Feature Review

CellPress

Logic, probability, and human reasoning

P.N. Johnson-Laird[1,2], Sangeet S. Khemlani[3], and Geoffrey P. Goodwin[4]

[1] Princeton University, Princeton, NJ, USA
[2] New York University, New York, NY, USA
[3] Naval Research Laboratory, Washington, DC, USA
[4] University of Pennsylvania, Philadelphia, PA, USA

This review addresses the long-standing puzzle of how logic and probability fit together in human reasoning. Many cognitive scientists argue that conventional logic cannot underlie deductions, because it never requires valid conclusions to be withdrawn – not even if they are false; it treats conditional assertions implausibly; and it yields many vapid, although valid, conclusions. A new paradigm of probability logic allows conclusions to be withdrawn and treats conditionals more plausibly, although it does not address the problem of vapidity. The theory of mental models solves all of these problems. It explains how people reason about probabilities and postulates that the machinery for reasoning is itself probabilistic. Recent investigations accordingly suggest a way to integrate probability and deduction.

The nature of deductive reasoning

To be rational is to be able to make deductions – to draw valid conclusions from premises. A valid conclusion is one that is true in any case in which the premises are true [1]. In daily life, deductions yield the consequences of rules, laws, and moral principles [2]. They are part of problem solving, reverse engineering, and computer programming [3–6] and they underlie mathematics, science, and technology [7–10]. Plato claimed that emotions upset reasoning. However, individuals in the grip of moderate emotions, even those from illnesses such as depression or phobia, reason better than controls, although only about matters pertaining to their emotion [11,12]. Deductive reasoning is an ability that varies vastly from one person to another, correlating with their intelligence and with the processing capacity of their working memory [13–15]. Our topic is the cognitive foundation of deductive reasoning, and we ask two fundamental questions:

(i) Does deduction depend on logic [16–20]?
(ii) How does deduction fit together with probabilities?

The first question is timely because of proposals that probability is the basis of human reasoning [21–23]. The second question has engaged theorists from the economist John Maynard Keynes [24] onward. Here we address both questions. We begin with logic (see Glossary) and present the arguments that logic alone cannot characterize deductive competence. These arguments motivated the turn to probability – a pivot that its proponents refer to as the 'new paradigm' [25–29]. Next, we outline the theory of mental models [30–34], which combines set theory with psychological principles. Finally, we present recent studies that suggest how to integrate deduction and probability.

Problems for logic as a theory of deductive reasoning

An ancient proposal is that deduction depends on logic (see also [16–20]). Sentential logic concerns inferences from premises such as conjunctions ('and') and disjunctions ('or'). Like most logics, it has two parts: proof theory and model theory [35]. Proof theory contains formal rules of inference for proofs. One major rule of inference in most formalizations is:

$A \rightarrow C$
A
therefore, C

where A and C can be any sentences whatsoever, such as: '2 is greater than 1' $\rightarrow$ '1 is less than 2'.

Proof theory specifies rules containing logical symbols such as $\rightarrow$, but not their meanings. Model theory defines their meanings. It specifies the truth of simple sentences such as '2 is greater than 1' with respect to a model, such as the natural numbers, and the truth of compound sentences containing connectives, such as $\rightarrow$, which is known as material implication. The meaning of $A \rightarrow C$ is defined as follows: it is true in any case except when A is true and C is false [1] and so it is analogous to 'if A then C'. This definition can be summarized in a truth table (Table 1). Model theory therefore determines the validity of inferences: a valid inference is one in which the conclusion is true in all cases in which the premises are true.

Logic is extraordinarily powerful and underlies the theory of computability [35–37]. Many cognitive scientists have accordingly supposed that human reasoning depends on unconscious formal rules of inference [16–20]. The hypothesis is plausible, but it runs into three difficulties.

First, conventional logic is monotonic; that is, if an inference is valid, its conclusion never needs to be withdrawn – not even when a new premise contradicts it. A contradiction validly implies any conclusion whatsoever [1]. However, human reasoners faced with a solid fact tend

Corresponding author: Johnson-Laird, P.N. (phil@princeton.edu).

