Psychological Bulletin
2012, Vol. 138, No. 3, 427–457

0033-2909/12/$12.00 DOI: 10.1037/a0026841

Theories of the Syllogism: A Meta-Analysis

Sangeet Khemlani and P. N. Johnson-Laird
Princeton University

Syllogisms are arguments about the properties of entities. They consist of 2 premises and a conclusion, which can each be in 1 of 4 "moods": *All A are B*, *Some A are B*, *No A are B*, and *Some A are not B*. Their logical analysis began with Aristotle, and their psychological investigation began over 100 years ago. This article outlines the logic of inferences about syllogisms, which includes the evaluation of the consistency of sets of assertions. It also describes the main phenomena of reasoning about properties. There are 12 extant theories of such inferences, and the article outlines each of them and describes their strengths and weaknesses. The theories are of 3 main sorts: heuristic theories that capture principles that could underlie intuitive responses, theories of deliberative reasoning based on formal rules of inference akin to those of logic, and theories of deliberative reasoning based on set-theoretic diagrams or models. The article presents a meta-analysis of these extant theories of syllogisms using data from 6 studies. None of the 12 theories provides an adequate account, and so the article concludes with a guide—based on its qualitative and quantitative analyses—of how best to make progress toward a satisfactory theory.

Keywords: logic, mental models, reasoning, rules of inference, syllogisms

Consider this inference:

> In some cases when I go out, I am not in company.
> Every time I am very happy I am in company.
> Therefore, in some cases when I go out, I am not very happy. (1)

It is an example of a syllogism, though it is not in the usual formal dress used in textbooks of logic. It is also *valid*, where "a valid inference is one whose conclusion is true in every case in which all its premises are true" (Jeffrey, 1981, p. 1). Among Western thinkers, Aristotle was the first to analyze syllogisms, and they were central to logic until the second half of the 19th century. Scholastic logicians thought that almost all arguments purporting to be logical could be expressed in syllogisms, and this tradition continued into the 20th century. The BBC once broadcast a monastic debate, and the monks were most adept at formulating their arguments in syllogisms. Even though modern logicians relegate syllogisms to one small part of logic, psychological studies of reasoning with determiners, such as *some* and *all*, have almost all concerned syllogistic reasoning.

The ability to reason is at the core of human mentality, and many contexts in daily life call for inferences, including decisions about goals and actions; the evaluation of conjectures and hypotheses; the pursuit of arguments and negotiations; the assessment of evidence and data; and above all science, technology, and culture. Reasoning based on quantifiers enters into all these pursuits, as the following sorts of assertion illustrate:

> Any experiment containing a confound is open to misinterpretation.
> No current word processor spontaneously corrects a user's grammar.
> Every chord containing three adjacent semitones is highly dissonant. (2)

If such premises are combined with others, they validly imply various sorts of conclusion. In daily life, individuals reason in a variety of contexts, and often so rapidly that they are unaware of having made an inference. Consider the discussion in the following example:

> Speaker 1: If you drop this cup it'll break.
> Speaker 2: It looks pretty solid to me.
> Speaker 1: Yes, but it's made from porcelain. (3)

The first speaker has made a tacit inference from the beliefs that porcelain is fragile and that fragile things break if they are dropped. The context of a potential inference affects the likelihood with which it will be made. For instance, when individuals reason about a matter that elicits a moderate emotion, they tend to outperform individuals who are not in an emotional state, including their own performance on matters that do not elicit an emotion (Johnson-Laird, Mancini, & Gangemi, 2006). In psychological experiments, participants are happy to make inferences about hypothetical individuals. One exception occurs with people from a

This article was published Online First January 30, 2012.

Sangeet Khemlani, Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, Washington, D.C.; P. N. Johnson-Laird, Department of Psychology, Princeton University.

This research was supported by a National Science Foundation Graduate Research Fellowship to Sangeet Khemlani and by National Science Foundation Grant SES 0844851 to P. N. Johnson-Laird to study deductive and probabilistic reasoning. We are grateful to the following individuals for taking the time to correct our accounts of their theories, where necessary, and for other helpful comments: Nick Chater and Mike Oaksford (probability heuristics), Marilyn Ford (verbal substitution), Bart Geurts (the monotonicity theory), Ken Gilhooly (matching), Thad Polk (the verbal reasoning model), Russ Revlin (illicit conversion), Lance Rips (the PSYCOP model), and Keith Stenning (the source-founding model). We thank Hua Gao, Catrinel Haught, Niklas Kunze, Max Lotstein, Gorka Navarrete, Janani Prabhakar, and Marco Ragni for their help and criticisms.

Correspondence concerning this article should be addressed to Sangeet Khemlani, Intelligent Systems Section, Code 5515, Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC 20375. E-mail: skhemlani@gmail.com

nonliterate culture, who are biased to make inferences from personal knowledge, but who will reason hypothetically provided that no one can have the relevant knowledge, such as inferences about a remote planet (Dias, Roazzi, & Harris, 2005).

Some psychologists demand that studies of reasoning should be ecologically valid: "Those who study first-order logic or variants thereof, such as mental rules and mental models, ignore the ecological and social structure of environments" (Hertwig, Ortmann, & Gigerenzer, 1997, pp. 105–106). In fact, abstract mathematics and logic are often outside the ecological and social context of environments. Much the same applies to Sudoku puzzles, which are popular in both East Asia and the Western world. Their solution depends on pure deduction—arithmetical calculations play no part—and on understanding the quantified principle that every row, column, and member of each of the nine 3×3 groups of cells contains each of nine digits. The puzzles derive from Euler's study of Latin squares, and their domain is open-ended (i.e., the puzzles can be larger than 9×9 squares) and computationally intractable (see Lee, Goodwin, & Johnson-Laird, 2008). The puzzles have no linguistic context beyond the statement of the puzzle itself, and they have no ecological validity either for us or for our protohominid ancestors. They show, however, that logically untrained individuals can make deductions from triply quantified assertions, and their popularity establishes that some people not only exercise logical ability in abstract problems, but also enjoy doing so. Without this ability, as Piaget realized, it is impossible to account for the development of logic and mathematics (see, e.g., Beth & Piaget, 1966).

Another context in which logically untrained individuals are called on to reason explicitly is when taking a test, such as the Scholastic Assessment Test or the Graduate Record Examination. And, of course, some people have to reason in the psychological laboratory as participants in studies of reasoning. In such cases, the instructions often call for them to draw conclusions that follow of necessity from the premises (see, e.g., Table 5). Most people have little difficulty in understanding the task, and even children of 9 years of age are able to make appropriate inferences (Johnson-Laird, Oakhill, & Bull, 1986). Undoubtedly, the instructions and other aspects of an experiment, including the social context that it creates, are likely to affect performance. But perhaps remarkably (as we will show), studies of syllogistic reasoning in the psychological laboratory yield highly correlated results. It seems that if individuals are allowed to think for themselves in relative quiet, without interruption, and with no undue stress or extraneous emotion, they are able to make a wide variety of systematic deductive inferences.

The first psychological study of syllogisms was carried out over 100 years ago: Störring (1908) studied the introspections of four participants tackling various inferences, including some syllogisms. Thirty years ago the task of understanding syllogistic reasoning seemed to be a good test for the feasibility of cognitive science (Johnson-Laird, 1983, Chapter 2): Syllogisms are a small closed set of inferences that lend themselves to psychological experimentation. The task of explaining their psychology appeared both feasible and worthwhile, because theories then had described the factors vitiating reasoning rather than underlying mental processes. If psychologists could agree on an adequate theory of syllogistic reasoning, then progress toward a more general theory of reasoning would seem to be feasible. If, however, researchers were unable to account for syllogistic reasoning, then they would have little hope of making sense of reasoning in general. This argument seems pertinent today.

There are now 12 main sorts of theory of syllogistic reasoning and reasoning based on *monadic* premises, that is, premises in which assertions assign only properties to individuals. One reason for the plethora of theories is that theorists have different goals. Early theorists, for instance, presupposed some sort of deductive mechanism and sought only to explain causes of error (e.g., Chapman & Chapman, 1959; Woodworth & Sells, 1935). Current theorists aim to characterize heuristics that yield conclusions and disregard logical validity (e.g., Chater & Oaksford, 1999). Other theories provide an account of logical reasoning with less regard for the intuitive responses that naive individuals make (e.g., Johnson-Laird & Steedman, 1978; Rips, 1994). Likewise, some accounts of logical reasoning focus on the evaluation of monadic conclusions within the scope of the first-order predicate calculus—the logic in which quantifiers, such as "some artists," range only over individuals (e.g., Rips, 1994). But other theories consider a broader variety of monadic inferences, such as the following example:

More than half the people in the room are English.
More than half the people in the room are American.
∴ At least someone in the room is both English and American. (4)

Quantifiers, such as "more than half the people," cannot be represented in first-order logic (Barwise & Cooper, 1981), because they call for quantification over sets rather than individuals. Hence, theories that invoke only standard first-order logic cannot hope to explain these inferences, or any that hinge on related determiners, such as "more than a third," "more than a quarter," "more than a fifth," etc.

The existence of 12 theories of any scientific domain is a small disaster. If psychologists cannot converge on a single theory of monadic reasoning, then the last 30 years of research have failed or at best made so little progress that skeptics may think that cognitive science itself is not feasible. A major challenge is therefore to develop a single comprehensive theory of monadic reasoning. Our aim in what follows is to make progress toward such a theory. We present a summary of robust data on syllogistic reasoning—a body of results that provide a touchstone for any putative theory. We then evaluate, where possible, each of the existing theories in relation to these data. The results show that the theories differ in the accuracy of their predictions, but that none of them does an adequate job in explaining syllogistic reasoning. We accordingly conclude with a synopsis of what needs to be done in order to establish a satisfactory theory.

Toward a Unified Theory of Syllogistic Reasoning

What should be the goals of such a theory? We list them here. The theory should cope with everyday determiners, such as *all*, *some*, *none*, *most*, *at least three*, *no more than four*, *less than a third*, which commonly occur in Indo-European languages, and with generic assertions, such as "ducks lay eggs," that is, assertions in English with plural subjects but no determiner (Khemlani, Leslie, & Glucksberg, 2008). It should cope with all common inferential tasks, including the spontaneous formulation of conclusions, the evaluation of given conclusions, and the evaluation of whether a set of assertions is consistent. The latter task is the other side of the deductive coin: Many systems of modern logic rely on a test of consistency to determine whether or not a conclusion follows from premises. It does if the conjunction of its negation with the premises yields an inconsistent set of assertions (Jeffrey, 1981). The theory should explain what is computed in reasoning (a theory of competence at the "computational level") and how it is computed (a theory of performance at the "algorithmic level"). An algorithmic theory must accordingly account for the interpretation of monadic assertions, for their mental representation, for reasoning, and for cognate tasks such as the verification of monadic assertions in relation to the world. The algorithmic account should explain both the differences in difficulty from one monadic inference to another and the source of common errors in performance, including the effects of beliefs and prejudices. The theory should be "explanatorily adequate," to use Chomsky's (1965) term; that is, it should explain how human beings develop the ability to reason from monadic assertions from childhood and how they develop strategies to cope with the exigencies of particular inferential tasks, both in daily life and in the laboratory. This aspect of the theory should explain the origins of the large differences in accuracy in monadic reasoning from one individual to another. Ideally, the theory should also give an account at the "implementational level," that is, how the brain implements the algorithms in neuronal systems. At present, we concede that this latter goal is utopian and that its achievement probably awaits further developments in cognitive neuroscience. Indeed, for some readers, the current prospectus for the theory might also seem utopian. It is beyond cognitive science at present, but it is just as well to try to describe these goals, because they will help psychologists to decide which theoretical paths lead in the right direction and which lead to dead ends.

In what follows, the article gives a brief account of the logic of monadic inferences. It then outlines the main empirical phenomena of syllogistic reasoning, and it describes phenomena that occur in other sorts of monadic reasoning. Next, it outlines the 12 extant theories of the domain, dividing them into three broad categories: heuristic theories, such as the theory based on the "atmosphere" of premises; theories based on formal rules of inference akin to those of logic; and theories based on set-theoretic accounts (i.e., on mental models or diagrams such as Euler circles). It presents a meta-analysis that evaluates the predictions of the theories against the results of six experiments examining all 64 sorts of syllogistic premises. None of the theories gives a satisfactory account of the data, and so, finally, the article uses its analyses of them to provide guidelines for the development of a comprehensive theory of monadic reasoning.

The Logic of Monadic Assertions

Aristotle was by his own account the first logician in the Western world, and he formulated an analysis of syllogisms. They were the heart of the logic of quantifiers until the invention of first-order logic in the late 19th century. Not until the 20th century did logicians prove that some monadic assertions cannot be represented in this calculus. Most psychological studies of monadic assertions have concerned Aristotelian syllogisms, and so we will outline their logic before considering other sorts of monadic inferences.

In the traditional Scholastic account of syllogisms in the Middle Ages (see, e.g., Cohen & Nagel, 1934), a syllogism consists of two premises and a conclusion, and each assertion is in one of four *moods*, two of which are affirmative and two of which are negative. We state the moods with their traditional abbreviations that derive from the Latin words for "I affirm" and "I deny," *affirmo* and *nego*, where the first two vowels in each word respectively denote the two relevant moods:

All A are B.	Affirmative universal (abbreviated as "A")
Some A are B.	Affirmative existential (abbreviated as "I")
No A are B.	Negative universal (abbreviated as "E")
Some A are not B.	Negative existential (abbreviated as "O")

(5)

A typical textbook example of a syllogism is

Some artists are bakers.
All bakers are chemists.
∴ Some artists are chemists. (6)

In this example, the terms in the premises are in the following figure:

A–B (artists–bakers)
B–C (bakers–chemists) (7)

There are four possible figures for the premises, where *B* is the middle term, which occurs in both premises, and *A* and *C* are the end terms, which occur in conclusions:

Figure 1	Figure 2	Figure 3	Figure 4
A–B	B–A	A–B	B–A
B–C	C–B	C–B	B–C

(8)

Aristotle recognized only Figures 1, 3, and 4, because any syllogism in Figure 1 becomes a syllogism in Figure 2 by switching the order of the two premises. For example, Example 6 above becomes *All bakers are chemists*, *Some artists are bakers* (see Aristotle's *Prior Analytics*, Book 1, in Barnes, 1984; see also Kneale & Kneale, 1962, p. 64). Scholastic logicians organized the figures differently because they also included the terms in the conclusion, which they constrained to one particular order, *C–A*. Given conclusions of this sort, the figures above correspond to the Scholastic numbering of the figures as 4, 1, 2, and 3, respectively.

Because each premise can be in one of four moods, there are 64 distinct pairs of premises (16 moods $\times$ 4 figures). Granted that there are four moods for conclusions, Scholastic logicians argued that there are 256 possible sorts of syllogism, and many psychologists reiterate this claim. The drawback of this account is illustrated in this example:

All A are B.
All B are C.
∴ All A are C. (9)

Individuals spontaneously draw this conclusion, which corresponds to Aristotle's perfect syllogism, but the only correct response in Scholastic logic is *Some C are A*, because the two end terms must be in this order in a conclusion. The moral for psychologists is that they should take into account eight possible conclusions to the 64 forms of premises: the four moods relating *A* to *C* and the four moods relating *C* to *A*. Hence, in principle, there are 512 syllogisms for psychologists to worry about. Given that *A*s, *B*s, and *C*s exist, 27 of the 64 pairs of premises yield at least one valid conclusion establishing a definite relation between the end terms. Of course, any set of premises yields infinitely many valid conclusions, and so when we refer to the other 37 syllogisms as "invalid," we mean that they do not yield a simple definite conclusion about the relation between the end terms. The 27 valid syllogisms fall into five categories in terms of the moods of the premises regardless of their order (Wason & Johnson-Laird, 1972, pp. 137–139): AA (three syllogisms), AI (four syllogisms), AE (eight syllogisms), AO (four syllogisms), and IE (eight syllogisms). Aristotle's procedure for evaluating the forms of syllogism yielding valid conclusions rested on his semantic intuitions (see, e.g., Kneale & Kneale, 1962, p. 67 et seq.). If he could find premises with a particular content in which the inference was invalid (i.e., it led from true premises to a false conclusion), then he eliminated it from his list. And his proof procedure, in essence, was to transform a given syllogism into one known to be valid.

Later, the Scholastic logicians developed a series of rules that act as a sieve retaining only valid conclusions (Cohen & Nagel, 1934, p. 79):

1. If both premises are affirmative, the conclusion must be affirmative.

2. If one premise is negative, the conclusion must be negative.

3. If both premises are negative, then no valid conclusion follows establishing a definite relation between the end terms.

4. The middle term must be *distributed* in at least one premise.

5. No term can be distributed in the conclusion if it is not distributed in the premise in which it occurs.

The notion of "distribution" in Rules 4 and 5 means that the assertion concerns the entire set of entities, and so the italicized terms are distributed in the following sorts of assertion: All *A* are B, Some A are B (neither term is distributed), No *A* are *B*, Some A are not *B*. The notion of distribution, as we show presently, also corresponds to the modern idea of a downward entailing term (Makinson, 1969).

