, weight gain, and fatigue. They consistently preferred broad scope explanations (e.g., "she is pregnant") that explained all of the facts to narrow scope explanations (e.g., "she has a stomach virus") that explained only a subset of the facts. In situations of complete, certain information, broad scope explanations seem sensible. But, those situations are rare, and abduction often serves to resolve situations of uncertainty and inconsistency. Hence, good explanations often explain only what needs to be explained (manifest scope) and not unobserved phenomena (latent scope), reasoners appear to prefer explanations of narrow latent scope (Johnson, Rajeev-Kumar, & Keil, 2014, 2016; Khemlani, Sussman, & Oppenheimer, 2011; Sussman, Khemlani, & Oppenheimer, 2014). In experiments due to Khemlani, Sussman, and Oppenheimer (2011), participants were given problems of the following form:

> A causes X and Y.
> B causes X, Y, and Z.
> Nothing else is known to cause X, Y, or Z.
> X occurred; we don't know whether or not Y or Z occurred.

In the study, A, B, X, Y, and Z were replaced by sensible contents. As the problem makes evident, Explanation A has narrower latent scope than Explanation B, because it can account for fewer effects. Individuals judged A to be more satisfying and more probable relative to B. Children exhibit a similar bias (Johnston, Johnson, Koven, & Keil, in press), and a recent account suggests that people base their preferences for A over B on an inference concerning the uncertain, unverified prediction (i.e., Y or Z; see Johnson, Rajeev-Kuman, & Keil, 2016). In any case, the bias toward narrow scope explanations can conflict with a bias toward simplicity, because some explanations can be both more complex and more narrow than simpler alternatives.

Table 5 provides an overview of the different biases that exist in human explanatory reasoning. The list itself appears internally inconsistent: how can reasoners maintain biases for both simplicity and complexity and narrower scope? These fundamental conflicts in explanatory reasoning stand in need of resolution. Perhaps a bias for simplicity is too simple an account of explanatory evaluation, but its alternative – that reasoners prefer complexity – is strictly false, because reasoners tend to avoid infinite regress when constructing explanations. These conflicts result may come from a scarcity of theoretical accounts of explanatory reasoning. Few accounts exist that explain how explanations are generated in the first place, and those that do tend to emphasize the role of retrieving explanatory hypotheses from memory (but cf. Johnson-Laird et al., 2004; Thomas, Dougherty, Sprenger, & Harbison, 2008). A recent proposal assumes that reasoners rely on a set of heuristics to construct explanations.

Insert Table 5 about here.

Explanatory heuristics

How do people generate explanations? Studies that reveal preferences from some explanations over others lend little insight into the processes by which reasoners construct explanations. Apart from a lack of relevant data, there is an overarching theoretical paradox for why nobody has proposed any such process-level account: on the one hand, explanatory reasoning appears to be an enormously complex task to carry out. Reasoners must search through their semantic memories, i.e., through vast amounts of conceptual and relational knowledge; they must creatively combine relevant portions of that knowledge to produce a plausible causal mechanism; and they must assess their putative explanation against their knowledge about the phenomenon. Much of our understanding of the complexity of the operations needed to build explanations suggests that they should demand extensive computation. On the other hand, people construct explanations rapidly. Reasoners do not appear to be flummoxed when constructing an explanation, and many are capable of offering multiple explanatory guesses. Why is a friend of yours running late? Perhaps he is stuck in traffic, perhaps his prior meeting ran late, perhaps he is injured or sick. And so, the speed with which you construct and then evaluate these explanations suggests that explanations should are easy to generate. This is what I call the "paradox of fast explanations", and without resolving it, an account of how people generate explanations will remain elusive.

Cimpian and Salomon (2014) recently proposed a theory that may resolve the paradox of fast explanations. They argue that people generate initial explanations heuristically and that these heuristic explanations concern inherent features, i.e., features internal to the elements of a phenomenon. Consider this question: "Why do people drink orange juice in the morning?" Cimpian and Saloman argue that when reasoners first identify patterns in the world (e.g., that people drink orange juice for breakfast) they spontaneously explain those patterns in terms of, say, the tanginess of orange juice instead of the promotion of oranges by the citrus lobby. The former is a property intrinsic of orange juice and the latter is a property external to the phenomena to be explained.

What happens during the construction process? Cimpian and Salomon propose that the main entities of the phenomenon to be explained (orange juice as a breakfast beverage) become active in working memory, and that the activation of these memories spreads to inherent properties that are a central to the representation of the entities in semantic memory (see McRae & Jones, 2013), such as the tanginess of orange juice. Hence, reasoners are likely to retrieve those properties and base their explanations on them. Thus, when the cognitive system assembles an explanation, its output will be skewed towards explanatory intuitions that appeal to the inherent features of the relevant focal entities. And recent studies by Cimpian and his colleagues corroborate reasoners' spontaneous use of heuristics in explanation (e.g., Cimpian & Steinberg, 2014; Horne, 2017; Hussak & Cimpian, 2015; Sutherland & Cimpian, 2015).

Summary

Psychologists have begun to catalog a set of explanatory preferences, and in the coming years, researchers should discover new factors that separate compelling explanations from unconvincing ones. But, the conjunction of the existing preferences does not constitute a theory, and indeed, there presently exists no theory of domain-general explanatory reasoning. The quest for such a theory may prove chimerical: reasoners' preferences for certain kinds of explanations may be an artifact of the kinds of impoverished problems they face in laboratory settings. Still, two general – and conflicting – trends have emerged from recent investigations: first, reasoners prefer simple explanations that broadly account for observed information. And second, they prefer complex explanations that elucidate underlying mechanisms and do not explain more than what is observed or known about a particular phenomenon. It may be that these two general preferences mirror universal cognitive strategies for exploiting and exploring: in some situations, reasoners are motivated to exploit known information to save cognitive resources, and in others, they benefit from exploring the space of possibilities. One potentially productive line of investigation might attempt to characterize the scenarios under which reasoners choose to explore instead of exploit, and vice versa. Still, a general theory cannot account only for how people assess given explanations; it needs to describe the cognitive processes that generate explanations.