1364-6613/
 http://dx.doi.org/10.1016/j.tics.2015.02.006

Glossary

Bayesian net: a directed graph in which each node represents a variable and arrows from one node to another represent conditional dependencies. It captures the complete joint probability distribution in a parsimonious way.
Consistency: a set of assertions is consistent if they can all be true at the same time.
Counterexample: in an inference, a possibility to which the premises refer but which is inconsistent with the conclusion.
Deductive reasoning: a process designed to draw a conclusion that follows validly from premises; that is, the conclusion is true in any case in which the premises are true.
Defeasible logics: also known as 'non-monotonic' logics. Unlike conventional logic, they allow conclusions to be weakened or withdrawn in the face of facts to the contrary.
Defective truth table: a truth table for a conditional, 'if *A* then *C*', that has no truth value when *A* is false (also known as the de Finetti truth table).
The Equation: the probability of a conditional, 'if *A* then *C*', equals the conditional probability of '*C* given *A*'.
Fully explicit model: unlike a mental model, it represents a possibility depicting each clause in the premises as either true or not. The fully explicit models of a disjunction, '*A* or *B* but not both', accordingly represent a conjunction of two possibilities: possibly(*A* & not-*B*) & possibly(not-*A* & *B*).
Kinematic model: a mental model that unfolds in time to represent a temporal succession of events.
Logic: the discipline that studies the validity of inferences. There are many logics, normally comprising two main components: proof theory, which stipulates rules for the formal derivation of proofs; and model theory, which is a corresponding account of the meanings of logical symbols and of the validity of inferences. In sentential logic, each proof corresponds one to one with a valid inference, but for other, more powerful logics not every valid inference can be proved.
Logical form: the structure of a proposition that dovetails with the formal rules of inference in a logic. No computer program exists to recover the logical form of propositions in daily life.
Material implication: a compound assertion in logic whose truth table is presented in Table 1 in main text. It is sometimes taken to correspond to a conditional, 'if *A* then *C*'. This view leads to logically valid but unacceptable 'paradoxes' such as that *C* implies 'if *A* then *C*'.
Mental model: an iconic representation of a possibility that depicts only those clauses in a compound assertion that are true. The mental models of a disjunction, '*A* or *B* but not both' accordingly represent two possibilities: possibly(*A*) and possibly(*B*).
Model theory: the component of a logic that accounts for the meaning of sentences in the logic and for valid inferences.
Modulation: the process in the construction of models in which content, context, or knowledge can prevent the construction of a model and can add information to a model.
Monotonicity: the property in conventional logic in which further premises to those of a valid inference yield further conclusions.
New paradigm: see probabilistic logic.
Probabilistic logic (p-logic): a paradigm for reasoning that focuses on four hypotheses: Ramsey's test, the defective truth table, the Equation, and p-validity.
Proof theory: the branch of a logic that provides formal rules of inference that can be used in formal proofs of conclusions from premises.
P-validity: an inference is p-valid if its conclusion is not more informative than its premises.
Ramsey's test: to determine your degree of belief in a conditional assertion, add its if-clause to your beliefs and assess the likelihood of its then-clause.
Recursive process: a loop of sequential operations performed either for a predetermined number of times or while a particular condition holds. If it has to be conducted an indefinite number of times, as in multiplication, it needs a working memory to hold intermediate results.
Syllogism: a form of inference that Aristotle formulated based on two premises and a conclusion, which each contain a single quantifier, such as 'all *A*', 'no *A*', or 'some *A*'.
Systems 1 and 2: the two systems of reasoning postulated in dual-process theories of judgment and reasoning, in which system 1 yields rapid intuitions and system 2 yields slower deliberations. Many versions of the theory exist.
Truth table: a systematic table showing the truth values of a compound assertion, such as a conjunction, as a function of the truth values of its clauses.
Validity: in logic, an inference is valid if its conclusion is true in every case in which its premises are true. In everyday reasoning, its premises should also be true in every case in which its conclusion is true.
Vapid deductions: valid inferences that yield useless conclusions, such as the conjunction of a premise with itself.

to withdraw any conclusion that it contradicts. Some theorists therefore defend so-called 'non-monotonic' or 'defeasible' logics developed in artificial intelligence, which allow conclusions to be withdrawn [38–42].

Second, conditional assertions (e.g., 'If she insulted him then he's angry') occur in all sorts of reasoning. However, they do not correspond to any connective in sentential logic. Theorists have treated them as material implications [16–19], but this interpretation yields deductions of the following sort:

He's angry.
Therefore, if she insulted him then he's angry.

As the truth table for $A \rightarrow C$ shows (Table 1), whenever C is true, the material implication is true. The preceding inference is therefore valid on this interpretation. It is also valid to infer a material implication from the falsity of A; for example:

She didn't insult him.
Therefore, if she insulted him then he's angry.

However, people usually reject both sorts of inference [43], which are called the 'paradoxes' of material implication. They are a major motivation for alternative foundations for human reasoning [21–23,25–29].

Third, logic yields infinitely many valid conclusions from any set of premises but many of them are vapid, such as a conjunction of the same premise with itself some arbitrary number of times; for example, 'A, therefore, A and A and A'. Logic alone cannot characterize sensible inferences [8,30,31]. Psychological theories based on logic therefore resort to extralogical methods to prevent vapid inferences [18–20]. No one knows to what degree these methods work without preventing useful inferences.

A further practical difficulty is that formal rules of inference apply, not to sentences, but to logical forms that match those of the formal rules of inference. No computer program exists for extracting logical forms from sentences in natural language, let alone from the propositions that sentences express in context. No one knows in full how to identify these forms from their shadows cast in sentences [44].

Probability logic

As a consequence of the preceding arguments, some cognitive scientists propose that probability should replace logic. Their theories differ in detail but overlap enough to have a label in common – the new paradigm [25–29]. We refer to the paradigm as 'probability logic' or 'p-logic' for short. It presupposes that degrees of belief correspond to subjective probabilities [45–49], an idea that not all psychologists accept [50,51]. It focuses on conditionals, and one p-logician even allows that conventional logic could apply to other sorts of assertion [47]. P-logic's proponents engage with four main hypotheses.

First, individuals fix their degree of belief in a conditional, using Ramsey's test [45]. To assess, say, 'If she insulted him then he's angry', they add the content of the if-clause (she insulted him) to their beliefs and then assess the likelihood of the then-clause (he's angry).

Second, Ramsey's test or an analogous concept of a conditional event [46] defines the conditions in which a conditional is true or false. As Table 1 shows, they yield