First-order logic deals with quantification over individuals. It yields the following analysis of Example 6 above:

$$(\exists x)((\text{Artist } x) \,\&\, (\text{Baker } x))$$
$$(\forall y)((\text{Baker } y) \rightarrow (\text{Chemist } y))$$
$$\therefore (\exists x)((\text{Artist } x) \,\&\, (\text{Chemist } x)) \qquad (10)$$

"$\exists x$" denotes the existential quantifier ("there is at least some x"), "$\forall y$" denotes the universal quantifier ("for any y"), "&" denotes logical conjunction ("and"), and "$\rightarrow$" denotes material implication ("if __ then __"). As the analysis shows, each nonlogical term is, or can be treated as, a simple predicate that takes one argument, such as (Artist x), and so the premises are monadic. A common feature of accounts of the syllogism (and psychological theories) is to treat relations as though they were properties (or features). The following valid inference counts as a syllogism, because the relations that occur in it do not need to be analyzed in order to establish its validity:

Some composers hate Wagner.
All those who hate Wagner love Debussy.
∴ Some composers love Debussy. (11)

The logic of monadic sentences is decidable; that is, it can be formulated so that a finite number of steps suffice to establish whether or not an inference is valid. The proof of inference in Example 10 is straightforward. In formalizations of first-order logic, rules of inference are sensitive only to the logical forms of assertions. The rule of modus ponens, for example, has the following form, where *A* and *B* can be any sentences whatsoever:

If A then B.
A.
∴ B. (12)

This rule of inference can be used to prove that Baker e, and Baker e $\rightarrow$ Chemist e imply Chemist e, where "e" denotes an arbitrary individual, and conjunction and the restoration of quantifiers yields the conclusion to Example 10: $(\exists x)((\text{Artist } x) \,\&\, (\text{Chemist } x))$.

The analysis of syllogisms in modern logic differs from Aristotle's account (see, e.g., Strawson, 1952), because Aristotle took "all artists" to presuppose that artists exist, whereas the universal quantifier in predicate calculus has no such presupposition. Many psychological experiments, however, have eliminated the difference between language and logic in the experimental instructions, which tell the participants that individuals of all relevant sorts exist, and in their use of the definite article, as in "All the artists are beekeepers," which implies that artists do exist in the universe of discourse (cf. Boolos, 1984; Johnson-Laird & Bara, 1984a, 1984b). In contrast, an assertion of the form *All A are B* is often taken to have no such existential implication even in daily life. For example, "All deserters will be shot" can be true even if there are no deserters.

A simple view about the meaning of a quantifier, such as "all women," is that it refers to the set of all women. The difficulty with this view is to determine what the quantifier "no women" refers to (see, e.g., Geach, 1962). An alternative is to propose analyses based on first-order logic (e.g., Rips, 1994), but this account is not powerful enough to represent many determiners in ordinary language, such as "more than half" and "most" (Barwise & Cooper, 1981). One way to represent these determiners is in the so-called second-order predicate calculus, which allows quantification over sets (equivalently properties) as well as over individuals. This calculus copes with "generalized quantifiers," which Montague (1974) argued are the proper way to treat quantifiers in English (for introductions to his account, see Johnson-Laird, 1983, Chapter 8; Partee, 1975; Peters & Westerståhl, 2006). The fundamental idea is that "all women" is a generalized quantifier, that is, one referring to the set of all sets containing all women, and "no women" refers to the set of all sets containing no woman. Likewise, "most artists" refers to the set of all sets containing most artists, and so on. Similarly, the noun phrase consisting of a proper noun, such as "Cezanne," refers to the set of all sets of which Cezanne is a member. In this way, a perfectly uniform semantic treatment of clauses consisting of a noun phrase and a verb phrase becomes possible: An assertion is true if and only if the predicate of the sentence refers to a set that is a member of the set of sets to which the noun phrase refers. This treatment is advocated by many linguists. Unfortunately, from a psychological standpoint, generalized quantifiers are infeasible: The computation of, say, the set of all sets containing all women is intractable and is likely to be too large to fit inside anyone's brain (Partee, 1979).

A long-standing and feasible alternative is to treat monadic assertions as stating relations between sets (see, e.g., Cohen & Nagel, 1934, pp. 124–125). This approach accommodates the quantifiers that cannot be expressed in first-order logic, and Table 1 presents examples of set-theoretic analyses of the four syllogistic moods and some of these other quantifiers. Cognitive scientists have argued that mental representations of quantified assertions are set-theoretic (e.g., Johnson-Laird, 1975, p. 49; 1983, p. 140; Johnson-Laird & Steedman, 1978), and this view has recent proponents, such as Geurts (2003) and Politzer, Van der Henst, Luche, and Noveck (2006). It is also consistent with various diagrammatic systems of syllogistic reasoning, which we consider presently.

The Phenomena of Monadic Reasoning

Syllogistic Reasoning

Many early studies of syllogistic reasoning called for participants either to evaluate a given conclusion or to choose an option from a list of choices, and as a consequence the studies could not examine all 512 possible sorts of syllogism. Since then, however, studies have established several robust effects, which any viable theory needs to explain. For purposes of comparison, we have carried out a meta-analysis of experiments, which we present below after our review of the theories. One prior result was that some syllogisms are so easy that children spontaneously draw valid conclusions to them (Johnson-Laird et al., 1986), whereas others are so difficult that hardly any adults can cope with them (Johnson-Laird & Steedman, 1978).

Studies of syllogisms have also shown a robust effect of individual differences. Johnson-Laird and Steedman (1978) tested students at a highly selective American university, who drew valid conclusions on 55% of inferences, whereas Johnson-Laird and Bara (1984b) tested students at a nonselective Italian state university, who drew valid conclusions on only 37% of problems, which is above chance. Performance in these two samples (of 20 participants each) ranged fairly continuously from the most accurate participant, who drew 85% correct valid conclusions, down to the least accurate participant, who drew only 15% correct valid conclusions (Johnson-Laird, 1983, pp. 118–119). In general, reasoning ability correlates with measured intelligence, or a proxy for it (Stanovich, 1999), but this correlation is not too revealing because many tests of intelligence include items that depend on reasoning. Nevertheless, as we argue, the predictions of a theory should reflect the variety of conclusions that different individuals tend to draw.

The figure of syllogisms has a robust effect on reasoning. This *figural effect* was first reported in Johnson-Laird (1975; see also Dickstein, 1978). It is illustrated in the following typical pattern of results for two syllogisms that are logically equivalent because their premises are merely stated in different orders, giving rise to different figures:

Some A are B.	All B are A.
All B are C.	Some C are B.
∴ Some A are C. (15 subjects)	∴ Some C are A. (16 subjects)
∴ Some C are A. (2 subjects)	∴ Some A are C. (1 subject)

(13)

As these examples illustrate, the participants in Johnson-Laird and Steedman's study (1978) tended to formulate *A–C* conclusions for Figure 1 and *C–A* conclusions for Figure 2. The same effect occurred whether the conclusions were valid, as in the examples above, or invalid. The result has been replicated many times, and there is usually a small bias toward conclusions of the form *A–C* in the other two figures. One corollary is that when the only valid conclusion is contrary to the figural bias, a valid inference is difficult. For example,

No A are B.
Some B are C.
∴ Some C are not A. (14)

When valid conclusions accord with the bias, they are frequent (e.g., around 80% in Johnson-Laird & Steedman, 1978). When they violate the bias, they are infrequent (around 20%). Ever since the discovery of the figural effect, theorists have considered alternative accounts, based either on an intrinsic ordering of terms in representations that reflect underlying semantic processes (e.g., Johnson-Laird & Steedman, 1978) or on the "first in, first out" properties of working memory (e.g., Johnson-Laird & Bara, 1984b). In our view, the definitive account of the phenomenon is a semantic one due to Oberauer and his colleagues (Oberauer, Hörnig, Weidenfeld, & Wilhelm, 2005; Oberauer & Wilhelm, 2000).

The contents of syllogisms affect reasoning. In particular, individuals are more likely to accept a believable conclusion than an unbelievable one. This effect is greater for invalid conclusions than

Table 1
Set-Theoretic Analyses of Quantified Monadic Assertions and Their Paraphrases in English

Quantified assertion	Set-theoretic relation	Its paraphrase
All A are B.	$A \subseteq B$	A is included in B.
Some A are B.	$A \cap B \neq \varnothing$	Intersection of A and B is not empty.
No A is a B.	$A \cap B = \varnothing$	Intersection of A and B is empty.
Some A are not B.	$A - B \neq \varnothing$	Set of A that are not B is not empty.
At least three A are B.	$\mid A \cap B \mid \geq 3$	Cardinality of the intersection ≥ 3.
Three A are B.	$\mid A \cap B \mid = 3$	Cardinality of the intersection = 3.
Neither A is B.	$\mid A \mid = 2 \;\&\; A \cap B = \varnothing$	Cardinality of A = 2, and intersection of A and B is empty.
Most A are B.	$\mid A \cap B \mid > \mid A - B \mid$	Cardinality of intersection > cardinality of the A that are not B.
More than half the A are B.	$\mid A \cap B \mid > \mid A \mid / 2$	Cardinality of intersection > ½ of cardinality of A.
The A is a B.	$A \subseteq B \;\&\; \mid A \mid = 1$	There is one A, which is a B.

for valid conclusions, whether individuals are evaluating a given conclusion (e.g., Evans, Barston, & Pollard, 1983) or drawing their own conclusions (Oakhill & Johnson-Laird, 1985).

The development of syllogistic reasoning begins when children learn to understand and to utter quantified assertions. Very young children have a fragile grasp of their meaning: Children who are presented with a collection of blue circles mixed with red and blue squares, and who are then asked, "Are all the circles blue?" often erroneously respond, "No" (Inhelder & Piaget, 1964). Recent evidence suggests that 3- and 4-year-olds first treat quantified assertions with "all" as generics that admit exceptions (Hollander, Gelman, & Star, 2002), but by the age of 5 they do differentiate the two (Tardif, Gelman, Fu, & Zhu, 2011). Once children can understand determiners, their syllogistic reasoning has a similar pattern to adult reasoning: Their conclusions show the figural effect, and inferences that call for a consideration of multiple possibilities usually defeat them, just as they cause difficulty for adults (Johnson-Laird et al., 1986). Accuracy continues to increase through adolescence, but certain inferences are rare even in adulthood (Bara, Bucciarelli, & Johnson-Laird, 1995).

One final phenomenon is noteworthy. Different individuals use different strategies in syllogistic reasoning—a point made first by Newstead (1989). Several studies have reported the use of both diagrammatic and verbal strategies (e.g., Ford, 1995; Störring, 1908). Many individuals can say little or nothing when they are asked to think aloud as they draw syllogistic conclusions, and they make no use of paper and pencil to make diagrams (Bucciarelli & Johnson-Laird, 1999). However, when participants are asked to construct counterexamples to a conclusion using external cut-out shapes to represent individuals, they are able to do so; that is, they can depict cases in which the premises hold, but the conclusion does not. This study also showed that the participants had a preference for certain sorts of external model. For example, *Some of the X are Y* elicits a variety of models, but the most frequent one has *Y* as a proper subset of *X*. Likewise, the preferred interpretation for *All X are Y* has the two sets as coextensive. These preferences were for the first premise, but they were affected by the second premise. Hence, the coextensive interpretation disappeared when the second premise was in the O mood, in which case the participants were more inclined to build models of *All X are Y* in which *X* was properly included within *Y*.

A further study compared the participants' use of external models with their normal reasoning, a week apart in a counterbalanced order (Bucciarelli & Johnson-Laird, 1999, Experiment 4). They drew their own conclusions for all 48 syllogisms in Figures 1, 3, and 4. There was no difference in accuracy between the two sessions, but the participants tended to make more diverse responses when they built external models. The two most striking phenomena were, first, the participants' use of alternative models and, second, the variety of their strategies. All of them constructed more than one model on at least one occasion, ranging from two participants who built such sequences on 75% of problems, down to one participant who built them on only 8% of the problems.

One factor that may guide the development of strategies is the number of distinct sorts of individual that occur in mental representations of premises—a variable that Zielinski, Goodwin, and Halford (2010) refer to as "relational complexity." For instance, in a representation of the syllogism Aab Abc, where we use the abbreviation "A" for the mood of a premise and the lowercase letters to represent the order of the terms—for example, *All the artists are beekeepers*, *All the beekeepers are chemists*—the premises call for the representation of three sorts of individual: an artist who is also a beekeeper and a chemist, a beekeeper who is a chemist but not an artist, and a chemist who is neither an artist nor a beekeeper. If participants represent all three sorts, then Zielinski et al. assigned a complexity value of 3. They identified some strategies in which reasoners could reduce the complexity of representations. In the example above, reasoners could "chunk" the latter two sorts of individual and represent a chemist who is not an artist and who may or may not be a beekeeper. This operation yields a complexity value of 2 for the full representation. In this way, as these authors argued, complexity could provide a metric for assessing the difficulty of each syllogism.

The principal moral of these results is that individuals use a variety of strategies in reasoning—a phenomenon that also occurs in reasoning based on sentential connectives, such as *if* and *or* (Van der Henst, Yang, & Johnson-Laird, 2002). Reasoners differ in which premise they interpret first, in how they interpret the premises, and in how they go about searching for alternative models. In many cases, their strategies were more variable than the experimenters had envisaged: The participants used different sequences of operations to reach the same result (or different results). They differed one from another and from one problem to another (see also Galotti, Baron, & Sabini, 1986, for similar phenomena).

A task that appears to be closely related to syllogistic reasoning is an immediate inference from one sort of syllogistic premise to another. Individuals commonly err in this task (Wilkins, 1928). In a more recent study, Newstead and Griggs (1983) showed that about a third of their participants wrongly evaluated this sort of inference as valid:

All A are B.
∴ All B are A. (15)

An even greater proportion wrongly evaluated this sort of inference as valid:

Some A are not B.
∴ Some B are not A. (16)

Another recent study has reported still higher rates of error (Evans, Handley, Harper, & Johnson-Laird, 1999), but errors were reduced when individuals identified which Euler circle diagrams matched assertions (Newstead, 1989). Evidence exists that the propensity to make these inferences or to hesitate to make valid conversions, such as *Some A are B*, and so *Some B are A*, predicts the accuracy of performance in syllogisms (see the account below of the source-founding theory).

One common view is that individuals are likely to infer, for example, *Some A are B* from the assertion *Some A are not B*. Such an inference rests on the following idea (due to Grice, 1975): A speaker would not assert *Some A are B* if the speaker knew for a fact that *All A are B*. Hence, according to this convention of conversation, the speaker's assertion suggests, or "implicates," that *Some A are not B*. If individuals make these Gricean implicatures, their conclusions are likely to diverge from conclusions that are valid in logic. The evidence, however, is that Gricean inferences alone are poor predictors of syllogistic reasoning (Newstead, 1995). What does occur, in contrast, is symmetric (or "reversible") interpretations; that is, *All A are B* is interpreted as meaning that the two sets are coextensive, and *Some A are not B* is interpreted as meaning that the two sets overlap each other or are disjoint (see Wason & Johnson-Laird, 1972, p. 149). The errors that are predicted from such interpretations, or from their conjunction with Gricean implicatures, tend to occur more often than purely Gricean errors (Roberts, Newstead, & Griggs, 2001).

Other Sorts of Monadic Reasoning

Beyond syllogisms, there have been several investigations of monadic reasoning. Studies have been made of the cardinalities ascribed to quantified expressions, their usage, and effects of focus (see especially the work of Moxey, Sanford, and their colleagues, e.g., Moxey & Sanford, 2000; Moxey, Sanford, & Dawydiak, 2001). These studies have consequences for the theory of the meanings of quantified expressions and their mental representation.

Studies have provided evidence for the spontaneous use of counterexamples in monadic reasoning. Consider this problem based on proportional quantifiers, which cannot be represented in first-order logic:

> More than half the people in the room speak English.
> More than half the people in the room speak Spanish.
> Does it follow that more than half the people in the room speak English and Spanish? (17)

When people are given this problem (and pen and paper), they tend to respond correctly, "No," and they base their response on the construction of a counterexample (Neth & Johnson-Laird, 1999). For example, they draw a diagram of 10 people in the room and represent six as speaking English and six as speaking Spanish but with a minimal overlap (of two) between them. In other words, they construct an external model of the sets that satisfies the premises but that refutes the putative conclusion. Again, this use of counterexamples is not unique to reasoning based on quantified assertions, and it also occurs in sentential reasoning (Johnson-Laird & Hasson, 2003). Another study of counterexamples used problems of the following sort, which it contrasted with a question calling for mental arithmetic based on the same premises:

> There are five students in a room.
> Three or more of these students are joggers.
> Three or more of these students are writers.
> Three or more of these students are dancers.
> Does it follow that at least one of the students in the room is all three: a jogger, a writer, and a dancer? (18)

If you think about this problem, you are likely to envisage a possibility in which the conclusion holds, but if you persevere, you should think of a counterexample: Two students are joggers and writers, another two students are writers and dancers, and the fifth student is a jogger and dancer. Hence, the conclusion does not follow from the premises. The study used functional MRI to contrast reasoning and mental arithmetic, and it also compared easy inferences, which follow immediately from just one of the premises, with more difficult inferences such as the preceding one, which call for a search for counterexamples (Kroger, Nystrom, Cohen, & Johnson-Laird, 2008). The results corroborated such a search. While the participants were reading the premises, Broca's and Wernicke's areas of their brains were active, but these areas ceased to be active as the participants started to reason (cf. Goel, Buchel, Frith, & Dolan, 2000). Regions in right prefrontal cortex and inferior parietal lobe were more active for reasoning than for mental arithmetic, whereas regions in left prefrontal cortex and superior parietal lobe were more active for mental arithmetic than for reasoning. Only the difficult inferences calling for counterexamples elicited activation in the right prefrontal cortex (i.e., in the right frontal pole; see also Tsujimoto, Genovesio, & Wise, 2011).

We have completed our survey of the main empirical results of studies of monadic reasoning, which we summarize in Table 2, though we do describe other studies later in the article. In our view, we can safely draw three main conclusions about such reasoning: (a) monadic inferences differ in their difficulty in systematic ways, (b) individuals differ in their accuracy, and (c) individuals differ in their strategies.