Conclusion

Towards the end of his chapter on speech and language in the first edition of the *Stevens' Handbook of Experimental Psychology*, George Miller described seminal research into the psychology of reasoning behavior. At the time, psychologists had proposed early accounts of syllogistic inference, but a sustained research program into reasoning wouldn't emerge for another few decades, and so his review comprised only a few paragraphs. Progress was swift. By the time the second edition of the *Stevens' Handbook* was published in the late 1980s, theorists had developed novel methodologies for studying thinking behavior and a community of researchers in artificial intelligence, psychology, and philosophy had developed new formal and computational frameworks to characterize human reasoning. And so, James Greeno and Herb Simon devoted an entire chapter to the burgeoning field (Greeno & Simon, 1988). In 2002, so many theories of reasoning had flourished that Lance Rips sought to categorize them into various families in his chapter in the third edition of the *Stevens' Handbook* (Rips, 2002).

Many of the advances that occurred in the years since the third edition produced important new controversies and debates, and this chapter reviewed three of them. First, while decades old skepticism that everyday inferences correspond to those sanctioned by conventional logic persists, reasoning researchers had few alternative options for formal accounts of what counts as a rational deduction. In the last decade, theories of rational inference matured into two competing research programs, and the first section of this chapter explored the debate. One account, commonly referred to as the "new paradigm", proposes that reasoning is inherently probabilistic, i.e., that premises and conclusions –

particularly those that concern conditional assertions – are best formalized in the probability calculus, where inferences transform prior beliefs into posterior probabilities. Another account, the model theory, holds that reasoning depends on the mental simulation of iconic possibilities ("mental models"). Reasoners construct and scan these possibilities to draw conclusions, and inferences are difficult when they need to consider multiple possibilities. Both accounts agree on some fundamental assumptions: reasoning is uncertain and non-monotonic; the everyday use of compound assertions is not truth functional; reasoning depends on both the form and the contents of the premises. But they diverge in important ways, too (see Johnson-Laird et al., 2015a,b, and Baratgin et al., 2015). The probabilistic account explains human reasoning as an application of Bayesian inference, which had previously been used to explain learning. And so, the probabilistic account connects reasoning and learning in novel ways. The model theory, in contrast, explains human difficulty and inferential biases, and its goal is to characterize the constraints of the mental representations that reasoners typically build. At the time of writing, no single dataset or experimental paradigm seems sufficient enough to adjudicate the two accounts. Nevertheless, the two approaches are getting closer to one another, not farther away. Researchers have begun to merge notions of mental simulation and probabilistic inference (see, e.g., Battaglia, Hamrick, & Tenenbaum, 2013; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2012; Hattori, 2016; Khemlani et al., 2015b; Khemlani & Johnson-Laird, 2013, 2016; Oaksford & Chater, 2010; Sanborn, Mansighka, & Griffiths, 2013). Perhaps a hybrid approach may resolve the extant controversies between the two accounts, or perhaps a new framework for thinking about rationality is needed. Any sort of advance would to need to resolve the controversies raised by the probabilistic and the simulation-based accounts, however.

A second major debate concerns the difference between inductive reasoning and deductive reasoning. Scholars since antiquity had characterized induction and deduction as separate constructs, and that tradition carried into contemporary psychology. But, not until the early part of this century did researchers conduct rigorous tests of whether induction and deduction arise from separate mental processes. Initial work by Lance Rips triggered a debate between several communities of researchers. They sought to explain why reasoners judge compelling, but invalid, inductive inferences as strong and why they simultaneously accept unconvincing deductions as valid. The relation between induction and deduction remains unclear, but research shows that two accounts under investigation, i.e., the account that posits that induction and deduction rely on distinct mental processes and the account that supposes that induction and deduction rely on a unitary mental process, are both viable. Hence, a task for future research is to explain when and how inductive and deductive reasoning processes diverge from one another. And the interplay between inductive and deductive inference may be particularly pronounced when reasoning about unique probabilities and defaults, because both sorts of inference require reasoners to base their information on uncertain background knowledge and dynamic situations.

Despite some formal accounts of abductive reasoning in the artificial intelligence community, as well as many conceptual frameworks proposed by philosophers, empirical work into the processes by which people construct explanations prior to the turn of the century was rare and exceptional. In the years that followed, explanatory reasoning research grew into a sustained focus for cognitive scientists. Indeed, explanatory reasoning can be considered one of the field's major growth industries. Researchers have begun to investigate the biases by which some explanations are deemed more compelling than others, as well as the constraints on those preferences (see Lombrozo, 2016, for a review). One major advantage of researching explanatory reasoning seems to be that children utilize their burgeoning linguistic abilities to ask and understand their parents' explanations. Hence, it is sensible to study how explanatory reasoning shifts and changes across the lifespan. Nevertheless, explanations can be challenging to investigate, because reasoners spontaneously draw information from their background knowledge to build explanations. And so, explanations pose a challenge to empirical researchers and their desire for highly controlled laboratory studies. Perhaps that is one reason that, despite pronounced interest in explanatory reasoning, few theories exist that can explain a fundamental paradox of explanatory reasoning: reasoners appear to draw explanations rapidly, despite the need to search through vast amounts of conceptual knowledge. Recent work targets this "paradox of fast explanations", and suggests that explanatory reasoning, like other aspects of cognition, is subject to heuristics that yield rapid, but often inaccurate, hypotheses for observations.

George Miller had tentative (and accurate) conclusions about the state of reasoning research circa 1950. He wrote:

> "The importance of verbalization in thinking is an unsettled issue in psychology. … The most we can say is that many people converse with themselves, and if they are interrupted they will say they are thinking."
>
> (Miller, 1951, p. 804)

Plainly, an optimist should view recent advances with excitement. Scientists can now say much more about the underlying processes that lead from premises to conclusions. Consensus exists that reasoning in daily life diverges from the norms set by orthodox logic, and that many premises and conclusions are inherently uncertain. And a growing group of researchers views theoretical accounts as insufficiently constrained until and unless they are implemented computationally. A pessimistic outlook holds that the explosion of new theories produced an unweildy number of controversies for the reasoning community to tackle. A unified approach is needed, and extant debates must come to resolution. The not-so-hidden agenda of this chapter is to spur researchers to resolve existing controversies – or at the very least, discover new ones.

Definitions and Terms

Abduction

A process designed to infer a hypothesis about an observation or premise. Abductive reasoning is a form of inductive reasoning.

Deduction

A process designed to draw valid conclusions from the premises, i.e., conclusions that are true in any case in which the premises are true.

Default reasoning

Reasoning with assumptions that hold by default, but that can be overturned when new information is available.

Defective truth table

A truth table of a conditional assertion, "If A then C", that has no truth value when A is false. (Also known as the de Finetti truth table).