Twelve Theories of Syllogistic Reasoning

We now outline the 12 main sorts of theory of monadic reasoning, almost all of which focus on syllogisms, and then we report the results of the meta-analysis of syllogistic reasoning. The diverse theories can be loosely grouped together into three principal varieties. First, there are heuristic theories, which allow that some individuals may reason deductively, but which seek either to characterize the causes of error or to postulate heuristics in place of deductive processes. The heuristics are based on atmosphere, illicit conversion, matching, or probabilities. Second, there are theories based on formal rules of inference. They include theories akin to simple verbal substitutions, akin to the predicate calculus, or akin to more powerful logics. Third, there are theories based on diagrams or sets. They include theories based on Euler circles, Venn diagrams, and mental models. We have consulted with all the recent proponents of the theories to try to ensure the accuracy of our descriptions and predictions for them. Table 3 summarizes each theory.

Heuristic Theories of Syllogistic Reasoning

Atmosphere. An early and influential hypothesis about syllogistic reasoning is that reasoners are predisposed to accept a conclusion that fits the mood of the premises (Sells, 1936; Woodworth & Sells, 1935). A succinct reformulation of the theory is due to Begg and Denny (1969): Whenever at least one premise is negative, the most frequently accepted conclusion will be negative; whenever at least one premise contains *some*, the most frequently accepted conclusion will likewise contain *some*; otherwise the bias is toward affirmative and universal conclusions. Because the effect is stronger for valid conclusions, Woodworth and Sells (1935) argued for an independent inferential mechanism. Revlis (1975) developed the atmosphere hypothesis into a model in which errors can occur in working out the joint atmosphere of the two premises. The model postulates that participants will reject a putative conclusion if it does not fit the atmosphere of the premises.

The most plausible aspect of the atmosphere hypothesis is that the quantifiers in the premises may well cue quantifiers for the conclusion. The original version of the hypothesis, however, accounts only for such a bias, and—as its proponents realized—it leaves much to be explained about syllogistic reasoning. Most conclusions that are deductively valid happen to fit

Table 2
Summary of the Empirical Phenomena of Monadic and Syllogistic Reasoning

Phenomenon	Description
Differences in difficulty among problems	Some syllogisms are extremely difficult (1% correct solutions), and others are easy (90% correct conclusions). Differences in difficulty are reliable across studies.
Figural effect	The figure of a syllogism (i.e., the order in which terms are organized) affects the frequency and type of conclusions reasoners tend to draw.
Content of terms	Individuals are more like to accept believable conclusions and less likely to accept unbelievable ones, especially for invalid syllogisms.
Individual differences	Accuracy varies greatly from one reasoner to another; likewise, some reasoners tend to produce invalid immediate inferences, and others hesitate to produce valid immediate inferences.
Development	Children are unable to draw sensible inferences from syllogisms until they can understand and produce quantifiers. The ability develops through adolescence up to adulthood.
The acquisition of strategies	Reasoners acquire various strategies to solve syllogisms (e.g., which premise is interpreted first, how premises are interpreted, and how counterexamples are constructed).

the atmosphere hypothesis (see the Scholastic rules above, which contain the atmosphere principles), and so any theory that explains valid inferences is bound to overlap with the atmosphere effect. The crucial datum is accordingly what happens in the case of invalid syllogisms. In some cases, individuals respond with a conclusion that fits the atmosphere, but in many cases, they correctly respond that "nothing follows"—a phenomenon that the atmosphere hypothesis cannot explain, because there is always a conclusion that fits the atmosphere of the premises. Conversely, there are cases in which individuals fail to draw a valid conclusion and respond wrongly that nothing follows even though the valid conclusion fits the atmosphere (Johnson-Laird & Bara, 1984b, p. 7; see also Dickstein, 1978; Mazzocco, Legrenzi, & Roncato, 1974).

One datum that appears to be contrary to atmosphere is from a study of syllogisms (Johnson-Laird & Byrne, 1989) that included inferences such as

Only the authors are bookkeepers.
Only the bookkeepers are cyclists.
What follows? (19)

An assertion of the sort *Only the A are B*, is logically equivalent to *All the B are A*, but *only* is intrinsically negative—it suggests that entities that are *not A* are *not B*. The results showed that premises based on *only* were more difficult to reason from than the equivalent premises based on *all*. Contrary to atmosphere, when both premises contained *only*, the participants drew a mere 16% of conclusions containing it, and when one of the premises contained *only*, just 2% of the conclusions contained it. A related finding is that it is much easier to make the inference from *All A are B* to *Only B are A* than vice versa.

Matching. A hypothesis that is analogous to the atmosphere effect is due to Wetherick and Gilhooly (1995). They proposed that some individuals try to reason logically, whereas others rely on a *matching* strategy in which they choose a conclusion that interrelates the end terms but matches the mood of the more *conservative* premise, that is, the one that presupposes the existence of fewer entities. An assertion of the form *No A are B* is the most conservative because it does not presuppose any entities, whereas an assertion of the form *All A are B* is the least conservative, and *Some A are B* and *Some A are not B* are equally conservative and lie between the two extremes. Therefore, these authors claimed, the partial rank order of preferred moods of conclusions is E $>$ O $=$ I $\gg$ A. It follows from the matching hypothesis that an *A* conclusion can be drawn only if both premises are *A*; that is, in this case, but not in others, matching makes the same predictions as atmosphere. Like atmosphere, matching cannot explain "no valid conclusion" responses.

Wetherick and Gilhooly (1995) reported a study in which the participants drew conclusions from syllogistic premises and constructed premises that implied given conclusions. The results enabled them to divide their participants into three main groups: those who reasoned well, those who did not reason well and tended to make matching responses, and those who did not reason well and tended not to make matching responses. One puzzle is the nature of the underlying processes enabling individuals in the first group to reason well. A secondary issue is whether those who make many errors are trying to reason or relying on a heuristic process. In sum, matching is a hypothesis about just one sort of strategy that individuals can adopt (K. J. Gilhooly, personal communication, December 1, 2010).

Illicit conversion. Chapman and Chapman (1959) proposed an account of certain errors in syllogistic reasoning, and Revlis (1975) formulated a more explicit version of the hypothesis (see also Revlin, Leirer, Yopp, & Yopp, 1980). The idea is that individuals often make invalid conversions from *All A are B* to *All B are A* and from *Some A are not B* to *Some B are not A*. Such conversions frequently yield true conclusions in daily life, and so they have a basis in probabilities. They can also underlie the acceptance of the invalid syllogisms such as

All A are B.
All C are B.
∴ All C are A. (20)

Likewise, according to the Chapmans, everyday probabilities underlie the inference that entities with a predicate in common are the same sort of thing. For example,

Some A are B.
Some C are B.
∴ Some C are A. (21)

In Revlis's (1975) formulation, individuals always represent assertions and their illicit conversions.

Table 3
Synopsis of Each of the 12 Main Theories of Syllogistic and Monadic Reasoning

Theory	Description
	Heuristic theories
Atmosphere Woodworth & Sells (1936) Begg & Denny (1969) Revlis (1975) Revlin et al. (1980)	If a premise contains *some*, use it in the conclusion; if a premise is negative, use a negative conclusion. Otherwise, draw a universal affirmative conclusion.
Matching Wetherick & Gilhooly (1990)	Draw a conclusion in the same mood as the most "conservative" premise, which commits one to the fewest sorts of individual, that is, according to this rank order of moods: E > O = I >> A.
Illicit conversion Chapman & Chapman (1959) Revlis (1975)	Convert a premise and also assign identity to individuals asserted to have the same properties.
Probability heuristics Chater & Oaksford (1999)	Draw a conclusion in the same mood as the least informative premise, and if its subject is the end term, use the same end term as the subject of the conclusion. The order of informativeness, starting with the least informative mood, is O < E < I < A. The theory extends to *most* and *few*.
	Formal rule theories
The PSYCOP model Rips (1994)	Start with the logical form of assertions and use rules of inference to prove given conclusions, or guess a tentative conclusion and try to prove it. Invokes Gricean implicatures. For a similar approach, see Braine (1998) and Braine and Rumain (1983).
Verbal substitutions Störring (1908) Ford (1995)	Substitute one term in a premise for another term in another premise; for example, given *Some A are B* and *All B are C*, infer *Some A are C*.
Source-founding theory See below.	See below.
Monotonicity theory Geurts (2003) Politzer (2007)	Any upward entailing term, *A*, is replaceable by another, *B*, if *A* implies *B*; and any downward entailing constituent, *A*, is replaceable by another, *B*, if *B* implies *A*.
	Theories based on diagrams, sets, or models
Euler circles Erickson (1974) Guyote & Sternberg (1981) Ford (1995)	Each circle represents a set, and relations among sets are represented by relations among circles.
Venn diagrams Newell (1981)	Three overlapping circles represent possible relations among sets, and annotations show which intersections of sets are empty and which are not. Newell used strings of symbols to capture these diagrams.
Source-founding theory Stenning & Yule (1997) Stenning & Cox (2006)	Existential premises establish the necessary existence of individuals, and other premises can add properties. The system for valid syllogisms is implementable in Euler circles or sentential rules.
Verbal models Polk & Newell (1995)	Uses premises to construct models of individuals and formulates a conclusion, rejecting those that fail to relate end terms. It reencodes premises until either they yield a legal conclusion or it declares that nothing follows (making no use of counterexamples).
Mental models Johnson-Laird & Steedman (1978) Bucciarelli & Johnson-Laird (1999)	Sets are represented iconically as models of their members. Members are combined in a parsimonious way. Conclusions from such models can be refuted by counterexamples.

Note. A = *All* __ *are* __; I = *Some* __ *are* __; E = *No* __ *are* __; O = *Some* __ *are not* __.

The strong point of the theory is that individuals do make illicit conversions when they evaluate immediate inferences from a single premise, particularly with assertions using abstract predicates, such as *A* and *B* (Sells, 1936; Wilkins, 1928): for example,

Some A are not B.
∴ Some B are not A. (22)

They may do so, as the Chapmans argued, because such conclusions probably hold in daily life, though the acceptance of inferences from one assertion to another may mislead us about what happens in syllogistic reasoning. For instance, the widespread occurrence of conversions would eliminate the figural effect: A figure of the form *A–B*, *B–C* should be just as likely to elicit invalid conclusions of the form *C–A* as *A–C* (but cf. the source-founding theory below on differences between individuals in their propensity to accept the converse of various moods of premise).

Probability heuristics. As part of their rational analysis of human reasoning—a view owing much to Anderson (1990)—Chater and Oaksford (1999) described a probability heuristics model (PHM) of syllogisms based on the assumptions that, first, theories should generalize to the defeasible reasoning of everyday life and, second, orthodox logic does not match performance on deductive tasks in the psychological laboratory. The solution according to Chater and Oaksford is that the appropriate theory at the computational level (i.e., the theory of what is computed) should be not logic but the probability calculus, and they proposed some heuristics at the algorithmic level to explain the mental processes underlying syllogistic reasoning. The essence of their account is that people are not failing to be logical but applying heuristics that often, though not always, converge on "probabilistically valid" conclusions. These authors used an analogous approach in their influential analysis of Wason's selection task (e.g., Oaksford & Chater, 1996).

The PHM postulates—as do the atmosphere and matching hypotheses—that individuals use heuristics to generate putative conclusions but that some individuals may have processes for testing deductive validity. However, the authors wrote: "We assume that, in most people, these [processes] are not well developed, which enables us to explain why many people frequently produce conclusions which are not logically valid" (Chater & Oaksford, 1999, p. 196). One of the major virtues of their account is that it generalizes to quantifiers that cannot be represented in first-order logic, and in particular to the determiners *most* and *few*.

Chater and Oaksford (1999) postulated that quantified assertions have probabilistic meanings, and so a conclusion is probabilistically valid ("p-valid") provided that the premises place sufficient constraints on the conditional probability of one end term given the other. If $p(C \mid A) = 1$, then *All A are C* is p-valid; if $p(A \mid C) = 1$, then the converse, *All C are A*, is p-valid; if $p(A \,\&\, C) > 0$, then *Some A are C* and its converse are p-valid; if $p(A \,\&\, C) = 0$, then *No A are C* and its converse are p-valid; if $p(C \mid A) < 1$, then *Some A are not C* is p-valid; and if $p(A \mid C) < 1$, then *Some C are not A* is p-valid. A transparent example is the inference

All A are B. $\quad p(B \mid A) = 1$
All B are C. $\quad p(C \mid B) = 1$ (23)

The value of $p(C \mid A)$ equals 1 too, and so the conclusion *All A are C* is p-valid. The computation of the probability of a conclusion depends on the probabilities of each premise, but it also depends on the figure of the syllogism. The other parameters affecting the probability of the conclusion are $p(A)$, $p(B \mid \textit{not-}A)$, and $p(C \mid \textit{not-}B)$. A divergence occurs between p-valid conclusions and logically valid conclusions: Thirty-one pairs of orthodox syllogisms yield p-valid conclusions, but only 27 pairs of orthodox syllogisms yield deductively valid conclusions.

Chater and Oaksford (1999) did not suppose that naive reasoners compute p-validity in order to evaluate syllogisms. Instead, they argued, individuals have "fast and frugal" heuristics (see Gigerenzer & Goldstein, 1996, for this notion) that usually converge on p-valid conclusions. These heuristics depend on probabilistic entailments ("p-entailments"):

All/most/few A are B p-entail *Some A are B*, where the inference from *all* holds provided *A*s exist;
Most/few/no A are B p-entail *Some A are not B*;
Some A are B and *Some A are not B* mutually p-entail each other. (24)

Some of these entailments follow from pragmatic considerations (see Grice, 1975), but one worrying consequence is that *All A are B* p-entails *Some A are B*, which in turn p-entails *Some A are not B*, which contradicts *All A are B*.

In Shannon's information theory, the informativeness of an assertion is the inverse of its probability: The smaller the probability of a communication, the more informative it is (Shannon & Weaver, 1949). Granted that an assertion *P* implies *Q*, then *Q* cannot be more informative than *P*, and so the p-entailments above yield a partial rank order in informativeness. Chater and Oaksford (1999) assumed, however, that the properties referred to in quantified assertions typically apply to only a small proportion of possible objects—that is, they are rare. Hence, almost all assertions of the form *No A are B* are true a priori and therefore uninformative, because of this rarity assumption. Together, probabilistic entailments and rarity yield a rank order of informativeness, which, starting from the most informative assertion, is *all*, *most*, *few*, *some*, *none*, *some_not*. The last of these assertions is much less informative than the penultimate assertion.

Armed with the concepts of informativeness and p-entailment, Chater and Oaksford (1999) stated the following three heuristics:

- the min-heuristic: the preferred conclusion has the same quantifier as the least informative premise;
- the p-entailment heuristic: the next most preferred conclusion is a p-entailment of the conclusion generated by the min-heuristic;
- the attachment heuristic: if the least informative premise has an end term as its subject, it is the subject of the conclusion; otherwise, the end term in the other premise is the subject of the conclusion.

They also proposed two further heuristics governing conclusions:

- the max-heuristic: individuals' confidence in a conclusion generated by the three preceding heuristics is proportional to the informativeness of the most informative premise (the max-premise); with low confidence, individuals should tend to respond that nothing follows from the premises;
- the O-heuristic: individuals avoid producing or accepting O-conclusions (e.g., *Some A are not C*), because they are so uninformative in comparison with other conclusions.

The heuristics predict that individuals will draw conclusions to syllogisms that do not have a p-valid conclusion. In their meta-analysis of results on studies of orthodox syllogisms, Chater and Oaksford (1999) ignored the order of the two terms in conclusions of a given mood and lumped together *All A are C* with *All C are A* and *Some A are not C* and *Some C are not A*, where at most one of each pair is valid. Similarly, their meta-analysis includes two studies in which the participants were constrained to choose from only the four Scholastic conclusions (of the form *C–A*) and "nothing follows." The min-heuristic successfully picked out the two most frequent responses in the resulting meta-analysis. They also used four empirically estimated parameters to fit the model to the data, accounting for over 80% of the variance. They fit the model only to the mood of the conclusion. But they showed that the model gives a good account of the logically invalid syllogisms. They corroborated the max-heuristic by showing that the max-premise predicted the proportion of "nothing follows" responses in the appropriate way. They also corroborated p-entailments; the attachment heuristic, which predicts the order of the two terms in a conclusion; and various other aspects of the theory. They compared the PHM with other theories and argued that it does better than its rivals. They also argued that it gives a good account of two new experiments that they carried out to investigate syllogisms using premises based on *most* and *few*. They claimed that to cope with these determiners, other theories must either adopt their notion of p-validity or develop an alternative notion of validity (p. 233). In fact, the standard notion of validity, which we introduced earlier, suffices. As an example, consider the following inference:

Most of the artists are bakers.
All the bakers are chemists.
∴ Most of the artists are chemists. (25)

It is valid because the conclusion holds in every possibility in which the premise holds (see Jeffrey, 1981, p. 1). But as they rightly pointed out, a vital component of any theory of quantifiers must explain how individuals reach such conclusions. One important finding in their study is that participants are sensitive to logical validity (Chater & Oaksford, 1999, p. 214). Hence, as they said, at least some participants may sometimes infer logically valid conclusions.

The PHM has three main strengths. First, it offers the most comprehensive account of heuristics for syllogistic reasoning. Second, it accommodates syllogisms based on the determiners *most* and *few*, and two experiments corroborated this account. Third, and perhaps the most important aspect of the theory, is the role of informativeness—the max-heuristic—in determining individuals' confidence in the conclusions that they draw. This aspect of the theory appears to have wide application. However, even though the heuristics are designed to salvage rationality, individuals are irrational on the theory's own account of p-validity. The heuristics diverge in some cases from what is p-valid, and individuals overgeneralize the heuristics and draw conclusions where none is warranted, not even in terms of p-validity (Chater & Oaksford, 1999, p. 207). Hence, people are not wholly rational in their syllogistic reasoning. Conversely, the theory allows that logical validity may play a part in inference, but it offers no account of how it might do so. The PHM provides an explanation of the figural effect and even, indirectly, an explanation of the differences in difficulty from one syllogism to another. It is less successful in accounting for other results. When individuals reason syllogistically, their performance is logically more accurate the second time around—even though they received no feedback and had no knowledge that they were to be tested twice (Johnson-Laird & Steedman, 1978). An explanation in terms of heuristics that are not geared to delivering logically valid responses offers no ready account of this phenomenon. By far the most important individual difference in syllogistic reasoning is that some people are very accurate and others are not—at least as evaluated with respect to logical validity. The PHM seems unable to offer any explanation of this phenomenon, because logical validity plays no part in the theory.

The rarity assumption postulates that the properties referred to in quantified assertions usually apply to only a small proportion of possible objects. But this assumption is dubious for the contents of many studies. Consider, for example, these premises from Johnson-Laird and Steedman (1978):

All of the gourmets are storekeepers.
None of the storekeepers are bowlers. (26)

The use of the definite article establishes the existence of particular subsets of gourmets, storekeepers, and bowlers, and at the same time it insulates the assertions to some degree from the rarity assumption. The assertion that none of the storekeepers are bowlers is informative, and the probabilistic entailments do not establish the relative informativeness of the two premises. Likewise, the O-heuristic implies that reasoners should avoid conclusions in this mood, but the evidence suggests that they are not reluctant to draw such conclusions (e.g., Hardman & Payne, 1995; Roberts et al., 2001). Finally, one wonders about the origins of the five heuristics in intellectual development: Where do they come from, and how could they have developed?

Theories Based on Formal Rules

The PSYCOP model (based on logic). Braine and his colleagues (e.g., Braine, 1998; Braine & Rumain, 1983) and Rips (1994) have proposed general theories of quantified reasoning that are based on formal rules of inference akin to those in a so-called natural deduction formulation of first-order logic. The two theories are similar enough that a decisive empirical test between their accounts of syllogistic and monadic inferences is hard to envisage. Rips's PSYCOP theory, however, is more comprehensive and has been modeled computationally, and so we focus on its account.

Rips (1994) defends deduction as a central human ability; he defends formal rules as the basic symbol-manipulating operators of the mind; and he defends formal rules as the lower-level mechanism for deductive reasoning. His set of rules was the first to accommodate reasoning both with sentential connectives and with quantifiers in a single psychological theory. At the core of the formal conception of reasoning is the concept of a mental proof. As Rips (1994, p. 104) wrote:

> I assume that when people confront a problem that calls for deduction they attempt to solve it by generating in working memory a set of sentences linking the premises or givens of the problem to the conclusion or solution. Each link in this network embodies an inference rule . . . , which the individual recognizes as intuitively sound.

The theorist's task is therefore to formulate psychologically plausible rules of inference and a mechanism for using them to construct mental proofs. The inputs to the program—and in effect to Braine's account as well—are the logical forms of the premises. Unfortunately, no algorithm exists for extracting the logical form of assertions, which often depends on the context of the sentences used to make them.

One obvious problem with rules is that they can be applied recursively to their own consequences, leading the system to run amok by drawing longer and longer conclusions, as here: *A*, *B*, ∴ *A & B*, ∴ *A & B & A*, ∴ *A & B & A & B*, and so on, ad infinitum. However, a formal rule can be used in two ways: either to derive a step in a forward chain from premises to conclusion or to derive a step in a backward chain from a given conclusion to subgoals for what has to be proved in order to prove the conclusion. To prevent the system running amok, Rips (1994) constrained certain rules, including those that increase the length of a conclusion, so that they can be used only in backward chains. (Braine used a different method to the same end.) An important rule in Rips's system is one that makes suppositions, which can then be discharged later, as in this sort of inference: *If A then B*, *If B then C*; therefore, *If A then C*. The proof depends on a supposition of *A*, then the use of the rule of modus ponens to prove *B*, its further use to prove *C*, and finally the suppositional rule discharges the supposition by expressing it explicitly as the if-clause in the conditional conclusion. Rips postulated that this rule can be used only in backward chains of inference.

The orthodox treatment of syllogisms in predicate calculus has rules that eliminate quantifiers by instantiating their variables with the names of hypothetical individuals, rules that make inferences based on sentential connectives, and rules that reintroduce quantifiers after the completion of the sentential inferences (see, e.g., Jeffrey, 1981). PSYCOP forgoes the rules for instantiation, and instead its input is representations of the logical form of assertions. In these representations, quantifiers are replaced by names and variables. For example, the sentence *Every child has a mother* is represented in the following way:

IF Child(x) THEN Mother(a_x, x),

where x stands for a universally quantified variable and a_x stands for a temporary name with a value dependent on x, a so-called Skolem function, that is equivalent to an existential quantifier. PSYCOP contains rules for matching variables and names in these representations.

PSYCOP postulates forward rules for two easy sorts of syllogism:

All A are B.
All B are C.
∴ All A are C. (27)

All A are B.
No B are C.
∴ No A are C. (28)

The theory also introduces a rule of conversion:

No A are B.
∴ No B are A. (29)

The similar conversion of an existential premise—*Some A are B*, therefore, *Some B are A*—has no rule of its own, because it can be derived from the rules for conjunction.

Because Rips's theory is close to logic, he postulated two implicatures (Grice, 1975) in order to ensure that universal quantifiers imply the existence of the relevant entities:

All A are B implicates that *Some A are B*.
No A are B implicates that *Some A are not B*. (30)

With these implicatures, PSYCOP yields the 27 valid syllogisms in the standard account. Rips (1994, pp. 230–231) also endorsed two Gricean implicatures:

Some A are B implicates *Some A are not B*.
Some A are not B implicates *Some A are B*. (31)

He stipulated that implicatures are not transitive in order to prevent the following chain of inferences: *All A are B* has an implicature that *Some A are B*, which has an implicature that *Some A are not B*. The latter conclusion would be inconsistent with the initial premise.

PSYCOP makes no predictions about specific invalid conclusions that are likely to occur in deduction. It postulates that errors can result from a failure to recognize the applicability of a rule, to retrieve the rule, or to carry out the steps it requires (Rips, 1994, p. 153). Similarly, reasoners may be uncertain about the correctness of a rule or its appropriateness (p. 379). Hence, errors can arise from many sorts of failure with rules, especially structurally more complicated rules (p. 388). The theory can accordingly predict which rules are likely to yield errors, but it cannot predict what the resulting erroneous conclusions will be. Likewise, the theory has no guaranteed way to establish that nothing follows from premises. It allows that reasoners may search through all possible proofs and fail to find one, but owing to the particular formulation of the theory, even if such a search were exhaustive, it would not guarantee that nothing follows.

Rips (1994, p. 233 et seq.) reported an experiment in which subjects evaluated the validity of entire syllogisms presented in the Scholastic figures. He fit the theory to the data by using the data to estimate the probabilities that each rule is used appropriately. He also showed how the theory might account for the results of a study in which subjects drew their own conclusions (Johnson-Laird & Bara, 1984b). Because of the constraints on rules for suppositions, the only way for participants to draw their own conclusions for some syllogisms is to guess a tentative conclusion and then to prove it working backward from the conclusion to the premises.

Verbal substitutions. From Störring (1908) onward, psychologists have claimed that some reasoners use diagrams and others use verbal or rule-like procedures. We deal with theories based on Euler circles and other diagrammatic methods below, and here we consider verbal rules. Consider the following syllogistic premises:

Some A are B.
All B are C. (32)

One way to draw a conclusion from them is to realize that the second premise sanctions the substitution of *C* for *B* in the first premise to yield

Some A are C. (33)

Various psychologists have observed simple substitutions of this sort (e.g., Bacon, Handley, & Newstead, 2003; Störring, 1908). Likewise, Ford (1995) examined the reasoning of 20 members of the Stanford University community as they attempted to draw conclusions from the 27 pairs of syllogistic premises that yield valid conclusions. From their "thinking aloud" protocols and diagrams, she divided the participants into a group that used Euler circles and a group that used substitution rules. Ford claimed: "Neither group makes use of representations containing finite elements standing for members of sets" (p. 19). In fact, at least four of her participants made claims about individual members of sets according to her protocols. Here is one example: "Suppose there's two historians right that means there are two weavers who are also historians so we can say some of the weavers are historians . . . ." Moreover, plenty of monadic inferences depend on the representation of individuals (e.g., "Pat is a sculptor; all sculptors are artists; therefore, Pat is an artist").

Eight of the participants in Ford's (1995) study used verbal substitutions. That is, they spoke of replacing one term in a syllogism with another, crossed out one term and replaced it with another, rewrote a syllogism as an equation, or drew arrows between a syllogism's terms (see Ford, 1995, p. 18, Footnote 2). She wrote: "The subjects . . . take one premise as having a term that needs to be substituted with another term and the other premise as providing a value for that substitution" (p. 21). She proposed four principles for such substitutions. The first principle (p. 21) is

> If a rule [i.e., a premise] exists affirming of every member of the class C the property P, then whenever a specific object, O, that is a member of C is encountered it can be inferred that O has the property P.

The phrase "a specific object, O" refers to either "some of the O" or "all of the O." Hence, the principle translates into two rules of inference:

Some A are B.	All A are B.
All B are C.	All B are C.
∴ Some A are C.	∴ All A are C.

It also translates into the corresponding two rules for cases in which the order of the premises is swapped round.

We can translate Ford's other three principles into analogous rules in which the quantifier in the first premise is also the quantifier in the conclusion:

All/Some A are not B.
All C are B.
∴ All/Some A are not C.

All/Some A are B.
None of the B is C.
∴ All/Some A are not C.

All/Some A are B.
None of the C is B.
∴ All/Some A are not C.

Such rules have a long history in logic (see Politzer, 2004, who observed them in Aristotle's proofs), and they have a short history in psychology (e.g., Braine & Rumain, 1983, proposed similar rules, albeit using a different notation; see also Braine, 1998). Ford (1995) proposed some more sophisticated principles to capture other valid syllogisms, and her principles can be applied beyond syllogisms to other sorts of monadic reasoning. A strong point of her account is that these more sophisticated substitutions yield reliably poorer performance by those participants whom she classified as using them. Likewise, verbal substitutions predict which syllogisms are likely to yield errors, though they cannot predict what those resulting erroneous conclusions will be (M. Ford, personal communication, January 3, 2011). But is there something deeper going on in these substitutions? The next theory suggests a possible answer.

The monotonicity theory. Geurts (2003) proposed a general verbal substitution rule in his monotonicity theory of syllogistic reasoning, which is based on the monotonic properties of quantifiers and their so-called upward and downward entailments (see Barwise & Cooper, 1981). The contrast between the two sorts of entailment can be illustrated, first, with an upward entailment, such as

He was wearing a green tie.
∴ He was wearing a tie. (34)

The inference is upward entailing (or "monotone increasing") because green ties are included in the set of ties. Negation, however, switches the entailment downward (or "monotone decreasing") from a set to one of its proper subsets:

He was not wearing a tie.
∴ He was not wearing a green tie. (35)

The nature of the adjective in such inferences is important. The following inference, for example, is invalid, because imitation leather is not the real thing:

Her bag is made from imitation leather.
∴ Her bag is made from leather. (36)

The premises of a simple syllogism have the following entailments, where "↑" denotes upward entailing and "↓" denotes downward entailing:

Some A are B: Some(A↑, B↑).
All B are C: All(B↓, C↑). (37)

Because *B* is upward entailing in the first premise, it can be replaced by any term that it implies, and the second premise asserts that it implies *C*. Hence, the following inference is valid: therefore, *Some A are C*. The other moods of syllogistic premises have these assignments of entailments:

No A are B: No(A↓, B↓).
Some A are not B: Some(A↑, not (B↓)↑). (38)

Hence, terms that are downward entailing are distributed in the Scholastic sense (see the section on the logic of monadic assertions above). Geurts proposed a rule for monotonicity, which specifies substitutions for both upward and downward entailments:

> Any upward entailing constituent, α, is replaceable by another, β, if α implies β; and any downward entailing constituent, α, is replaceable by another, β, if β implies α.

Other rules allow for the conversion of *some* and *no* premises, and for the inference from *No A is a B* to *All A are not B*. And an axiom asserts that *all* implies *some*. The resulting theory yields all the valid syllogisms and subsumes the four verbal substitution rules in Ford (1995; see above). A strong point of the theory is that it generalizes to numerical determiners, such as *at least two*, and to proportional determiners that cannot be represented in first-order logic, such as *more than half*. But it gives no account of erroneous inferences, beyond allowing—for unspecified reasons—that individuals may make illicit conversions. It also gives no account of the erroneous conclusions that individuals frequently draw from two existential premises or from two negative premises. For example,

Some A are B.	No A are B.	
Some B are C.	No B are C.	
∴ Some A are C.	∴ No A are C.	(39)

Similarly, as Newstead (2003) pointed out, it cannot explain the differences in difficulty of invalid syllogisms. Not all quantifiers are monotone, and so the theory also fails to explain valid inferences that are drawn from such quantifiers as *exactly three artists*, which are not monotone. For example,

Exactly three artists are bakers.
∴ More than one artist is a baker. (40)

Just as the monotonicity theory yields an account of the verbal substitution rules above, so too, as we will see, the monotonicity theory is a special case of a still more general account. According to Geurts, "The monotonicity theory was never intended to give a full-blown account of syllogistic reasoning, but merely to show that certain aspects of the meanings of quantifiers play a role in reasoning" (personal communication, December 22, 2010).

Experiments have examined other effects of monotonicity on reasoning. Geurts and van der Slik (2005) argued that a system for monotonicity can be simple because it requires only a shallow understanding of assertions. Reasoners need not grasp the exact meaning of determiners, such as *more than half*, merely whether they are upward or downward entailing: "In so far as human reasoning is based on monotonicity inferences, it will be sensitive only to the logical bare bones of an argument" (Geurts & van der Slik, 2005, p. 106). These authors predicted that upward entailing inferences should be easier than downward entailing inferences on the grounds that the latter are "marked" terms in linguistics (see Clark, 1974) and that there seems to be a general biological preference for an upward direction. They examined upward entailing inferences, such as

Every nurse played against more than 2 foresters.
All foresters were socialists.
∴ Every nurse played against more than 2 socialists. (41)

They also examined the corresponding downward inferences based on *less than 2 foresters*. And the results corroborated the difference.

In contrast, Politzer (2007) used a task in which the participants had to fill in the missing quantifier and predicate in a conclusion from a premise containing a "blank" predicate symbolized as [ ]. A typical problem (translated from French) was, for example,

All animals are [ ].
∴ __ cats __ [ ], (42)

to which the correct completion of the conclusion was

∴ All cats are [ ]. (43)

The premises in the problems were of the form *All*, *Some*, *No*, *Some_not*; and on half the trials the premise had a subject and a blank predicate, as in the example above, and on half the trials the premise had a predicate but a blank subject, as in *All* [ ] *are cats*. The participants' completions showed that they were aware of both upward and downward entailments. But the results did not corroborate Geurts's prediction that upward entailing inferences should be easier than downward entailing ones (see also Newstead, 2003, for a similar failure to find this difference), but rather inferences from universal premises (*all* and *no*) were easier than those from existential premises (*some* and *some_not*).

Diagrams, Models, and Sets

In line with set-theoretic analyses, logicians have developed various diagrammatic systems for monadic assertions and for syllogistic inference. The diagrams most relevant to psychological theories are Euler circles and Venn diagrams. Euler circles represent a set as a circle, and the inclusion of one set within another by one circle spatially included within another. Each premise in a syllogism can accordingly be represented as one or more Euler diagrams (see, e.g., Politzer's, 2004, account of Gergonne's method of using Euler circles to represent the assertions in syllogisms). Originally, there was no systematic procedure for combining the diagrams, and informal methods can easily overlook a possibility; for example, the conjunction of *Some A are B* and *All B are C* calls for 16 different combinations. To see why, consider Table 4, which represents all the possible sorts of individual according to these premises. One sort of individual with the properties of *A*, *B*, and *C* must exist, and there are five other sorts of individual that may or may not exist. One of these sorts of individual has none of the properties *A*, *B*, and *C*, and Euler circles do not represent such cases. Hence, there are $2^4 = 16$ distinct Euler diagrams of the premises.

Venn diagrams are a more efficient method (Edwards, 2004). They represent three sets as overlapping circles within a rectangle representing the universe of discourse. Hence, there are eight distinct regions in the diagram corresponding to all possible sets based on instances and noninstances of the three sets. A premise such as *All B are C* establishes that the regions corresponding to *B*s that are not *C*s, which may or may not be *A*s, are empty. The premise *Some A are B* establishes that the regions corresponding to *A*s that are *B*s, whether or not they are *C*s, are not empty. But as the first premise establishes, only one of these two regions has members: *A*s that are *B*s must be *C*s. So, at least some *A*s are *C*s. This system of annotating which sets have members, and which do not, establishes whether or not any syllogism is valid.

Logicians have sometimes claimed that diagrammatic methods of reasoning are improper (cf. Tennant, 1986), but the late Jon Barwise and his colleagues have shown that these methods are feasible, and indeed yield complete systems that capture all valid syllogisms (Barwise & Etchemendy, 1994; Shin, 1992). Barwise and Etchemendy (1994) developed a computer program, Hyperproof, that helps users to learn logic. It uses diagrams to represent conjunctive information and sentences to represent disjunctive information. But the most powerful diagrammatic methods were developed by Peirce (1958). His existential diagrams transcend syllogisms because they accommodate the whole of the predicate calculus (for an account, see Johnson-Laird, 2002). Peirce wrote that his diagrams "put before us moving pictures of thought" (Vol. 4, para. 8), but they seem too sophisticated to be psychologically plausible.