Fully explicit model

A fully explicit model is a mental representation of a set of possibilities that depicts whether each clause in a compound assertion is true or not. The fully explicit models of a disjunction, "A or B, but not both" represents two possibilities: the possibility in which A occurs and B does not, and the possibility in which B occurs and A does not.

Induction

A process designed to draw plausible, compelling, or likely conclusions from a set of premises. The conclusions drawn from an inductive inference are not always true in every case in which the premises are true.

Logic

The discipline that studies the validity of inferences, and the formal systems produced by the discipline. Many logical systems exist, and each system is built from two main components: proof theory and model theory.

Mental model

An iconic representation of a possibility that depicts only those clauses in a compound assertion that are possible. Hence, the mental model of a disjunctive assertion, "A or B, but not both" represents two possibilities: the possibility in which A occurs and the possibility in which B occurs.

Model theory (logic)

The branch of logic that accounts for the meaning of sentences in the logic and explains valid inferences.

Model theory (psychology)

The psychological theory that humans build iconic models of possibilities in order to think and reason.

Monotonicity

The property of many formal systems of logic in which the introduction of additional premises leads to additional valid inferences.

Non-monotonicity

The property of everyday reasoning (and some formal systems of logic) in which additional information can lead to the withdrawal of conclusions.

Probabilistic logic

A paradigm for reasoning that focuses on four hypotheses: Ramsey's test embodies conditional reasoning; truth tables for conditionals are defective; the probability of a conditional is equal to a conditional probability; and rational inferences are probabilistically valid.

Probabilistic reasoning

Reasoning about premises that are probabilistic, or else reasoning that produces probabilistic conclusions.

Probabilistic validity

P-valid inferences concern conclusions that are not more informative than their premises.

Proof theory

A branch of logic that stipulates the formal rules of inference that sanction the formulas that can be derived from other formulas. The system can be used to construct proofs of conclusions form a set of premises.

Ramsey test

A thought experiment designed to determine a degree of belief in a conditional assertion, “If A then C”. To carry out the experiment, a reasoner adds *A* to her set of beliefs and then assesses the likelihood of *C*.

Truth functional

A compound assertion, e.g., “If A then C”, is truth functional if its truth value is defined as a function of the truth of its constituent assertions, i.e., the truth of A and the truth of C.

Truth table

A systematic table that depicts the truth values of a compound assertion, such as a conjunction, that hold as a function of the truth values of its clauses.

Validity

A logical inference is valid if its conclusion is true in every case in which its premises are true.

References

Adams, E. W. (1975). *The logic of conditionals.* Dordrecht, the Netherlands: Reidel. doi:10.1007/978-94-015-7622-2

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

Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.). *Explanation and cognition* (pp. 199−225). Cambridge, MA: MIT Press.

Ahn, W., Kim, N.S., Lassaline, M.E., & Dennis, M.J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41,* 361–416.

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional reasoning: Mental models or causal models. *Cognition*,*119*, 403-418.

Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science*, *10*, 790-812.

Amini, M. (2003). Has foundationalism failed? A critical review of *Coherence in thought and action* by Paul Thagard. *Human Nature Review, 3,* 119–123.

Anderson, C.A., & Ross, L. (1980). Perseverence of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049.

Baggio, G., Lambalgen, M., & Hagoort, P. (2015). Logic as Marr's computational level: Four case studies. *Topics in Cognitive science*, *7*, 287-298.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences, 30,* 241–297.

Barnes, J. (Ed.) (1984). *The complete works of Aristotle.* Princeton, NJ: Princeton University Press.

Barrouillet, P., & Gauffroy, C. (2015) Probability in reasoning: a developmental test on conditionals. *Cognition*, 137, 22–39.

Barrouillet, P., Gauffroy, C., & Leças, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, *115*, 760-771.

Barrouillet, P., Grosset, N., and Leças, J. F. (2000) Conditional reasoning by mental models: chronometric and developmental evidence. *Cognition, 75,* 237-266.

Barwise, J. (1993). Everyday reasoning and logical inference. *Behavioral and Brain Sciences,* 16, 337–338.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*, 18327-18332.

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.

Besnard, D., & Bastien-Toniazzo, M. (1999). Expert error in trouble-shooting: An exploratory study in electronics. *International Journal of Human Computer Studies, 50,* 391-405.

Bibel, W. (2013). *Automated theorem proving*. Springer Science & Business Media.

Birch, S. A., & Bloom, P. (2003). Children are cursed an asymmetric bias in mental-state attribution. *Psychological Science*, *14*, 283-286.

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*, 1156-1164.

Bonnefon, J. F., & Sloman, S. A. (2013). The causal structure of utility conditionals. *Cognitive science*, *37*, 193-209.

Boolos, G., & Jeffrey, R. (1989). *Computability and Logic* (3rd edn). Cambridge University Press.

Brandone, A. C., Cimpian, A., Leslie, S. J., & Gelman, S. A. (2012). Do lions have manes? For children, generics are about kinds rather than quantities. *Child Development*, *83*, 423-433.

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

Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic.* Mahwah, NJ: Erlbaum.

Brockman, J., Ferguson, A., & Ford, M. (2014). *This explains everything*. New York: Harper Perennial.

Callanan, M. A. & Oakes, L. M. (1992). Preschoolers’ questions and parents’ explanations: Causal thinking in everyday activity. *Cognitive Development, 7,* 213–233.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press, Bradford Books.

Carlson, G., & Pelletier, F. J. (1995). *The generic book.* Chicago, IL: University of Chicago Press.

Carnap, R. (1962). *Logical foundations of probability (2nd Ed.).* Chicago, IL: Univeristiy of Chicago Press.

Case, K. E., Shiller, R. J., & Thompson, A. (2012). *What have they been thinking? Home buyer behavior in hot and cold markets* (No. w18400). National Bureau of Economic Research.

Ceraso, J., & Provitera, A. (1971). Sources of error in syllogistic reasoning. *Cognitive Psychology, 2,* 400–410.

Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology, 58,* 220–226.

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., & Vitanyi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Science, 7,* 9–22.

Cheng, P.W. (2000). Causal reasoning. In R. Wilson & F. Keil (Eds.), *The MIT Encyclopedia of Cognitive Sciences* (pp. 106−108). Cambridge, MA: MIT Press.

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive science*, *34*, 1452-1482.

Cimpian, A. & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*, 461-480.

Cimpian, A., & Steinberg, O. D. (2014). The inherence heuristic across development: Systematic differences between children’s and adults’ explanations for everyday facts. *Cognitive psychology*, *75*, 130-154.

Clark, H. H. (1975). Bridging. In *Proceedings of the 1975 workshop on Theoretical issues in natural language processing* (pp. 169-174). Association for Computational Linguistics.

Clark, H.H. (1996). *Using language.* London, UK: Cambridge University Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317-331.

Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science*, *13*, 1-49.

Corner, A., & Hahn, U. (2009). Evaluating science arguments: Evidence, uncertainty & argument strength. *Journal of Experimental Psychology: Applied*, 15, 199–212.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition, 58,* 1–73.

Craik, K. (1943). *The nature of explanation.* Cambridge, UK: Cambridge University Press.

Davis, M. (2000). *Engines of Logic: Mathematicians and the Origin of the Computer*. Norton.

de Finetti, B. (1995). The logic of probability (translation of 1936 original). Translated in R. B. Angell, The logic of probability. Philosophical Studies, 77, 181–190.

Dewar, K., & Xu, F. (2010). Induction, overhypothesis, and the origin if abstract knowledge: Evidence from 9-month-old infants. Psychological Science, 21, 1871–1877.

Doyle, J., & Wellman, M. P. (1991). Impediments to universal preference-based default theories. *Artificial Intelligence*, *49*, 97-128.

Einhorn, H. J. & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99,* 3-19.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning*, *19*, 249-265.

Elqayam, S., Thompson, V.A., Wilkinson, M.R., Evans, J.St.B.T., & Over, D.E. (2015). Deontic introduction: A theory of inference from Is to Ought. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*, 1516-1532.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Evans, J. St. B. T. (2007). *Hypothetical thinking.* Hove, UK: Psychology Press.

Evans, J. St. B.T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning,* 18, 5–31.

Evans, J. St. BT, Ellis, CE, & Newstead, SE (1996). On the mental representation of conditional sentences. *Quarterly Journal of Experimental Psychology A*, *49*, 1086-1114.

Evans, J. St. B. T., & Over, D. E. (2004). *If.* Oxford: Oxford University Press.

Evans, J. St. B. T., & Over, D. E. (2013). Reasoning to and from belief: Deduction and induction are still distinct. *Thinking & Reasoning, 19,* 267–283.

Fantino, E., Kulik, J., Stolarz-Fantino, S., & Wright, W. (1997). The conjunction fallacy: A test of averaging hypotheses. Psychonomic Bulletin and Review, 4, 96–101.

Fernbach, P. M., Macris, D. M., & Sobel, D. M. (2012). Which one made it go? The emergence of diagnostic reasoning in preschoolers. *Cognitive Development*, *27*, 39-53.

Florian, J. E. (1994). Stripes do not a zebra make, or do they: Conceptual and perceptual information in inductive inference. *Developmental Psychology, 30,* 88-101.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998.

Franklin, J. (2001). *The science of conjecture: Evidence and probability before Pascal.* Baltimore, MD: Johns Hopkins University Press.

Frosch, C.A., & Johnson-Laird, P.N. (2011). Is everyday causation deterministic or probabilistic? *Acta Psychologica*, *137*, 280–291.

Fugard, A. J., Pfeifer, N., Mayerhofer, B., & Kleiter, G. D. (2011). How people interpret conditionals: shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 635-648.

Geiger, S.M., & Oberauer, K. (2010). Toward a reconciliation of mental model theory and probabilistic theories of conditionals. In M. Oaksford and N. Chater (Eds.) *Cognition and Conditionals: Probability and Logic in Human Thinking* (pp. 289–307). Oxford, UK: Oxford University Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press, USA.

Gentner, D., & Stevens, A.L., Eds. (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M. A., & Kaschak, M. P. (2003). *Language comprehension.* John Wiley & Sons, Ltd.

Gerstenberg, T., Goodman, N., Lagnado, D., & Tenenbaum, J. (2012) Noisy Newtons: Unifying process and dependency accounts of causal attribution. In N. Miyake, D. Peebles, and R.P. Cooper (Eds.) *Proceedings of the 34th Conference of the Cognitive Science Society* (pp 378–383). Austin, TX: Cognitive Science Society.

Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, *28*, 843-900.

Gilio, A. (2012). Generalizing inference rules in a coherence-based probabilistic default reasoning. *International Journal of Approximate Reasoning*, *53*(3), 413-434.

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

Girotto, V., & Johnson-Laird, P.N. (2004) The probability of conditionals. *Psychologia*, 47, 207–225.

Glymour, C. (2001). *The mind's arrows.* Cambridge, MA: The MIT Press.

Goel, V., 2009. Cognitive Neuroscience of Thinking. In: Berntson, G., Cacioppo, J.T. (Eds.), *Handbook of Neuroscience for the Behavioral Sciences.* Wiley, New York

Goldvarg, Y., & Johnson-Laird, P. N. (2001). Naive causality: a mental model theory of causal meaning and reasoning. *Cognitive Science, 25,* 565-610.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*, 173-184.

Goodwin, G.P. (2014). Is the basic conditional probabilistic? *Journal of Experimental Psychology: General, 143,* 1214–1241.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: the function, evolution, and phenomenology of the theory-formation system. In F. Keil, F. and R. A. Wilson (Eds.) *Explanation and cognition.* Cambridge, MA: MIT Press.

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science*, *7*, 99-108.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences*, *14*(8), 357-364.

Hacking, I. (2006). *The emergency of probability* (2nd Ed.).Cambridge, UK: Cambridge University Press.

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

Handley, S. J., Evans, J. S. B., & Thompson, V. A. (2006). The negated conditional: A litmus test for the suppositional conditional?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 559-569.

Harman, G. H. (1965). The inference to the best explanation. *Philosophical Review, 74,* 88-95.

Harris, A. J. L., Hsu, A. S., & Madsen, J. K. (2012). Because Hitler did it! Quantitative tests of Bayesian argumentation using “ad hominem”. *Thinking and Reasoning*, *18,* 311–343.

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

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley interdisciplinary reviews: Cognitive science*, *1*, 278-292.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*, 280-285.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248-274). Oxford: Oxford University Press.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, *7*, 569-592.

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

Henle, M. (1978). Foreword to Revlin, R., and Mayer, R.E. (Eds.) *Human reasoning.* Washington, DC: Winston.

Horne, Z. (2017). *Inherent thinking in scientific explanation.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaigne, Champaign, IL.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children’s causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology, 5,* 668–683.

Hilton, D. J. & Erb, H. (1996). Mental models and causal explanation: judgments of probable cause and explanatory relevance. *Thinking & Reasoning, 2,* 273–308.