Euler circles. Several early information-processing accounts of quantified assertions were set-theoretic in nature (e.g., Ceraso & Provitera, 1971; Johnson-Laird, 1970, 1975; Neimark & Chapman, 1975; Wason & Johnson-Laird, 1972, pp. 56–57). And the earliest explicit attempt to specify the mental processes underlying syllogistic inference is Erickson's (1974) account based on Euler cir-

Table 4
Set of Individuals to Which a Syllogism—Some A Are B, All B Are C—*Can or Must Refer*

Properties of an individual			Status
A	B	C	+
A	B	Not-C	−
A	Not-B	C	±
A	Not-B	Not-C	±
Not-A	B	C	±
Not-A	B	Not-C	−
Not-A	Not-B	C	±
Not-A	Not-B	Not-C	±

Note. Each row corresponds to an individual. + = the corresponding individual must exist; − = the corresponding individual cannot exist; ± = the corresponding individual may or may not exist.

cles. When individuals draw Euler circles, they do not depict all possibilities (see, e.g., Bucciarelli & Johnson-Laird, 1999), and Erickson anticipated this failure as a likely source of error. He postulated that *All A are B* is often interpreted in a coextensive way and that individuals construct only one combination of diagrams representing syllogistic premises, selected at random from the set of possibilities. This procedure always delivers a conclusion, and so the theory cannot predict the response that nothing follows. Erickson assumed that atmosphere determines the mood of the conclusion used to express the resulting diagram, because a diagram of an overlap between the sets corresponding to the end terms is appropriately interpreted sometimes as *Some A are C* and sometimes as *Some A are not C*.

Erickson's (1974) account is important as the first information-processing theory of syllogistic reasoning, but it suffers by comparison with later accounts in its inability to predict responses that nothing follows from the premises, which even occur for premises that yield a valid conclusion (see the meta-analysis below). A more tractable variant of Euler circles uses strings of symbols in place of diagrams. Guyote and Sternberg (1981) took this step, and their method is in essence set-theoretic. That is, they introduced novel symbols to represent the four moods of syllogistic assertion, but the symbols correspond to those in Table 1. The procedure for combining the representations depends on two rules of inference. First, if one set is included in a second set, and the second set is included in a third set, then the first set is included in the third set. Second, if one set is included in the complement of a second set, and the second set itself is included in a third set, then the first set is either included or not included in the third set. The first rule is a statement of upward entailment. The second rule is true of any pair of otherwise contingent sets, and so it serves only a weak general purpose.

Guyote and Sternberg's (1981) process of combining the representations of premises is complex and calls for four steps. The first step constructs transitive chains of symbols, and the next two steps eliminate those combinations that are inconsistent with one or other of the premises. The final step combines what survives into complete representations and selects a matching conclusion, if any, from the list of given conclusions. The theory postulates that these processes are carried out correctly, and it locates the cause of errors in the selection of putative conclusions that match final representations. Guyote and Sternberg were therefore obliged, like Erickson (1974), to invoke the atmosphere effect in order to ensure an appropriate mood for certain conclusions. The most efficient method of using Euler circles is due to Stenning and his associates, and we consider it below. But a problem with such diagrams, as Rips (2002, p. 387) observed, is that reasoners may not use them unless they have already been taught to do so at school.

Venn diagrams. No psychologist has proposed that monadic reasoning is based on visual images of the three overlapping circles used in Venn diagrams. Newell (1981), however, described a theory of syllogisms based on strings of symbols representing the different areas in a Venn diagram and formal rules to combine the strings corresponding to the two premises. The premise *All A are B* is represented by a string:

Nec A+B+, Pos A−B+, Pos A−B−.

The string is akin to rows in a table of possible individuals, such as Table 4 above, that represents the individuals that are necessary and those that are possible according to the premise. The string means that there are necessarily *A*s that are *B*s (Newell followed Aristotle in taking universals to establish existence), that possibly there are *non-A*s that are *B*s, and that possibly there are *non-A*s that are *non-B*s. There is no symbol representing the possibility of *A*s that are not *B*s, because such entities are impossible given the premise. As Newell pointed out, this latter convention makes the notational system vulnerable to errors of omission. The system uses heuristic rules to combine strings into new strings and to compare the result with a given conclusion. The theory accounts for underlying competence in evaluating given conclusions rather than providing a theory of performance. It makes no predictions about errors, and it is aimed at illustrating how a theory of reasoning can be developed within the framework for studying problem solving devised by Newell and Simon (1972). The theory is a step toward mental models of different sorts of individuals, and it was superseded by the theory in the next section. Euler circles and Venn diagrams work for monadic assertions, but they need, at the very least, to be adapted for numerical or proportional quantifiers, and they fail completely for multiply quantified assertions, such as "All philosophers have read some books." An ideal theory calls for a more powerful representation.

The verbal models theory. Polk and Newell (1995) argued that reasoning is verbal, and they implemented a computer program called VR (for Verbal Reasoning) that constructs mental models from syllogistic premises and either formulates a conclusion from them or declares that nothing follows. For example, given premises of this sort:

Some B are A.
All B are C,

their VR program constructs an initial model of two sorts of individual:

B'		C
B'	A	C'

where the apostrophe denotes an "identifying" property, which is more accessible than other properties, because it derives from the subject of a premise (Polk & Newell, 1995, p. 539, Figure 5). The program uses this property to generate the putative conclusions *Some B are A* and *All B are C*. The program rejects these conclusions because they do not interrelate the end terms. What happens next depends on the version of the program. VR1 is the simplest version, and it does not use any indirect knowledge in reencoding the premises. Hence, it responds that nothing follows from these premises. It often fails to relate the end terms, and so it yields a large number of "nothing follows" responses (52 for the 64 pairs of syllogistic premises). VR2 has the ability to extract information from I and E premises equivalent to encoding their converses. It first attempts to extract information about *C* from each of them and then attempts to extract information about *A*. Its reencoding of the first premise yields information about *A*, that is, *Some A are B*, and so it can now construct an augmented model in which *A* is marked as an identifying property:

B'		C
	A'	
B'	A'	C

The program constructs various conclusions that fail to interrelate the end terms, and a legal conclusion, *Some A are C*, from the third individual in the preceding model. In general, VR2 yields many conclusions, and so it yields a smaller number of "nothing follows" responses than VR1 does (39 for the 64 pairs of syllogistic premises). VR3 extracts even more indirect information, including the equivalents of invalid conversions of *All X are Y* and *Some X are not Y* premises, and so it produces still fewer responses that nothing follows (eight for the 64 pairs of syllogistic premises).

Polk and Newell (1995) argued that the linguistic processes of encoding and reencoding are crucial to deduction, whereas other processes such as searching for alternative models are not. They wrote: "The point is that syllogism data can be accurately explained without positing a falsification strategy, in keeping with the hypothesis that such a reasoning-specific strategy is less important than verbal processes in explaining deduction" (p. 553). Hence, their theory is a fundamental departure from the use of falsification to test validity (cf. the mental model theory below). But evidence shows that individuals can and do use counterexamples (see our earlier account).

The source-founding theory. Stenning and Yule (1997) proposed an account of syllogisms that is neutral about the nature of mental representations and that aims to justify both formal rule accounts and model-based accounts, such as Euler circles. It provides a theory of competence and performance, and its central assumption is that reasoners represent those individuals that are necessary according to the premises. For example, consider again the premises

Some A are B.
All B are C.

The first premise necessitates the existence of at least one entity that is both *A* and *B* (see Table 4). The second premise necessitates that this entity is also *C*. And so it follows from the premises that *Some A are C*. The source premise of the critical individual that must exist, *Some A are B*, provides the linguistic foundation for the conclusion.

Stenning and Yule (1997) argued that their abstract algorithm can be implemented, with a few modifications, in both diagrammatic and logical systems; that is, the operations that their account calls for can be applied to the manipulation of Euler diagrams and to simplified propositional representations of quantified assertions. As we saw earlier, syllogistic assertions are normally represented in predicate calculus with quantifiers. In contrast, Stenning and Yule converted them into simpler representations akin to those in Newell's (1981) system for representing Venn diagrams (see above), by using negation (¬) and the sentential connectives of conjunction (&) and implication (→). The four moods of assertion are accordingly represented as follows:

All A are B:	A → B
Some A are B:	A & B
No A are B:	A → ¬B
Some A are not B:	A & ¬B

The source premise in the example above yields the individual *A & B*, and the implication into which the second premise is translated yields the individual *A & B & C*, under the rule for modus ponens. The source premise can then be used as a foundation for the linguistic form of the conclusion *Some A are C*. According to Stenning (personal communication, April 20, 2011), he and Yule "provided a novel task which allows figural effects to be studied in isolation from complexities introduced by syllogistic quantifiers in the conclusions. This task provides by far the most detailed data on figural effects available."

Stenning and Cox (2006) amplified the theory of performance by adding heuristics and implicatures to the source-founding account. They also contrasted classical logic, which they regarded as adversarial, with defeasible logic, which they regarded as cooperative. They argued: "We share much with Oaksford and Chater (2001) and with Bonnefon (2004). We agree that the classical logical competence model is an insufficient basis for modelling most subjects in these tasks [of syllogistic reasoning], and that their reasoning is often defeasible" (p. 1455). According to the source-founding theory, a substantial proportion of participants in experiments on syllogistic reasoning at least initially interpret the task as cooperative rather than in terms of the adversarial classical logic intended by the experimenter. As an example, consider the premises *Some A are B*, *Some B are C*, and their figural variations. They have no classically valid conclusion, and yet individuals often draw the conclusion *Some A are C*. So, their performance is "a matter of having a different goal—trying to cooperatively find the model that the author of the problem intends" (K. Stenning, personal communication, April 20, 2011).

The crux of the source-founding theory is the heuristics that individuals use to identify the source premise, and Stenning argued that there are many similarities here with Chater and Oaksford's (1996) PHM, including its attachment heuristic (see above). The performance model accordingly predicts that reasoners differ in their strategies. These individual differences and errors in performance occur as a result of factors that affect the choice of source premise and issues surrounding the selection of quantifiers for the conclusion (K. Stenning, personal communication, December 15, 2010). The most important heuristic treats any unique existential premise (I or O) as the source premise. If the quantifier contains an end term, it becomes the end term of the conclusion. Another heuristic prefers affirmative to negative premises as the source premise. A corollary is that premises based on *no* are the least preferred as sources. But according to Stenning and Cox (2006), "The source-founding model is a 'shell' process model, which abstracts over different logics and representations and in which different strategies can be expressed by changing the heuristics for source premise identification" (pp. 1641–1642). They then adopted a descriptive approach to finding "synoptic" patterns by examining performance in a task that required participants to evaluate immediate inferences from a single syllogistic premise. The participants had to decide, given the truth of the premise, whether the conclusion, which was also a syllogistic assertion, was definitely true, definitely false, or possibly true and possibly false. Some participants were rash in that they tended to respond that a conclusion was definitely true or definitely false, when in fact it was only possibly true given the premise. Other participants were hesitant in that they tended to respond that a conclusion was possibly true when it was definitely true, or possibly false when it was definitely false. Rashness and hesitancy also depended in part on the form of the conclusions. Some participants were rash when the order of subject and predicate differed from premise to conclusion. For example,

All of the A are B.
∴ All of the B are A.

Others were rash only when the order of subject and predicate was the same from premise to conclusion. For example,

Some of the A are B.
∴ Some of the A are not B.

Stenning and Cox categorized their participants into four main sorts: those who were logical (i.e., they were neither rash nor hesitant on any inference); those—the most frequent category—who were rash; those who were rash for inferences with the same order of subject and predicate in premise and conclusion, but otherwise logical; and those who were hesitant when the order of the subject and predicate differed from premise to conclusion, but otherwise rash. The participants then drew their own conclusions given a template of the eight possible sorts of conclusion from syllogistic premises. Their performance could be predicted, in part, by considering these individual differences.

The performance theory is complicated, and no computer implementation of it exists in the public domain. It promises a detailed account of differences in reasoning from one individual to another, which Stenning and his colleagues argued depend on different logics (see, e.g., Stenning & Van Lambalgen, 2008). It is not easy to see how naive individuals acquire either the underlying competence or the performance heuristics. Nor is it easy to determine empirically whether individuals have a family of logics for syllogisms or no coherent logic at all. What is clearly correct is to take seriously how individuals interpret assertions, including the monadic assertions in syllogisms, but not to treat all possible interpretations as equally justified (Stenning & Cox, 2006, p. 1477). What is missing is an account of which conclusions reflect errors in reasoning rather than an alternative logic, why some syllogisms are very easy and others very difficult, and the predictions of the theory for each of the 64 syllogisms, which have never been published. The general approach, however, is at odds with theories that seek to explain group data (Stenning & Cox, 2006, p. 1455). Our meta-analysis accordingly takes into account predictions about the different conclusions that different individuals should draw to each syllogism. The key question is whether the source-founding theory, by postulating adversarial and defeasible interpretations, gives a better account of the experimental results than other theories. In principle, the union of these predictions can be compared with the data in our meta-analysis.

The mental model theory. The mental model theory was formulated first for syllogisms (Johnson-Laird, 1975). Its account of competence posits that individuals grasp that an inference is no good if there is a counterexample to it, and that in making deductions they aim to maintain semantic information, to be parsimonious, and to reach a new conclusion (e.g., a relation that is not explicitly stated in the premises; Johnson-Laird & Byrne, 1991). The model theory postulates that individuals can represent a set *iconically* (in the sense of Peirce, 1958, Vol. 4): They build a mental model of its members, which is based on meaning and knowledge For example, *All of the artists are beekeepers* has a mental model, such as

```
artist   beekeeper
artist   beekeeper
artist   beekeeper
           . . .
```

Each row in this diagram denotes a representation of the properties of an individual, and the ellipsis denotes an implicit representation of the possibility of other sorts of individual who are not artists. This mental model can be fleshed out into various *fully explicit* models, such as

```
artist   beekeeper
artist   beekeeper
artist   beekeeper
         beekeeper
         beekeeper
           . . .
```

The artists are here a proper subset of the beekeepers; in another fully explicit model the two sets are coextensive. The initial mental model captures what is common to these two cases (Barwise, 1993).

How does the mental model above capture the fact that the set of artists is represented exhaustively, whereas the set of beekeepers is not? How does it avoid confusion with a mental model of three artists who are beekeepers? How is negation represented in models? One option is to use special symbols in models to represent these matters (see, e.g., Bucciarelli & Johnson-Laird, 1999; Khemlani, Orenes, & Johnson-Laird, in press). They are used in the computer program simulating the model theory (its source code in Lisp is available at http://mentalmodels.princeton.edu/programs/Syllog-Public.lisp). But another option is to maintain a separate *intensional* representation of the meaning of an assertion, which can be used to check that modifications to models are consistent with the assertion (Johnson-Laird, 2006, p. 129; Khemlani, Lotstein, & Johnson-Laird, 2012).

The model theory adopts a simple principle of parsimony: It maximizes the number of properties of each individual to try to keep the number of distinct sorts of individual to a minimum. The same number of tokens of the middle term occurs in the models of the two premises, and so it is easy to conjoin the two models to make a single initial model of both premises. Some syllogistic premises have only a single mental model: for example,

All A are B.
No B are C.

These premises yield the mental model (in the output of the computer program)

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

The square brackets are symbols indicating that a set is exhaustively represented (i.e., distributed), and "¬" is a symbol for negation. This model yields the conclusion *No A are C* or its converse. No model of the premises refutes these conclusions, and so they are both valid. The figural effect makes the preceding conclusion much more likely to be drawn than its converse.

In contrast, some premises have more than one mental model, and so a conclusion based on an initial model can be refuted by an alternative model. For example, the premises

All A are B.
Some B are C,

yield the mental model

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

This model yields the conclusion *Some A are C*. But the premises allow that the first individual in the model can be "broken" into two individuals:

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

Both premises hold in this model, but it is a counterexample to the conclusion, and no definite relation between the end terms holds in the two models above. Hence, the correct response to these premises is that nothing follows; that is, no simple definite conclusion holds between the end terms. Still other premises yield three models. In general, the model theory correctly predicts that one-model valid syllogisms are easier than two-model ones, which in turn are easier than three-model ones.

As Johnson-Laird (1983, p. 121) wrote: "The theory of mental models at least provides a framework suitable for describing individual differences, and even suggests some explanations for them." Such differences could arise in forming an integrated model of the premises, in searching for alternative models to ensure that a conclusion has no counterexamples, and in the ability to formulate verbal conclusions. Johnson-Laird reported a study with Bruno Bara showing that almost all the participants could cope with one-model syllogisms. The main difficulty in forming an integrated model is to hold a model in working memory while integrating the information from another premise. (A separate unpublished study with Jane Oakhill reported a reliable correlation, $\rho = .7$, between a measure of the processing capacity of working memory and accuracy in syllogistic reasoning.) In the study with Bara, two participants did not do well even with one-model syllogisms. Three participants appeared not to search for alternative models, as shown by their hardly ever responding that nothing followed from premises—instead they drew invalid conclusions consistent with just a single model of the premises. Other participants did search for alternatives but could not formulate conclusions common to them and their initial models. Hence, they almost invariably responded that nothing followed in the case of multiple-model problems. Another individual difference occurs in susceptibility to the figural effect. And there are also, as we discussed earlier, differences in individuals' strategies for syllogistic reasoning (Bucciarelli & Johnson-Laird, 1999). The focus of the model theory, however, has not been to give a detailed account of individual differences, but rather to explain how, in principle, individuals are able to reason with quantifiers, and the various sorts of responses that they are likely to make to syllogistic premises.

Iconic representations of sets have striking emergent properties. They yield upward entailing, downward entailing, and nonmonotone inferences. For example, Johnson-Laird and Steedman (1978, p. 92) showed how the following inference emerges from models that represent the relative sizes of the sets of fascists and authoritarians:

Most neofascists are authoritarian.
Most authoritarians are dogmatic.
∴ Many neofascists are dogmatic. (44)

In contrast, the following inference is not valid because the number of geniuses is tiny in proportion to the number of insane individuals:

Most geniuses are insane.
Most insane individuals are in asylums.
∴ Most geniuses are in asylums. (45)

Likewise, the following sort of nonmonotone inference is straightforward with models to represent sets:

Exactly three of the dogs are poodles.
All the dogs are feral.
∴ Exactly three of the poodles are feral. (46)

In short, mental models can represent any sort of relations between sets, and they yield as emergent consequences both monotone (upward and downward entailments) and nonmonotone inferences.