Holyoak, K.J. (2012) Analogy and relational reasoning. In K.J. Holyoak and R.G. Morrison (Eds.) *The Oxford Handbook of Thinking and* (pp. 234–259). Oxford, UK: Oxford University Press.

Hussak, L. J., & Cimpian, A. (2015). An early-emerging explanatory heuristic promotes support for the status quo.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge, Chapman & Hall.

Jeffrey, R. (1981). *Formal logic: Its scope and limits* (2nd Ed). New York: McGraw-Hill.

Johnson, S. G. B., Rajeev-Kumar, G., & Keil, F. C. (2014). Inferred evidence in latent scope explanations. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 707–712). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Rajeev-Kumar, G., & Keil, F. C. (2016). Sense-making under ignorance. *Cognitive Psychology.*

Johnson-Laird, P. N. (1975). Models of deduction. In R. Falmagne (Ed.), *Reasoning: Representation and process* (pp. 7–54). Springdale, NJ: Erlbaum.

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

Johnson-Laird, P. N. (2006). *How we reason.* Oxford, UK: Oxford University Press.

Johnson-Laird, P. N. (2010). Against logical form. *Psychologica Belgica,* 50, 193-221.

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

Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646-678.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111,* 640-661.

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory & Cognition, 31,* 1105–1113.

Johnson-Laird, P. N. & Khemlani, S. (forthcoming). Mental models and causation. In M. Waldmann (Ed.), *Oxford Handbook of Causal Reasoning*.

Johnson-Laird, P. N., Khemlani, S., & Goodwin, G. (2015). Logic, probability, and human reasoning. *Trends in Cognitive* Sciences, 19, 201-214.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62–88.

Johnson-Laird, P. N., Lotstein, M., & Byrne, R. M. J. (2012). The consistency of disjunctive assertions. *Memory & Cognition, 40,* 769–778.

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *The American Journal of Psychology*, *82*, 367-373.

Johnston, A.M., Johnson, S.G.B., Koven, M.L., & Keil, F.C. (in press). Little Bayesians or little Einsteins? Probability and explanatory virtue in children's inferences. *Developmental Science*.

Kassirer, J. P. (1989). Diagnostic reasoning. *Annals of Internal Medicine, 110,* 893-900.

Kindleberger, C. P. (1978). *Manias, panics and crashes: a history of financial crises*. Springer.

Keil, F. C. (2006). Explanation and understanding. *Annual review of psychology*, *57*, 227.

Kelemen, D. (1999). Function, goals, and intention: children's teleological reasoning about objects. *Trends in Cognitive Sciences, 3,* 461–468.

Kelemen, D., & DiYanni, C. (2005). Intuitions about origins: purpose and intelligent design in children's reasoning about nature. *Journal of Cognitive Development, 6,* 3–31.

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

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy, 30,* 1–45.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The ego-centric basis of language use: insights from a processing approach. *Current Directions in Psychological Science, 7,* 46– 50.

Khardon, R., & Roth, D. (1995). Default-reasoning with models. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 319-327).

Khemlani, S. (2016). Automating human inference. In U. Furbach and C. Shon (Eds.), *Proceedings of the 2nd IJCAI Workshop on Bridging the Gap between Human and Automated Reasoning* (pp. 1–4). CEUR Workshop Proceedings.

Khemlani, S., Barbey, A., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience*, 8, 849, 1-15.

Khemlani, S., Byrne, R.M.J, Goodwin, G., & Johnson-Laird, P.N. (under review). A unified theory of sentential reasoning about facts and possibilities. Manuscript under review.

Khemlani, S., & Johnson-Laird, P.N. (2009). Disjunctive illusory inferences and how to eliminate them. *Memory & Cognition, 37,* 615-623.

Khemlani, S. & Johnson-Laird, P.N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, 64, 2276-88.

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

Khemlani, S., & Johnson-Laird, P. N. (2013). The processes of inference. *Argument & Computation, 4,* 1-20.

Khemlani, S., & Johnson-Laird, P.N. (2015). Domino effects in causal contradictions. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Khemlani, S., & Johnson-Laird, P. N. (2016). How people differ in syllogistic reasoning. In A. Papafragou, D. Grodner, D. Mirman, and J. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Khemlani, S., & Johnson-Laird, P.N. (under review). Illusions in reasoning. Manuscript under review in *Minds and Machines.*

Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2012). Inferences about members of kinds: The generics hypothesis. *Language and Cognitive Processes*, 27, 887-900.

Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2015a). Naive probability: Model-based estimates of unique events. *Cognitive Science*, 39, 1216–1258.

Khemlani, S., Lotstein, M., Trafton, J.G., & Johnson-Laird, P. N. (2015b). Immediate inferences from quantified assertions. *Quarterly Journal of Experimental Psychology*, 68, 2073–2096.

Khemlani, S., Mackiewicz, R., Bucciarelli, M., & Johnson-Laird, P. N. (2013). Kinematic mental simulations in abduction and deduction. *Proceedings of the National Academy of Sciences*, 110, 16766-16771.

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, 24, 541–559.

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D.M . (2011). *Harry Potter* and the sorcerer's scope: Latent scope biases in explanatory reasoning. *Memory & Cognition, 39,* 527–535.

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. MIT Press.

Krauss, R. M., & Glucksberg, S. (1969). The development of communication: competence as a function of age. *Child Development, 40,* 255–66.

Kroger, J. K., Nystrom, L. E., Cohen, J. D., & Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. *Brain Research*, 1243, 86–103.

Kuhn, T. S. (1977). *Objectivity, value judgment, and theory choice. In the essential tension: Selected studies in scientific tradition and change.* Chicago: University of Chicago Press.

Lagnado, D. (1994). *The psychology of explanation: A Bayesian approach.* Masters Thesis. Schools of Psychology and Computer Science, University of Birmingham.

Lassiter, D., & Goodman, N. D. (2015). How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition*, *136*, 123-134.

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Development, 83,* 173-185.

Legrenzi, P., & Johnson-Laird, P. N. (2005). The evaluation of diagnostic explanations for inconsistencies. *Psychologica Belgica, 45,* 19-28.

Leslie, S. J. (2008). Generics: Cognition and acquisition. Philosophical Review, 117, 1–47.

Leslie, S.-J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language, 65,* 15–31.

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

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10,* 464-470.

Lombrozo, T. (2007). Simplicity and probability in causal explanations. *Cognitive Psychology, 55,* 232-257.