When individuals are asked to evaluate a given conclusion, they can search for counterexamples (i.e., alternative models of the premises that are inconsistent with the conclusion; Bucciarelli & Johnson-Laird, 1999; Kroger et al., 2008; Neth & Johnson-Laird, 1999). Models can accordingly explain the effects of beliefs on syllogistic reasoning (Cherubini, Garnham, Oakhill, & Morley, 1998; Newstead, Pollard, Evans, & Allen, 1992). When the content of the premises elicits an initial mental model that yields a preposterous conclusion, individuals search harder for a counterexample and are more likely to find it (Oakhill & Garnham, 1993; Oakhill & Johnson-Laird, 1985). But beliefs do affect one-model syllogisms too (Oakhill, Johnson-Laird, & Garnham, 1989; Gilinsky & Judd, 1994). One reason for this effect is that individuals examine the plausibility of their conclusions and are likely to be more cautious in accepting incredible conclusions than in accepting credible ones (Ball, Philips, Wade, & Quayle, 2006; Oakhill et al., 1989; Thompson, Striemer, Reikoff, Gunter, & Campbell, 2003). Their caution is a kind of response bias, but as Garnham and Oakhill (2005) showed, it is not of a kind that Klauer, Musch, and Naumer (2000) included in their multinomial model of individual differences.

The model theory postulates a principle of *truth*: Mental models represent what is true at the expense of what is false. This principle reduces the load that models place on working memory, but it has an unexpected consequence. Consider a monadic assertion, such as

Either all the coins are copper or some of the coins are copper. (53)

Granted that the disjunction is exclusive (both clauses cannot be true), it has two mental models: one representing that all the coins are copper and one representing that some but not all of the coins are copper. In contrast, its fully explicit models represent both what is true and what is false. An exclusive disjunction means that if the first clause is true (all the coins are copper), then the second clause is false (i.e., none of the coins is copper). This inconsistency establishes that the first clause cannot be true. Hence, the only fully explicit model is one in which the second clause is true and the first clause is false: some, but not all, of the coins are copper. The principle of truth leads participants into systematic fallacies in reasoning (e.g., Johnson-Laird, 2006, 2010; Johnson-Laird & Savary, 1999; Khemlani & Johnson-

Laird, 2009). Fallacies akin to the one above also occur in reasoning from quantified premises (Yang & Johnson-Laird, 2000a, 2000b).

An important part of monadic reasoning is the assessment of the consistency of a set of assertions, such as

Some of the artists are bakers.
Some of the bakers are chemists.
None of the artists are chemists. (54)

Participants are asked, "Could all of these assertions be true at the same time?" The correct answer is, "Yes." The model theory has a straightforward explanation of how individuals carry out the task: They try to construct a single model that satisfies all the assertions. One cause of difficulty is the principle of truth (Kunze, Khemlani, Lotstein, & Johnson-Laird, 2010). None of the other sorts of theory of monadic inference seems to explain how individuals are able to carry out the task of evaluating the consistency of sets of assertions. The best candidates among them appear to be theories based on first-order logic: Individuals choose an assertion in the set and then try to prove its negation from the remaining assertions. If, and only if, they succeed, then the original set is consistent. In our view, this procedure is too sophisticated for logically naive individuals, and in any case fails to make the correct predictions about the difficulty of the task (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000).

The weaknesses of the model theory fall into two categories: disputable and indisputable. Critics have claimed that the theory does not explain individual differences (e.g., Ford, 1995, p. 3; Stenning & Cox, 2006); that it cannot distinguish between representations of complete sets, proportions, and specific numbers of individuals (e.g., Geurts, 2003); and that models cannot represent large numbers of tokens (e.g., Newstead, 2003). Yet the discussion above shows how the model theory handles individual differences, and the other criticisms overlook that the theory has always allowed that models can be tagged with numerical and other sorts of information (Johnson-Laird, 1983, pp. 442–444) or accompanied by intensional representations. To forestall another potential criticism, the theory also allows that sets themselves can be directly represented by single mental tokens, with a constraint designed to prevent paradoxes (Johnson-Laird, 1983, p. 427 et seq.).

An indisputable weakness of the theory is that it originally postulated that individuals search for counterexamples as a matter of course. They do not, though they do have the ability to search for them (Bucciarelli & Johnson-Laird, 1999; Newstead, Handley, & Buck, 1999). And like other theories, it does not predict that syllogistic reasoning is sensitive to the phonological similarity of terms in the premises, at least when they are nonsense syllables (Ball & Quayle, 2009). Perhaps the only way to retain nonsense syllables in working memory is to rehearse their sound, with the well-established chance of confusions between similar-sounding syllables (Baddeley, Eysenck, & Anderson, 2009). But another possibility is that working memory for syllogistic reasoning is multimodal and does not depend solely on models of the world (see also Gilhooly, 2004; Gilhooly, Logie, & Wynn, 1999). Finally, a major gap in the theory is that it provides no account of how individuals develop strategies to cope with a series of monadic inferences. Although the theory provides a powerful account of monadic reasoning, this and another shortcoming to be described in the Discussion suffice to rule it out as the correct theory of the domain.

A Meta-Analysis of Theories of Syllogistic Reasoning

The aim of our meta-analysis was to evaluate seven current theories of syllogistic reasoning. In what follows, we describe the theories' predictions that we examined in the meta-analysis, our search in the literature for pertinent experimental results, an analysis of the reliability of these individual sets of data, and the outcome of the meta-analysis of the seven theories.

Theoretical Predictions

To conduct a meta-analysis, we needed a uniform way to compare each theory to the relevant data. Theories differ about what counts as correct; for example, Chater & Oaksford's (1999) probabilistic theory treats a wider variety of responses as correct than traditional logic, and so predictions about correctness were inappropriate. Similarly, some theories make predictions about the relative difficulty of different syllogisms, but others do not. The most general question that we could examine was, what conclusions does a theory predict that individuals should draw spontaneously to each of the 64 sorts of syllogism? For example, for premises of this sort:

All the A are B.
All the B are C,

one theory predicts these two responses:

All the A are C.
All the C are A,

whereas another theory predicts these responses:

All the A are C.
All the C are A.
Some of the A are C.
Some of the C are A.

The virtue of this dependent measure is that it is less susceptible to guessing than a task in which participants evaluate given conclusions, or select options from a multiple choice. Moreover, the dependent measure was used most often in the experiments in the literature that investigated all 64 pairs of syllogistic premises.

An accurate theory should predict the responses that occur reliably to each pair of premises, but a theory that predicts all nine possible responses (the eight possible conclusions and "nothing follows") to each syllogism would do perfectly on this measure. Hence, an accurate theory should also not predict responses that do not occur. Our measure for each syllogism was therefore the sum of the proportion of predicted responses that occurred (the hits) and the proportion of nonpredicted responses that did not occur (the correct rejections). As long as investigators can enumerate the possible responses, this measure can be applied to any sorts of inference. It also obviates the criticism that an analysis fails to take into account individual differences. Suppose, for example, that a theory postulates that some individuals aim solely for logical validity, whereas others draw conclusions that are merely possible. The theory as a whole predicts the union of these two sorts of predictions, and our meta-analysis can compare this union of responses to the data.

The Literature Search and Effect Size

To assess the predictions of the various theories, we needed to find experiments in which the participants drew their own conclusions to all 64 sorts of syllogism. We knew of several such studies, and we also searched the literature using PsycINFO, Google Scholar, and Medline for entries through 2011 using the following terms: *syllogis**, (*quant** and (*reasoning* or *inference*)), (*monad** and (*reasoning* or *inference*)), (*all* and *some* and (*none* or *no* or *not*)), (*figur** and *effect*). We likewise searched for dissertations, book chapters, and unpublished articles by consulting both reference sections and experts in the field. We excluded studies that did not examine all 64 pairs of premises or that examined only syllogisms with conclusions constrained to the Scholastic order *C–A* (e.g., Dickstein, 1978; and Rips, 1994, p. 233).

Any study of the conclusions that participants draw to syllogisms is likely to include some responses that are nothing more than "noise," that is, unsystematic errors that are attributable to temporary aberrations of one sort or another. Hence, we recoded the data in each experiment to isolate two sorts of response: those responses that occurred reliably more often than chance, which a theory should predict will occur, and those responses that did not occur reliably more often than chance, which a theory should predict will not occur. Conservatively, reasoners can make nine possible responses to syllogistic premises: conclusions in one of four moods and in one of two figures, and the response that nothing follows from the premises. The results in our analysis came from six experiments (see below), and so in each study we counted a response as reliable if it occurred significantly often, that is, in at least 16% of trials, assuming a prior chance probability of 1/9.

Table 5 summarizes the provenance of the six experiments that we included in our analysis, and Table 6 summarizes the aggregate results of these six studies. It presents each of the 27 valid syllogisms and each of the 37 invalid syllogisms with their respective percentages of the nine possible conclusions that occurred overall in the six studies.

The Reliability of the Data

Given a small number of studies carried out in Italy and in the United States, it was quite possible that their results would be too heterogeneous to provide an informative test of the theories. We therefore assessed their overall homogeneity in three independent ways. First, we examined their respective rank orders of difficulty in terms of logically correct responses (also shown in Table 6) using Kendall's coefficient of concordance (W), which ranges from 0 (*no consensus whatsoever*) to 1 (*perfect consensus*). The rank order of the difficulty of the 27 valid syllogisms over the six studies was high (Kendall's $W = .86$, $p < .001$), and the rank order was also high for the 37 problems with no valid conclusions interrelating the end terms ($W = .64$, $p < .001$). Second, we assessed the diversity of conclusions to each of the 64 syllogisms. Some syllogisms tended to elicit only one or two conclusions over the six studies, whereas others tended to elicit a greater variety of conclusions. We measured this diversity for each syllogism in each of the six studies using Shannon's information-theoretic measure of entropy (Shannon & Weaver, 1949). For each syllogism in each study, we computed the probability (p) with which each sort of conclusion was drawn and then aggregated the probabilities for a syllogism using Shannon's equation: $-\Sigma_i p_i \log_2 p_i$. We measured the correlation in the rank orders of these entropy scores over the six studies using Kendall's W. There was again a reliable concordance over the six studies (Kendall's $W = .45$, $p < .001$). Table 6 also shows the overall entropy of each of the 64 syllogisms. It correlated inversely with accuracy: The harder a problem, the greater the diversity of responses ($r = -0.52$). Finally, we conducted separate analyses that compared each theory's predictions against the data and found no violations of homogeneity in the overall results (with one exception, which we describe below).

We conclude that syllogisms show robust and consistent differences in difficulty over studies carried out in the United States and Italy. For instance, the easiest valid syllogism was *All the B are A*, *Some of the C are B*, which elicited 90% valid conclusions, and the hardest valid syllogism was *All the B are A*, *None of the C is B*, which elicited only 1% of valid conclusions (*Some of the A are not C*), with most participants inferring invalidly *None of the C is A* or its converse. Likewise, the easiest invalid syllogism was *None of the A is B*, *None of the C is B*, and it elicited 76% of correct responses that nothing follows interrelating the end terms, and the hardest invalid syllogism was *Some of the B are A*, *All the C are B*, which elicited only 12% of correct responses (nothing follows), with most participants inferring that *Some C are A*. In general, easy syllogisms elicit only one or two predominant responses (with an entropy measure of just over 1 bit), whereas difficult syllogisms elicit a greater variety of responses (with an entropy measure of over 2 bits). A syllogism is difficult when reasoners are uncertain about which conclusion, if any, to draw as opposed to certain about the wrong conclusion.

What both these measures establish is that the overall patterns of response in the six studies are robust. They occur despite the considerable differences from one person to another both in accuracy and in reasoning strategies. Overall, a by-materials analysis of the six studies shows that they elicit logically correct responses more often than chance in both valid syllogisms (Wilcoxon test, $z = 3.94$, $p < .001$) and invalid syllogisms (Wilcoxon test, $z = 5.30$, $p < .001$). It was therefore appropriate to assess how well the various theories predicted the responses to each of the 64 syllogisms in each study separately.

The Meta-Analysis of Seven Theories

For various reasons, we examined the predictions of only seven of the 12 theories. We did not include Euler circles, because they are more a system of representation than a specific psychological theory. We did not include Venn diagrams, because they offered no predictions about erroneous conclusions, and because their only proponent—the late Allen Newell—replaced them with an alternative theory (the verbal models of Polk & Newell, 1995). We did not include verbal substitution, because it does not make predictions about the sorts of erroneous responses people are likely to produce (M. Ford, personal communication, January 3, 2011). We did not include monotonicity theory, because the theory was not intended to be a full-blown account of syllogistic reasoning (B. Geurts, personal communication, December 22, 2010). We were not able to include the predictions of the source-founding theory: They have never been published for the 64 syllogisms, and we were unable to obtain them perhaps because its proponents are

Table 5
Provenance of the Six Experiments Included in the Meta-Analysis, Their Sample Sizes, Populations From Which the Participants Were Sampled, and Key Instructions

Provenance	Sample size	Population	Key instructions
Johnson-Laird & Steedman (1978), Experiment 2, first test	20	Students at Teachers College, Columbia University	What could be deduced with absolute certainty.
Johnson-Laird & Steedman (1978), Experiment 2, second test	20	Students at Teachers College, Columbia University	What could be deduced with absolute certainty.
Johnson-Laird & Bara (1984b), Experiment 3	20	Undergraduates at University of Milan, Italy	What followed necessarily from the premises.
Bara et al. (1995), 14- to 15-year-old group	20	High school in Florence, Italy	What followed necessarily from the premises.
Bara et al. (1995), adult group	20	Undergraduates at University of Florence, Italy	What followed necessarily from the premises.
Roberts et al. (2001)	56	Undergraduates at University of Florida, Gainesville	What followed necessarily from the premises.

critical of analyses of group results (K. Stenning, personal communication, April 20, 2011). The predictions of the remaining seven theories are summarized in Table 7, and we checked these predictions with the proponents of each of the theories. The verbal models theory's predictions were derived from the default versions of the authors' computational implementation. The theory does, in fact, generate other sets of predicted conclusions with different settings of the parameters of the implementation, but it is not possible to list them all (T. A. Polk, personal communication, December 29, 2010). They could improve the performance of the theory.

We compared the predictions of the theories to the data in each of the studies in Table 5, and we weighted the results of the comparison by the sample size of the study. The meta-analysis was conducted with the metafor package (Version 1.60; Viechtbauer, 2010) in the R statistical environment (Version 2.13.1; R Development Core Team, 2011). The results of the meta-analysis are presented in Table 8. It shows three point estimates for each theory: the proportion of hits, the proportion of correct rejections, and the overall proportion of correct predictions (the hits and correct rejections combined). Each analysis includes confidence intervals from a lower bound to an upper bound and results of a test showing that one can reject the null hypothesis that the theory made no correct predictions. We also tested whether each theory's performance was homogeneous, that is, whether it produced the same proportion of correct predictions no matter which experiment the data came from. Any deviation from homogeneity, as indicated by a high Higgins I^2 (Higgins & Thompson, 2002), which ranges from 0% to 100%, or by a significant Cochran's Q (Cochran, 1954), which is the traditional measure of heterogeneity, implies that the theory's predictions hold in some studies but not others. However, as Table 8 shows, all the outcomes of the meta-analysis of correct predictions were homogeneous, with the exception of marginal heterogeneity for the verbal models theory. Hence, researchers should exercise some caution in interpreting the predictions of the verbal models theory. Finally, we tested whether the results exhibit a publication bias, that is, whether the results of experiments with smaller sample sizes differ significantly from experiments that tested larger samples. Publication bias is a potential concern, because many of the studies under analysis came from the same laboratory. The reliability analysis presented earlier suggests that the results of the experiments are fairly consistent, but we used a more stringent regression test of publication bias (Egger, Davey Smith, Schneider, & Minder, 1997) on each theory's proportion of correct predictions. A significant result of the regression test for a particular theory implies that successful predictions depend on sample size. However, as Table 8 shows, no theory exhibited any reliable publication bias for the correct predictions. This result corroborates our reliability analysis and suggests that the experimental data are reasonably consistent across sample sizes.

The separate analyses of the hits, the correct rejections, and the overall correct predictions revealed differences among the theories. For instance, the mental model theory outperforms all other theories in its ability to predict the responses that participants produce (95%, $z = 47.09$, $p < .001$). However, the theory's capacity to reject responses that participants do not produce was relatively weak (74%, $z = 36.42$, $p < .001$). Likewise, the analysis of the theory's hits revealed a significant publication bias (regression test, $z = 2.03$, $p < .05$). The combined results suggest that the theory predicts too many conclusions, and one clear case is that the program from which the predictions derive draws conclusions both consistent and contrary to figural bias. A contrasting theory is the illicit conversion. It predicted too few conclusions (44% of hits, $z = 9.82$, $p < .001$), but it correctly rejected participants' nonresponses better than any other theory (93%, $z = 9.82$, $p < .001$). The theory that achieved the best balance between hits and correct rejections was the verbal models theory (63% hits, $z = 9.55$, $p < .001$; 90% correct rejections, $z = 80.50$, $p < .001$), and it accordingly yielded the highest proportion of correct predictions (84%, $z = 37.82$, $p < .001$). However, its predictions were marginally heterogeneous ($I^2 = 52\%$, $Q = 10.69$, $p = .06$), which indicates that its accuracy depended on the set of data with which it was compared.

In sum, all the theories perform much better than a baseline of zero correct predictions: They make more correct than incorrect predictions. We conducted two additional analyses that aggregated the predictions of the different families of theories. Across the seven theories in our analysis, four are based on heuristics, one is based on formal rules, and two are based on models. Since only one theory was based on formal rules, we omitted it from our aggregate analysis. In general, theories based on models performed better than heuristic theories (76% vs. 66%). However, the aggregated predictions performed worse than the highest performing

Table 6
Each Valid and Invalid Syllogism, Valid Conclusions, and Percentage of Correct Responses for the Syllogism Across the Six Studies in the Meta-Analysis, the Rank of a Syllogism's Difficulty, the Entropy of Participants' Responses, and the Percentages of the Nine Possible Conclusions

			Accuracy and rank order			Percentage of conclusions[a]								
Syllogism	Premises	Valid conclusions	% correct	Rank	Entropy (in bits)	Aac	Eac	Iac	Oac	Aca	Eca	Ica	Oca	NVC
Valid syllogisms														
AA1	Aab, Abc	Aac, Iac, Ica	88	2	1.08	81	2	6		1		1		1
AA2	Aba, Acb	Aca, Iac, Ica	54	16	1.83	35	1	3		48	1	4		1
AA4	Aba, Abc	Iac, Ica	16	53	2.02	49	1	12		10		4		22
AI2	Aba, Icb	Iac, Ica	90	1	1.33	2	1	20				71		4
AI4	Aba, Ibc	Iac, Ica	83	5	1.75	1	1	54	2	1		29		6
AE1	Aab, Ebc	Eac, Eca, Oac, Oca	87	4	1.08		78	1	1		8			1
AE2	Aba, Ecb	Oac	1	64	1.84		26	1	1		53		3	13
AE3	Aab, Ecb	Eac, Eca, Oac, Oca	81	7	1.77	1	33	1			47			10
AE4	Aba, Ebc	Oac	8	61	2.09	3	51		8		13		3	19
AO3	Aab, Ocb	Oca	40	34	2.44	1	3	6	13			9	40	20
AO4	Aba, Obc	Oac	54	17	2.14	1	3	9	54			4	7	14
IA1	Iab, Abc	Iac, Ica	88	3	1.05	6		82	1			6		1
IA4	Iba, Abc	Iac, Ica	81	6	1.83	1	1	44		1	1	37		5
IE1	Iab, Ebc	Oac	44	29	2.16		24	8	44		8			9
IE2	Iba, Ecb	Oac	13	57	2.49		16	3	13		29	1	5	27
IE3	Iab, Ecb	Oac	20	50	2.36		15	5	20		29		1	26
IE4	Iba, Ebc	Oac	28	47	2.48	1	15	6	28		14		3	29
EA1	Eab, Abc	Oca	3	63	1.52	1	71	1	1		11		3	10
EA2	Eba, Acb	Eac, Eca, Oac, Oca	78	9	1.89		28	2	1	1	51			8
EA3	Eab, Acb	Eac, Eca, Oac, Oca	80	8	1.67		63	1	1		17		3	10
EA4	Eba, Abc	Oca	9	60	2.19	1	45	3	1		21		9	16
EI1	Eab, Ibc	Oca	8	61	2.14		53	3	4		6	2	8	19
EI2	Eba, Icb	Oca	37	35	2.53		16	4	1		11	6	37	17
EI3	Eab, Icb	Oca	21	49	2.46		33	5	1		10	3	21	21
EI4	Eba, Ibc	Oca	15	55	2.50		28	4	5		8	2	15	32
OA3	Oab, Acb	Oac	36	39	2.29		1	19	36			1	10	22
OA4	Oba, Abc	Oca	42	33	2.42		3	13	12			8	42	13
Invalid syllogisms														
AA3	Aab, Acb	NVC	31	43	1.94	47	1	6		7		1		31
AI1	Aab, Ibc	NVC	16	53	1.46	2	1	70	1			4		16
AI3	Aab, Icb	NVC	37	38	1.85	2	1	13	1			43		37
AO1	Aab, Obc	NVC	14	56	1.80		3	10	62			2	6	14
AO2	Aba, Ocb	NVC	17	52	2.43	1	3	6	6			24	37	17
IA2	Iba, Acb	NVC	12	59	1.89	2	1	27	2	1		52		12
IA3	Iab, Acb	NVC	28	46	1.82		1	50	3			12		28
II1	Iab, Ibc	NVC	33	41	1.28		1	61				3		33
II2	Iba, Icb	NVC	30	45	1.85			25	1			39	1	30
II3	Iab, Icb	NVC	51	20	1.63	1	1	37	2			1		51
II4	Iba, Ibc	NVC	61	14	1.62	1	4	26	3			3		61
IO1	Iab, Obc	NVC	33	41	1.64		1	10	51		1			33
IO2	Iba, Ocb	NVC	49	22	2.22		1	6	10		1	13	15	49
IO3	Iab, Ocb	NVC	53	19	2.14	1	1	12	10		1	3	13	53
IO4	Iba, Obc	NVC	54	17	1.72		2	8	30		1		1	54
EE1	Eab, Ebc	NVC	44	29	1.48		46	1			3			44
EE2	Eba, Ecb	NVC	44	29	1.78	1	28				22			44
EE3	Eab, Ecb	NVC	76	10	1.12	1	17	1			2			76
EE4	Eba, Ebc	NVC	66	11	1.48	1	22	1	3		3			66
EO1	Eab, Obc	NVC	28	47	2.64	1	19	6	24		10	1	4	28
EO2	Eba, Ocb	NVC	47	27	2.27		12	1	3		4	13	15	47
EO3	Eab, Ocb	NVC	49	22	2.24		20	2	6		2	8	9	49
EO4	Eba, Obc	NVC	57	15	1.95		15	3	10		1		5	57
OA1	Oab, Abc	NVC	20	50	2.08	1		19	46			1	4	20
OA2	Oba, Acb	NVC	13	58	2.06		2	10	6		1	3	56	13
OI1	Oab, Ibc	NVC	36	39	2.20		1	24	25		1	2	5	36
OI2	Oba, Icb	NVC	31	43	2.32		1	7	10	1	1	8	38	31
OI3	Oab, Icb	NVC	49	22	2.12		1	13	19			4	8	49

Table 6 (*continued*)

Syllogism	Premises	Valid conclusions	Accuracy and rank order: % correct	Accuracy and rank order: Rank	Entropy (in bits)	Percentage of conclusions[a]: Aac	Eac	Iac	Oac	Aca	Eca	Ica	Oca	NVC
OI4	Oba, Ibc	NVC	47	27	2.24		3	6	11			13	16	47
OE1	Oab, Ebc	NVC	37	35	2.44		10	17	22		5	3	2	37
OE2	Oba, Ecb	NVC	51	20	2.22		10	4	3		18	2	6	51
OE3	Oab, Ecb	NVC	47	26	2.23		11	3	13		15		3	47
OE4	Oba, Ebc	NVC	49	22	2.28		7	4	13		11		6	49
OO1	Oab, Obc	NVC	37	35	1.73		1	7	47				3	37
OO2	Oba, Ocb	NVC	42	32	2.34		2	13	9	1	1	7	22	42
OO3	Oab, Ocb	NVC	64	13	1.53			8	22		1	1	2	64
OO4	Oba, Obc	NVC	66	11	1.53		3	3	21	1			3	66

Note. The sum of the various correct conclusions differs from the percent correct as a result of rounding errors. In AA1, the letters refer to the mood of the premises, and the number refers to the their figure. Aac = *All of the A are C*; Iac = *Some of the A are C*; Eac = *None of the A is C*; Oac = *Some of the A is not C*; NVC = no valid conclusion.
[a] The balance of the percentages is miscellaneous error.

individual model theory or the individual heuristic theory, which suggests that many of the aggregated predictions do not occur significantly often in the experiments. One potential route to a better theory may be to integrate models and heuristics, and we come back to this point in the Discussion.

There are caveats to our meta-analysis. An assessment based on the *possible* conclusions that a theory predicts rather than the *probable* conclusions that it predicts could be biased against certain sorts of theory. More important, what matters is not which theory yields the best predictions, but which theory is correct. And the answer to this question, sadly, is none of the above. Even the verbal models theory—which makes the best overall correct predictions—is far from fully accurate. The result of our meta-analysis is accordingly that none of the existing theories is correct. Investigators of reasoning need to develop a better theory of monadic reasoning.

Discussion

When individuals arrive in a psychological laboratory to participate in a study of monadic reasoning (i.e., reasoning about the properties of entities), they come equipped with knowledge of the meanings of quantifiers and with some deductive ability. During the course of a study, say, of syllogistic reasoning, their performance is likely to reflect the processing capacity of their working memories, and perhaps other differences that they bring to the laboratory. Their performance is also likely to improve—as it did in two sessions a week apart (Johnson-Laird & Steedman, 1978). Anyone confronted with the task of drawing conclusions from 64 pairs of syllogistic premises is likely to develop a strategy for coping. These strategies tend to differ from one person to another, though one-model syllogisms remain easier than multiple-model syllogisms for everyone whom we have ever tested. And some individuals even begin to grasp certain general principles (e.g., no valid syllogistic conclusion follows from two negative premises; see, e.g., Galotti et al., 1986). The moral is clear: Experiments on monadic reasoning do not aim at a stationary target. How, then, are cognitive scientists to converge on a comprehensive theory of monadic reasoning?

Among existing theories, those based on heuristics tend to downplay, or even to deny, systematic processes of valid reasoning (see, e.g., Revlis, 1975). Chater and Oaksford (1999) allowed that some individuals may reason deductively, but they also wrote:

> Everyday rationality does not depend on formal systems like logic and only formal rationality is constrained and error prone. . . . Everyday reasoning is probabilistic and people make errors in so-called logical tasks because they generalize these strategies to the laboratory. (Oaksford & Chater, 2001, p. 349).

Most adults are in fact capable of some monadic deductions, such as

> More than half of the artists are painters and sculptors.
> ∴ More than half of the artists are sculptors and painters.

The deduction could hardly be simpler, because it merely reorders the description of two properties (Geurts, 2003). People not only make such inferences, but also realize that the truth of the premise guarantees the truth of the conclusion. However, no one has complete deductive competence, because most domains of reasoning are computationally intractable (see, e.g., Garey & Johnson, 1979; Ragni, 2003). The inference above is trivial, but it is also remarkable, because none of the theories based on first-order logic can account for it.

Just as heuristic theories downplay deduction, theories based on logic downplay the role of heuristics. Yet, in several domains of higher cognition, a case can be made for so-called dual-process theories. This idea goes back at least to Pascal, who distinguished between intuitive and mathematical thinking (Pascal, 1670/1966, p. 211). Cognitive psychologists similarly distinguish between System 1, which makes rapid automatic inferences based on heuristics, and System 2, which makes slower conscious deliberations based on systematic and perhaps normative principles (see, e.g., Evans, 2003; Kahneman, 2011; Sloman, 1996; Stanovich, 1999; Verschueren, Schaeken, & d'Ydewalle, 2005). What is missing from these theories is an algorithmic account of the processes on which System 1 relies. From the beginnings of the model theory, however, it invoked two systems of reasoning, intuitive and deliberative, and it distinguished between them in terms of compu-

Table 7
Predicted Responses for Each Syllogism From Seven Theories of Syllogistic Reasoning

Syllogism	Atmosphere	Matching	Conversion	PHM	PSYCOP	Verbal models	Mental models
			Valid syllogisms				
AA1	Aac, Aca	Aac, Aca	Aac, Aca	Aac, Aca, Iac, Ica	Aac, Iac, Ica	Aac	Aac, Aca, Ica
AA2	Aac, Aca	Aac, Aca	Aac, Aca	Aac, Aca, Iac, Ica	Aca, Iac, Ica	Aca	Aca, Aac, Iac
AA4	Aac, Aca	Aac, Aca	Aac, Aca	Aac, Aca, Iac, Ica	Iac, Ica	NVC, Aca	Aac, Aca, Iac, Ica
AI2	Iac, Ica	Iac, Ica, Oac, Oca	Iac, Ica	Ica, Oca	Iac, Ica, Oac, Oca	Ica	Ica, Iac
AI4	Iac, Ica	Iac, Ica, Oac, Oca	Iac, Ica	Iac, Oac	Iac, Ica, Oac, Oca	NVC, Ica	Iac, Ica
AE1	Eac, Eca	Eac, Eca	Eac, Eca	Eac, Oac	Eac, Eca, Iac, Ica, Oac, Oca	Eac	Eac, Eca
AE2	Eac, Eca	Eac, Eca	Eac, Eca	Eca, Oca	Oac, Iac, Ica	NVC, Eca	Eac, Eca, Oca, Oac, NVC
AE3	Eac, Eca	Eac, Eca	Eac, Eca	Eca, Oca	Eac, Eca	NVC, Eac, Eca	Eac, Eca
AE4	Eac, Eca	Eac, Eca	Eac, Eca	Eac, Oac	Oac, Iac, Ica	NVC, Eac	Eac, Eca, Oac, Oca, NVC
AO3	Oac, Oca	Iac, Ica, Oac, Oca	Oac, Oca	Oca, Ica	Oca, Ica, Iac	NVC, Oca	Oac, Oca, NVC
AO4	Oac, Oca	Iac, Ica, Oac, Oca	Oac, Oca	Oac, Iac	Oac	NVC, Oac	Oac, Oca, NVC
IA1	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Iac, Oac	Iac, Ica, Oac, Oca	Iac	Iac, Ica
IA4	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Ica, Oca	Iac, Ica, Oac, Oca	NVC, Iac, Ica	Iac, Ica
IE1	Oac, Oca	Eac, Eca	Oac, Oca	Eac, Oac	Oac, Iac, Ica	Oac	Eac, Eca, Oac, Oca, NVC
IE2	Oac, Oca	Eac, Eca	Oac, Oca	Eca, Oca	Oac, Iac, Ica	NVC, Oac, Ica	Eca, Eac, Oca, Oac, NVC
IE3	Oac, Oca	Eac, Eca	Oac, Oca	Eca, Oca	Oac, Iac, Ica	NVC, Oac, Iac, Ica	Eac, Eca, Oac, Oca, NVC
IE4	Oac, Oca	Eac, Eca	Oac, Oca	Eac, Oac	Oac, Iac, Ica	NVC, Oac	Eac, Eca, Oac, Oca, NVC
EA1	Eac, Eca	Eac, Eca	NVC	Eac, Oac	Oca, Iac, Ica	NVC, Eca	Eac, Eca, Oac, Oca, NVC
EA2	Eac, Eca	Eac, Eca	NVC	Eca, Oca	Eac, Eca, Oac, Oca	Eca	Eca, Eac
EA3	Eac, Eca	Eac, Eca	NVC	Eac, Oac	Eac, Eca, Oac, Oca	NVC, Eca	Eac, Eca
EA4	Eac, Eca	Eac, Eca	NVC	Eca, Oca	Oca, Iac, Ica	NVC, Eca	Eac, Eca, Oac, Oca, NVC
EI1	Oac, Oca	Eac, Eca	NVC	Eac, Oac	Oca, Ica, Iac	NVC, Oca	Eac, Eca, Oac, Oca, NVC
EI2	Oac, Oca	Eac, Eca	NVC	Eca, Oca	Oca, Ica, Iac	Oca	Eca, Eac, Oca, Oac, NVC
EI3	Oac, Oca	Eac, Eca	NVC	Eac, Oac	Oca, Ica, Iac	NVC, Oca	Eac, Eca, Oac, Oca, NVC
EI4	Oac, Oca	Eac, Eca	NVC	Eca, Oca	Oca, Ica, Iac	NVC, Oca	Eac, Eca, Oac, Oca, NVC
OA3	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Iac	Oac, Iac, Ica	NVC, Oca	Oac, Oca, NVC
OA4	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oca, Ica	Oca, Ica, Iac	NVC, Oca	Oac, Oca, NVC
			Invalid syllogisms				
AA3	Aac, Aca	Aac, Aca	Aac, Aca	Aac, Aca, Iac, Ica	NVC	NVC, Iac, Aca	Aac, Aca, Iac, Ica, NVC
AI1	Iac, Ica	Iac, Ica, Oac, Oca	Iac, Ica	Iac, Oac	NVC	Iac	Iac, Ica, NVC
AI3	Iac, Ica	Iac, Ica, Oac, Oca	Iac, Ica	Ica, Oca	NVC	NVC, Iac, Ica	Iac, Ica, NVC
AO1	Oac, Oca	Iac, Ica, Oac, Oca	Oac, Oca	Oac, Iac	NVC	Oac	Oac, Oca, NVC
AO2	Oac, Oca	Iac, Ica, Oac, Oca	Oac, Oca	Oca, Ica	NVC	NVC, Oca	Oca, Oac, NVC
IA2	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Ica, Oca	NVC	NVC, Ica	Iac, Ica, NVC
IA3	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Iac, Oac	NVC	NVC, Iac, Ica	Iac, Ica, NVC
II1	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Iac, Ica, Oac, Oca	NVC	Iac	Iac, Ica, NVC
II2	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Iac, Ica, Oac, Oca	NVC	NVC	Ica, Iac, NVC
II3	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Iac, Ica, Oac, Oca	NVC	NVC	Iac, Ica, NVC
II4	Iac, Ica	Iac, Ica, Oac, Oca	NVC	Iac, Ica, Oac, Oca	NVC	NVC, Ica	Iac, Ica, NVC
IO1	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Iac	NVC	Oac, Iac	Oac, Oca, NVC
IO2	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oca, Ica	NVC	NVC, Ica	Oca, Oac, NVC
IO3	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oca, Ica	NVC	NVC, Iac, Ica	Oac, Oca, NVC
IO4	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Iac	NVC	NVC, Oac	Oac, Oca, NVC
EE1	Eac, Eca	Eac, Eca	NVC	Eac, Eca, Oac, Oca	NVC	NVC, Eac	Eac, Eca, NVC
EE2	Eac, Eca	Eac, Eca	NVC	Eac, Eca, Oac, Oca	NVC	NVC, Eca	Eca, Eac, NVC
EE3	Eac, Eca	Eac, Eca	NVC	Eac, Eca, Oac, Oca	NVC	NVC, Iac, Aca	Eac, Eca, NVC
EE4	Eac, Eca	Eac, Eca	NVC	Eac, Eca, Oac, Oca	NVC	NVC	Eac, Eca, NVC
EO1	Oac, Oca	Eac, Eca	NVC	Oac, Iac	NVC	NVC, Oac	Eac, Eca, Oac, Oca, NVC
EO2	Oac, Oca	Eac, Eca	NVC	Oca, Ica	NVC	NVC, Oca	Eca, Eac, Oca, Oac, NVC
EO3	Oac, Oca	Eac, Eca	NVC	Oca, Ica	NVC	NVC, Iac, Ica	Eac, Eca, Oac, Oca, NVC
EO4	Oac, Oca	Eac, Eca	NVC	Oac, Iac	NVC	NVC	Eac, Eca, Oac, Oca, NVC
OA1	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Iac	NVC	NVC, Oca	Oac, Oca, NVC
OA2	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oca, Ica	NVC	NVC, Oca	Oca, Oac, NVC
OI1	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Iac	NVC	NVC, Oca	Oac, Oca, NVC
OI2	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oca, Ica	NVC	NVC	Oca, Oac, NVC
OI3	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Iac	NVC	NVC, Oca	Oac, Oca, NVC
OI4	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oca, Ica	NVC	NVC, Oca	Oac, Oca, NVC
OE1	Oac, Oca	Eac, Eca	NVC	Oac, Iac	NVC	NVC, Oac	Eac, Eca, Oac, Oca, NVC
OE2	Oac, Oca	Eac, Eca	NVC	Oca, Ica	NVC	NVC, Oca	Eca, Eac, Oca, Oac, NVC
OE3	Oac, Oca	Eac, Eca	NVC	Oac, Iac	NVC	NVC, Oca	Eac, Eca, Oac, Oca, NVC
OE4	Oac, Oca	Eac, Eca	NVC	Oca, Ica	NVC	NVC	Eac, Eca, Oac, Oca, NVC