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

Lombrozo, T. & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99,* 167-204.

López, A. (1995). The diversity principle in the testing of arguments. *Memory & Cognition*, **23**, 374-382.

Mackay, C. (1869). *Memoirs of extraordinary popular delusions and the madness of crowds*. George Routledge and Sons.

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition?. *Psychological science*, *24*, 2351-2360.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

McRae, K., & Jones, M. N. (2013). Semantic memory. *The Oxford Handbook of Cognitive Psychology* (pp. 206-219). Oxford, UK: Oxford University Press.

McGill, A. L. (1990). Conjunctive explanations: The effect of comparison of the target episode to a contrasting background instance. *Social Cognition*, *8*, 362.

Medin, D, & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge, UK: Cambridge University Press.

Mehlhorn, K., Taatgen, N. A., Lebiere, C., & Krems, J. F. (2011). Memory activation and the availability of explanations in sequential diagnostic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1391.

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

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40-48.

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proceedings of the National Academy of Sciences*, *106*, 12554-12559.

Murphy, G. L. (2000). Explanatory concepts. In R. A. Wilson & F. C. Keil (Eds.), *Explanation and cognition* (pp. 361-392). Cambridge, MA: MIT Press.

Murphy, G.L. & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nickerson, R.S. (2001). The projective way of knowing. *Current Directions in Psychological Science, 10,* 168–72.

Nickerson, R. S. (2015). *Conditional reasoning: The unruly syntactics, semantics, thematics, and pragmatics of "If".* New York Oxford University Press.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339-363.

O'Brien, D. P. (2014). Conditionals and disjunctions in mental-logic theory: A response to Liu and Chou (2012) and to López-Astorga (2013). *Universum*, *29*, 221-235.

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.

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

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

Oaksford, M., & Chater, N. (2009). Precis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral & Brain Sciences, 32*, 69-120.

Oaksford, M., & Chater, N. (2010). Conditional inference and constraint satisfaction: Reconciling mental models and the probabilistic approach. *Cognition and conditionals: Probability and logic in human thinking*, 309-333.

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

Oaksford, M., & Hall, S. (2016). On the source of human irrationality. *Trends in Cognitive Sciences*, *20*, 336-344.

Oaksford, M., & Hahn, U. (2007). Induction, deduction, and argument strength in human reasoning and argumentation. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches* (pp. 269–301). Cambridge University Press.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 835-854.

Oberauer, K., & Wilhelm, O. (2003) The meaning(s) of conditionals: conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory, & Cognitoin, 29,* 688–693.

Osherson, D. N. (1974–1976). *Logical abilities in children* (Vols. 1–4). Hillsdale, NJ: Erlbaum.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.

Over, D. E., Hadjichristidis, C., Evans, J. S. B., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, *54*, 62-97.

Patalano, A. L., Chin-Parker, S., & Ross, B. H. (2006). The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of Memory and Language, 54,* 407-424.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge, UK: Cambridge University Press.

Peirce, C.S. (1931-1958). *Collected papers of Charles Sanders Peirce*. 8 vols. Hartshorne, C., Weiss, P., & Burks, A. (Eds.) Cambridge, MA: Harvard University Press.

Pelletier, F. J., & Elio, R. (2005). The case for psychologism in default and inheritance reasoning. *Synthese*, *146*, 7-35.

Pepperberg, I. M., Koepke, A., Livingston, P., Girard, M., & Hartsfield, L. A. (2013). Reasoning by inference: Further studies on exclusion in grey parrots (Psittacus erithacus). *Journal of Comparative Psychology, 127,* 272–281.

Politzer, G. (2004). Some precursors of current theories of syllogistic reasoning. In K. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (pp. 213–240). Hove, England: Psychology Press.

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

Prado, J., Chadha, A., & Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *Journal of Cognitive Neuroscience, 23,* 3483–3497.

Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*.

Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science, 33,* 401–448.

Prasada, S., Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition, 126,* 405-422.

Quelhas, A.C., & Johnson-Laird, P.N. (2016). The modulation of disjunctive assertions. *Quarterly Journal of Experimental Psychology*, in press.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological review*, *120*, 561-588.

Ramsey, F. P. (1990). General propositions and causality (originally published 1929). In D. H. Mellor (Ed.), *Philosophical Papers* (pp. 145–163). Cambridge: Cambridge University Press.

Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W.B. Rouse (Eds.), *Human Detection and Diagnosis of System Failures*. New York: Plenum Press.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition*, *91*(2), 113-153.

Reiter, R. (1978). On reasoning by default. In *Proceedings of the 1978 workshop on Theoretical issues in natural language processing* (pp. 210-218). Association for Computational Linguistics.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of verbal learning and verbal behavior*, *14*, 665-681.

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

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, *12*(2), 129-134.

Rips, L. J. (2002). Reasoning. In D. Medin (Ed.), *Stevens' Handbook of Experimental Psychology: Vol. 2. Memory and cognitive processes* (3rd ed., pp. 317–362). New York, NY: Wiley.

Ross, L., Lepper, M. R., Strack, F., Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology, 35,* 817-829.

Rouse, W. B. & Hunt, R. M. (1984). Human problem solving in fault diagnosis tasks. In W.B. Rouse (Ed.), *Advances in Man-machine Systems Research*. Greenwich, CT: JAI Press.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*, 883-893.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, *120*, 411.

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

Schlottman, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 44(A),* 321–342.

Schroyens, W., Schaeken, W., & Dieussaert, K. (2008). "The" interpretation (s) of conditionals. *Experimental Psychology*, *55*, 173-181.

Selman, B., & Kautz, H. (1989). The complexity of model-preference default theories. In *Non-Monotonic Reasoning* (pp. 115-130). Springer Berlin Heidelberg.

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

Shafto, P. & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: novices to experts, naïve similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29,* 641-649.

Skyrms, B. (1986). *Choice and chance: An introduction to inductive logic* (3rd ed.). Belmont, CA: Wadsworth.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231-280.

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition, 52,* 1-21.

Sloman, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking & Reasoning, 2,* 81-110.

Sloman, S. A. (2005). *Causal models: How we think about the world and its alternatives*. New York: Oxford University Press.

Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition, 49,* 67-96.

Steinberg, D.D. (1970). Analyticity, amphigory, and the semantic interpretation of sentences. *Journal of Verbal Learning and Verbal Behavior*, *9*, 37-51.

Steinberg, D.D. (1975) Semantic universals in sentence processing and interpretation: A study of Chinese, Finnish, Japanese, and Slovenian speakers. *Journal of Psycholinguistic Research*, *4*, 169-193.

Stenning, K., & van Lambalgen, M. (2016). Logic programming, probability, and two-system accounts of reasoning: a rejoinder to Oaksford and Chater (2014). *Thinking & Reasoning*, *22*, 355-368.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*, 453-489.

Sutherland, S. L., & Cimpian, A. (2015). An explanatory heuristic gives rise to the belief that words are well suited for their referents. *Cognition*, *143*, 228-240.

Sussman, A. B., Khemlani, S. S., & Oppenheimer, D. M. (2014). Latent scope bias in categorization. *Journal of Experimental Social Psychology, 52,* 1–8.

Selman, B., & Kautz, H. (1989). The complexity of model-preference default theories. In Reinfrank et al. (Eds.), *Non-monotonic reasoning* (pp. 115-130). Springer Verlag: Berlin.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The Rationality Quotient: Toward a Test of Rational Thinking*. MIT Press.

Störring, G. (1908). Experimentelle Untersuchungen über einfache Schlussprozesse [Experimental investigations of simple inference processes]. *Archiv für die gesamte Psychologie*, *11*, 1–27.

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.

Thagard, P. (1992). *Conceptual revolutions.* Princeton, NJ: Princeton University Press.

Thielscher, M., & Schaub, T. (1995). Default reasoning by deductive planning. *Journal of Automated Reasoning*, *15*, 1-40.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. (2008). Diagnostic hypothesis generation and human judgment. *Psychological review*, *115*, 155-185.

Turing, A.M. (1937). On computable numbers, with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, 42, 230-265.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567

Van Overwalle, F. (1998). Causal explanation as constraint satisfaction: A critique and a feedforward connectionist alternative. *Journal of Personality and Social Psychology*, *74*, 312.

Waldmann, M. (in press). (Ed.) *The Oxford Handbook of Causal Reasoning.* Oxford, UK: Oxford University Press.

Walker, C.M., Bonawitz, E., & Lombrozo, T. (in press). Effects of explaining on children's preference for simpler hypotheses. Manuscript in press at *Psychonomic Bulletin & Review.*

Wallsten, T. S., Budescu, D. V., Erev, I., & Diederich, A. (1997). Evaluating and combining subjective probability estimates. *Journal of Behavioral Decision Making, 10,* 243–268.

Wallsten, T. S., Budescu, D. V., & Tsao, C. J. (1997). Combining linguistic probabilities. *Psychlogische Beiträge, 39,* 27–55.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.

Wason, P. C., & Brooks, P. G. (1979). THOG: The anatomy of a problem. *Psychological Research*, *41*, 79–90.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Cambridge, MA: Harvard University Press.

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

Wellman, H.M., Hickling, A.K., & Schult, C.A. (1997). Young children’s psychological, physical, and biological explanations. *New Directions for Child Development, 75,* 7-25.

Wellman, H.M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik and L.E. Schulz (Eds.) *Causal learning: Psychology, philosophy, and computation.* Oxford, UK: Oxford University Press.

Whewell, W. (1840). *The philosophy of the inductive sciences: Founded upon their history.* London: John W. Parker.

White, P.A. (1999). Toward a causal realist account of causal understanding. *American Journal of Psychology, 112,* 605–642.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136*, 82−111.

Zemla, J., Sloman, S., & Lagnado, D. (in press). Evaluating everyday explanations. Manuscript in press at *Psychonomic Bulletin & Review.*

Zhao, J., Shah, A., & Osherson, D. (2009). On the provenance of judgments of conditional probability. *Cognition, 113,* 26–36.

Figures and Tables

Figure 1. Percentages of various responses to the question “What proportion of As are Cs?” (panel A) and “What proportion of As must be Cs?” (panel B) for basic and probabilistically marked conditionals (reproduced from Goodwin, 2014, Figure 1 and 2, with permission).

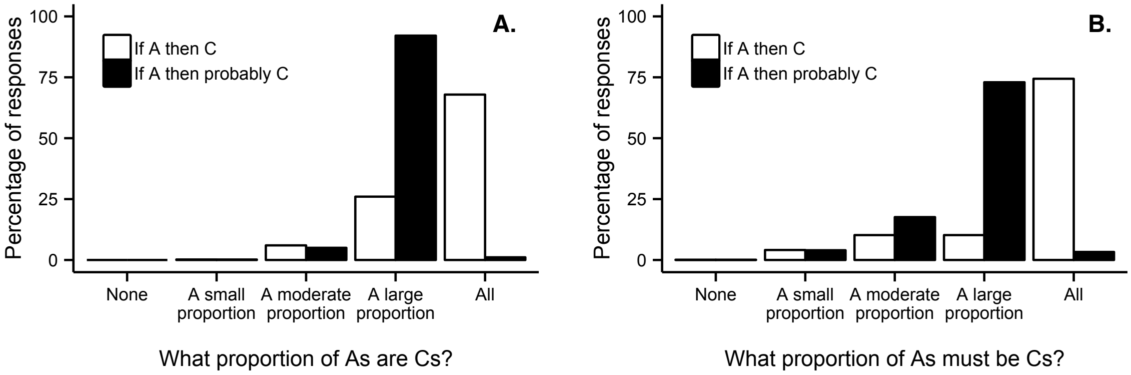

Table 1. Three overarching frameworks of rationality, their primary principles and hypotheses, the kinds of representational structures on which proposed mechanisms operate, the manner in which content and background knowledge affect the mechanisms, and the accounts of validity the frameworks espouse.

Framework of rationality	**Principle**	**Central hypotheses**	**Structure**	**Content**
Mental logic	People manipulate propositional representations of beliefs and facts by applying syntactic transformations (rules of inference) to those propositions. The mind is equipped with a finite set of such rules, which dictate whether new propositions can be introduced or old ones eliminated.	• The more rules of inference it takes to solve a problem, the harder the problem • People make unsystematic errors in reasoning when they fail to apply logical rules of inference	Propositions and proofs	Meaning postulates that axiomatize various domains of inference, e.g., spatial and causal reasoning
Probabilistic logic	People rarely hold any belief with absolute certainty, and so uncertainty is present in all scientific reasoning and decision making. The probability calculus and its identities (e.g., Bayes's rule) serves as a mathematical account of uncertainty, and it is central to understanding reasoning.	• Subjective probabilities are an index of belief strength • Reasoners apply the Ramsey test to assess conditionals • The probability of a conditional is its conditional probability • Conditionals have a defective (de Finetti) truth table	Conditional probabilities; Bayesian networks	Prior probability distributions that represent belief strength, posterior probabilities
Mental models	People mentally simulate the world when they reason. The more simulations ("models") they consider, and the richer those models are, the more accurate their responses are. Humans are rational in principle, but err when they fail to consider possibilities.	• People make systematic errors in reasoning • They correct errors by considering counterexamples in which the premises are true but the conclusion false • The more models it takes to solve a problem, the harder the problem	Models, i.e., iconic simulations of possibilities	The relational structure inherent within models; models in the form of background knowledge that eliminate possibilities and introduce relations

Table 2. Semantic definitions of logical connectives for two sentences, *A* and *C,* in formal logic, probability logic, and mental model theory. The first column illustrates the four states of affairs that can occur with *A* and *C* on separate rows. For example, the first row depicts the situation in which *A* and *C* are both true. The rest of the columns illustrate how various connectives are defined relative to the four contingencies. Proponents of probabilistic logic import many of the assumptions of orthodox logic, but interpret conditionals as having a "defective" truth table, i.e., one that has no truth values when the antecedent of a conditional, *A*, is false. Proponents of mental model theory interpret the four states of affairs as possibilities, not truth values.

	Formal logic				Probabilistic logic	Mental models	
	Conjunction	*Inclusive disjunction*	*Negation*	*Material conditional*	*Defective conditional*	*Conjunction*	*Basic conditional*
Contingency	*A* & *C*	*A* v *C*	¬*A*	*A* → *C*	If *A* then *C*	*A* and *C*	If *A* then *C*
A and C	True	True	False	True	True	Possible	Possible
A and not C	False	True	False	False	False	Impossible	Impossible
Not A and C	False	True	True	True	No truth value	Impossible	Possible
Not A and not C	False	False	True	True	No truth value	Impossible	Possible

Table 3. Summary of robust phenomena of inductive arguments that increase the propensity to generalize a given property. The table lists each phenomenon and its description, provides an illustrative example, and in brackets, provides a contrasting example of a situation that violates the phenomenon.

Phenomenon	The propensity to make an inductive inference increases when…	Example inference [and contrast]	Representative citation
Similarity	…the category of the premise is similar to the category of a conclusion	Rabbits have property X. Therefore, dogs [bears] have property X.	Florian (1994) Rips (1975)
Typicality	…the premise category is a more typical member of its superordinate category	Bluejays [geese] have property X. Therefore, Ys have property X.	Osherson et al. (1990) Rips (1975)
Variability	…there is less variance in the behavior of the premise category (i.e., it is more homogeneous)	One sample of a chemical [member of a tribe] has property X. Therefore, all instances of the chemical [members of the tribe] have property X.	Nisbett et al. (1983)
Sample size	…more instances of the premise category exhibit the property (interacts with variability effect)	Five members of a tribe [one member of a tribe] have property X. Therefore, all members of the tribe have property X.	Nisbett et al. (1983) Osherson et al. (1990)
The inclusion fallacy	…the premise and conclusion categories are similar regardless of violations of probability theory	Robins have property X. Therefore, all birds have property X. [Therefore, ostriches have property X.]	Osherson et al. (1990) Sloman (1993)
Diversity	…the categories in the premises are more diverse from one another, and the conclusion category is a superordinate	Horses, seals, and squirrels [horses, cows, and rhinos] have property X. Therefore, mammals have property X.	Carey (1985) López (1995) (but cf. Osherson et al, 1990; Sloman, 1993)
Explanations	…an explanation of the premise accords with an explanation of the conclusion	Many ex-cons are hired as bodyguards [unemployed]. Therefore, many war veterans are hired as bodyguards [unemployed].	Shafto & Coley (2004) Sloman (1994, 1997)

Table 4. Summary of formal models of inductive reasoning.

Model	Citation	Description
Similarity coverage model	Osherson et al. (1990)	Inductive strength is a function of the similarity of the conclusion to the premises as well as the "coverage", i.e., the ratio between the size of the category described by the premises and the size of the lowest possible superordinate category that includes both the premise and conclusion categories.
Feature-based model	Sloman (1993)	Inductive strength is a function of the amount of feature overlap between the premise and conclusion categories.
Gap model	Smith et al. (1993)	For blank features, inductive strength follows Osherson et al. (1990); for familiar features, strength is a function of both similarity and a "gap", i.e., a threshold initially derived from background knowledge that designates whether a conclusion category possesses the feature. The premise may shift the threshold upwards or downwards.
Bayesian models	Heit (1998) Kemp & Tenenbaum (2009)	Premises are treated as evidence, which is used to carry out Bayesian inference over prior beliefs to estimate the probability of the conclusion. Heit (1998) described the mathematical formalism of the inference but did not specify the representation of prior probabilities; Kemp and Tenenbaum (2009) extended the formalism to operate on prior probabilities derived from structured background knowledge.

Table 5. Various types of preferences for explanations, their descriptions, and empirical studies of the preference.

Factor	Description	Empirical studies
Simplicity	Simple explanations are those that concern relatively fewer causal relations and mechanisms than more complex alternatives.	(Bonawitz & Lombrozo, 2012; Lagnado, 1994; Lombrozo, 2007; Walker, Bonawitz, & Lombrozo, in press)
Coherence	Coherent explanations depend on causal relations from background knowledge (external coherence) or else those relevant to other links in the proposed causal mechanism (internal coherence). They avoid ad hoc causal relations.	(Read & Marcus-Newhall, 1993)
Completeness	Complete explanations posit causal mechanisms and relations that satisfy reasoners' expectations. Incomplete explanations leave expected causal relations unspecified.	(Johnson-Laird et al., 2004; Khemlani & Johnson-Laird, 2011; Legrenzi & Johnson-Laird, 2005; Zemla et al., in press)
Relevance	Relevant explanations provide information pertinent to a conversation, i.e., from the same domain of discourse or, via analogy, from a different domain as in with a similar relational structure. Irrelevant explanations provide too much or too little information, or else a relational structure that fails to map coherently to the domain of discourse.	(Hilton, 1996; McGill, 1990)
Latent scope	Explanations that explain many different observed phenomena have broad manifest scope. Explanations that do not explain unknown, uncertain, or unobserved phenomena have narrow latent scope.	(Johnson et al., 2014, 2016; Johnston et al., in press; Khemlani et al., 2011; Sussman et al., 2014)