Table 7 (*continued*)

Syllogism	Atmosphere	Matching	Conversion	PHM	PSYCOP	Verbal models	Mental models
OO1	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Oca, Iac, Ica	NVC	NVC	Oac, Oca, NVC
OO2	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Oca, Iac, Ica	NVC	NVC, Oca	Oca, Oac, NVC
OO3	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Oca, Iac, Ica	NVC	NVC, Oca	Oac, Oca, NVC
OO4	Oac, Oca	Iac, Ica, Oac, Oca	NVC	Oac, Oca, Iac, Ica	NVC	NVC	Oac, Oca, NVC

Note. PHM = probability heuristics model; PSYCOP = PSYCOP with Gricean implicatures; Aac = *All of the A are C*; Iac = *Some of the A are C*; Eac = *None of the A is C*; Oac = *Some of the A are not C*; NVC = no valid conclusion.

tational power (Johnson-Laird, 1983, Chapter 6). The intuitive system (System 1) makes no use of working memory for intermediate computations and is therefore equivalent in power to a finite-state automaton. It works with just a single mental model at a time and can make no comparisons between models. The deliberative system (System 2), in contrast, has access to working memory and can therefore carry out recursive processes, at least until they become too demanding on its processing capacity. It can therefore consider alternative models of the same premises.

Individuals arriving at the laboratory could be equipped with a single system for monadic reasoning—it would need to be a deliberative theory that fails often enough to cause systematic errors. Alternatively, individuals could rely on dual processes: System 1 to generate intuitive conclusions and System 2 to check them. They could even revert from difficult deliberations in System 2 to the intuitions of System 1 and make a guess constrained by relevant information, such as the mood of the premises. At present, no decisive evidence exists to decide between these two

Table 8
Meta-Analysis of Hits, Correct Rejections, and Correct Predictions From Seven Theories of Syllogistic Reasoning

		95% CI		Test of null		Heterogeneity			Publication bias	
Theory	Effect size	*LL*	*UL*	*z*	*p*	I^2 (%)	*Q*	*p*	*z*	*p*
Point estimate: Hits[a]										
Mental models theory	95	91	99	47.09	<.001	0	5.47	.36	2.03	.04
Verbal models theory	63	50	76	9.55	<.001	56	11.12	.05	0.02	.98
Matching	59	48	70	10.24	<.001	38	8.09	.15	1.56	.12
Probability heuristics model	52	43	60	11.34	<.001	0	0.75	.98	0.44	.66
Atmosphere	51	42	59	11.12	<.001	0	0.71	.98	0.78	.44
PSYCOP model	44	35	53	9.92	<.001	0	4.47	.48	0.88	.38
Conversion	44	35	53	9.82	<.001	0	2.66	.75	0.59	.56
Point estimate: Correct rejections[b]										
Conversion	93	91	96	80.60	<.001	0	0.88	.97	0.44	.66
Verbal models theory	90	87	93	64.29	<.001	0	3.01	.69	0.48	.63
Atmosphere	85	82	89	51.68	<.001	0	0.96	.97	0.92	.37
PSYCOP model	85	82	88	50.59	<.001	0	1.17	.95	0.11	.92
Probability heuristics model	79	75	82	40.91	<.001	0	0.64	.99	0.47	.64
Mental models theory	74	70	78	36.42	<.001	0	0.88	.97	0.83	.40
Matching	74	70	78	36.13	<.001	0	3.12	.68	1.08	.28
Point estimate: Correct predictions[c]										
Verbal models theory	84	80	89	37.82	<.001	52	10.69	.06	1.03	.30
Conversion	83	80	86	54.05	<.001	0	3.62	.60	0.22	.82
Mental models theory	78	75	81	45.46	<.001	0	2.80	.73	1.40	.82
Atmosphere	78	75	81	45.39	<.001	0	1.16	.95	1.03	.30
PSYCOP model	77	73	80	43.54	<.001	1	5.07	.41	0.16	.87
Probability heuristics model	73	69	77	39.43	<.001	0	0.84	.97	0.53	.60
Matching	71	66	75	29.32	<.001	34	7.61	.18	1.55	.11
Aggregate										
Model-based theories	76	73	80	42.97	<.001	0	2.25	.81	1.25	.21
Heuristic-based theories	66	62	70	33.38	<.001	0	1.80	.88	1.19	.23

Note. The table includes analyses of the aggregated correct predictions of model- and heuristic-based theories. Effect sizes are measured as proportions (of hits, correct rejections, and correct predictions). CI = confidence interval; *LL* = lower limit; *UL* = upper limit; I^2 = Higgins's measure of heterogeneity; *Q* = Cochran's statistic.
[a] A prediction that was corroborated by >16% of participants in a given experiment. [b] Any nonprediction that was corroborated by <16% of participants in a given experiment. [c] Hits and correct rejections.

alternatives, though dual systems seem plausible because of their evident role in other sorts of higher cognition (e.g., Kahneman & Frederick, 2002). What complicates matters is that the processes of System 2 are imperfect: Individuals can reason in a conscious and deliberate way and still err (see, e.g., Johnson-Laird & Byrne, 1991). Likewise, they err in deliberating about probabilities. Hence, systematic errors are not in themselves definitive cues to the operations of System 1. Perhaps a better guide is the set of responses that individuals make when they have only a brief time to make inferences. Hence, we need studies of the rapid guesses that people make given premises about the properties of individuals.

One such study allowed participants only 10 s in which to respond to each of the 64 pairs of premises (Johnson-Laird & Bara, 1984b, Experiment 1). Many of their responses to the valid responses were correct, and so a better clue to heuristics is likely to be the predominant responses to the 37 invalid syllogisms. The results showed considerable variation in the frequencies of the correct "nothing follows" responses (from 100% down to 15%). The figure of the premises when their mood is held constant had a reliable effect on the proportion of these correct responses: Figure 1, 50%; Figure 2, 57%; Figure 3, 70%; and Figure 4, 86% (Jonckheere test, $z = 3.32$, $p < .001$). This phenomenon makes atmosphere and matching less likely heuristics for System 1, because they have no machinery to explain why figure affects the response that nothing follows. In contrast, the PHM includes the attachment heuristic, which postulates that if the least informative premise has an end term as its subject, it is the subject of the conclusion; otherwise, the end term in the other premise is the subject of the conclusion. It also includes the max-heuristic, which postulates that individuals' confidence in a conclusion is proportional to the informativeness of the most informative premise (the max-premise), and so they will respond that nothing follows if the max-premise has a low informativeness.

Whatever its fundamental mechanism, System 1 needs to extend to rapid inferences from premises, such as

Exactly two of the artists are cartographers.
All the artists are lithographers.

Individuals are likely to guess the following invalid conclusion or its equivalent converse:

Exactly two of the cartographers are lithographers.

A comprehensive account of System 1 accordingly calls for a set of heuristics akin to those of the PHM, and perhaps a system in which individuals generate possible conclusions or perhaps probable conclusions based on mental models (for analogous suggestions about conditional reasoning, see Geiger & Oberauer, 2010; Oaksford & Chater, 2010).

Once individuals have a putative conclusion, they can examine it in a slower, more deliberative way using System 2. Common to logical and set-theoretic accounts is the assumption that individuals are capable of valid deductive reasoning. The two sorts of theory diverge on whether the principles of reasoning are implemented in terms of formal rules of inference or in terms of semantic—that is, model-theoretic—principles of logic (see Jeffrey, 1981). On this point, we argue that monadic reasoning is better explained in semantic terms, and in particular in terms of relations among sets. Such a claim is founded on several arguments.

First, set-theoretic relations generalize to accommodate quantifiers such as *more than half*, which cannot be represented in first-order logic (Barwise & Cooper, 1981; Johnson-Laird, 1983, p. 140). Set-theoretic representations can accommodate them, but they need to go beyond the orthodox use of Euler circles and Venn diagrams, which are likewise inadequate if the theory is to be generalized to inferences from premises containing multiple quantifiers.

Second, two sorts of inference are fundamental in reasoning about sets, and they are an emergent consequence of the iconic representation of sets. One sort of inference is that an individual is a member of a set: for example,

Every composer is a musician.
Adams is a composer.
∴ Adams is a musician.

The other sort of inference is that an individual is not a member of a set: for example,

Every composer is a musician.
Koons is not a musician.
∴ Koons is not a composer. (64)

Once reasoners can make these inferences about individuals, they can cope in principle with monadic inferences about properties (Johnson-Laird, 2006, Chapter 10). When entities are in a set, they inherit any properties that hold for the complete set. Indeed, given a premise asserting that a property holds for an entire set, another premise based on almost any sort of affirmative quantifier yields a valid inference: for example,

Every composer is a musician.
All/most/many/more than half/at least two of the students are composers.
Therefore, all/most/many/more than half/at least two of the students are musicians.

The only exceptions are those quantifiers limiting the number of members of a set (i.e., nonmonotone quantifiers). For instance, the following inference is invalid:

Every composer is a musician.
Exactly two of the students are composers.
Therefore, exactly two of the students are musicians.

A counterexample is one in which a third student is not a composer but is nevertheless a musician, though the premises do yield the valid conclusion that at least some students are musicians.

Likewise, when entities are not in a set, they are not in any of its subsets:

Every composer is a musician.
None/few/less than half/no more than two of the students are musicians.
Therefore, none/few/less than half/no more than two of the students are composers.

We suspect that Aristotle was aware of these principles. Certainly, the Scholastic logicians were: They argued that at least one premise in a valid inference has to have a distributed middle term. And, of course, the preceding inferences are monotonic except for the valid deduction from *exactly two of the students* (Barwise & Cooper, 1981).

Third, another argument for the primacy of sets concerns claims about the membership, not the inclusion, of one set within another. A so-called generic assertion is a common way in which to express such claims, such as *Horses are of varied heights*. Class membership is not transitive, and so the further premise, *Secretariat is horse*, does not yield the inference that Secretariat is of varied heights. Universal quantifiers cannot normally be used for assertions that one set is a member of another (cf. the oddity of *All horses are of varied heights*). The generic assertion calls for a mental model of one set as a member of another (Johnson-Laird, 1983, p. 427).

One major shortcoming of all 12 of the theories of monadic reasoning is their lack of any system that determines how to carry out a given inferential task. As we have shown earlier, monadic assertions enter into a variety of tasks. They include the following, which have all been investigated experimentally:

- the evaluation of a stated conclusion to determine whether, given the premises, it follows that the conclusion is *necessarily* the case;
- the similar task of assessing whether it follows that the conclusion is *possibly* the case;
- the spontaneous formulation of such conclusions from given premises;
- the assessment of whether or not a set of assertions is consistent (i.e., whether they could all be true at the same time);
- the spontaneous formulation of a missing premise, which, taken in conjunction with a given premise, suffices to imply that a given conclusion follows validly;
- given a putative but invalid inference, the formulation of a counterexample that refutes it.

The list is hardly exhaustive, but most individuals have little difficulty in understanding any of these tasks, or in tacitly devising a mental procedure that will carry them out. This ability is largely taken for granted by students of reasoning, and as we have shown, many of the theories in Table 3 are geared to only one or two of these tasks. For several of the theories, we could not discern any obvious way in which their accounts could be extended to some of the tasks on the list (e.g., the evaluation of consistency). And we know of no algorithmic account of how a control system translates verbal instructions and examples into an appropriate procedure for carrying out a required inferential task.

One hundred years ago psychologists had no theory of syllogistic reasoning. Thirty-five years ago they had only heuristic accounts that explained biases and errors, and so the domain appeared to be an excellent test case for cognitive science. There are now 12 sorts of theories of syllogisms and monadic inferences, and so skeptics may well conclude that cognitive science has failed: It yields no consensus about a small, empirically tractable domain of reasoning. Several ways might lead out of the impasse. One way is to argue that all the theories are correct. The obvious retort is that they make inconsistent predictions. But different people develop different strategies, and even the same person may use different strategies on different occasions. Yet, to make a union of all existing theories is simplistic, and the result is likely to yield a small proportion of correct predictions for the data in Table 6. A better way out of the impasse is to frame constraints on a new theory. It will have roots in existing accounts, but it should do a better job than any of them.

A unified theory of monadic reasoning should explain

- the interpretation and mental representation of monadic assertions, including syllogistic premises;
- what the brain computes and how it carries out all inferential tasks with such assertions;
- the differences in difficulty from one inference to another, and common errors;
- how contents affect performance;
- how the ability to reason with monadic assertions develops;
- differences in performance from one person to another, which are likely to reflect the processing capacity of working memory, experience in deductive reasoning tasks, and motivation (these individual differences, however, call for much more research).

The theory should ideally be implemented in a computer program, its predictions for the 64 syllogisms should be clear, and it should outperform all other extant theories. The best guide to the construction of a new theory is likely to be the responses in Table 6 that occurred overall more often than chance. Any theory that can account for all and only those responses is also likely to outperform the theories in the meta-analysis. In a new theory of syllogistic reasoning (Khemlani, Lotstein, & Johnson-Laird, 2011), the present authors have tried to meet the six goals above in an account that combines both heuristics and mental models. However, there are additional constraints on an ideal computational implementation of the theory. In particular, when an individual arrives in the laboratory, a few measures of the individual's abilities not directly related to reasoning should set the parameters of the implementation. It should then make accurate predictions about the individual's performance in the relevant reasoning task. The measures themselves, such as the processing capacity of working memory, or the ability to understand quantifiers, should be based on a causal account of intellectual development. A theory of this sort, which explains differences in the difficulty of inferences, differences in accuracy from one person to another, and differences in their strategies of reasoning, will vindicate the cognitive science of higher cognition.

The choice of syllogisms as a test case may be a mistake—we heed the late Allen Newell's dictum that one cannot play "20 Questions" with nature and win (Newell, 1973). But we remain convinced that a theory of monadic reasoning is a feasible goal for the discipline. Its solution is likely to depend on a corroborated account of three main components: a control system, which maps the description of inferential tasks into appropriate procedures for the deliberative system to carry out; a heuristic system that comes into play at the outset or at a point where individuals find themselves forced to guess; and finally, a deliberative system that is likely to be set-theoretic in nature.

References

References marked with an asterisk indicate studies included in the meta-analysis.

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

Bacon, A. M., Handley, S. J., & Newstead, S. E. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking & Reasoning, 9,* 133–168. doi:10.1080/13546780343000196

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory.* Hove, England: Psychology Press.

Ball, L. J., Philips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology, 53,* 77–86. doi:10.1027/1618-3169.53.1.77

Ball, L. J., & Quayle, J. D. (2009). Phonological and visual distinctiveness effects in syllogistic reasoning: Implications for mental models theory. *Memory & Cognition, 37,* 759–768. doi:10.3758/MC.37.6.759

*Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). The development of syllogistic reasoning. *American Journal of Psychology, 108,* 157–193. doi:10.2307/1423127

Barnes, J. (Ed.). (1984). *The complete works of Aristotle* (Vol. 1). Princeton, NJ: Princeton University Press.

Barwise, J. (1993). Everyday reasoning and logical inference. *Behavioral and Brain Sciences, 16,* 337–338. doi:10.1017/S0140525X00030314

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy, 4,* 159–219. doi:10.1007/BF00350139

Barwise, J., & Etchemendy, J. (1994). *Hyperproof.* Stanford, CA: CSLI.

Begg, I., & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. *Journal of Experimental Psychology, 81,* 351–354. doi:10.1037/h0027770

Beth, E. W., & Piaget, J. (1966). *Mathematical epistemology and psychology.* Dordrecht, the Netherlands: Reidel.

Bonnefon, J.-F. (2004). Reinstatement, floating conclusions, and the credulity of mental model theory. *Cognitive Science, 28,* 621–631. doi:10.1207/s15516709cog2804_6

Boolos, G. (1984). On "syllogistic inference." *Cognition, 17,* 181–182. doi:10.1016/0010-0277(84)90018-0

Braine, M. D. S. (1998). Steps toward a mental-predicate logic. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 273–331). Mahwah, NJ: Erlbaum.

Braine, M. D. S., & Rumain, B. (1983). Logical reasoning. In J. H. Flavell & E. M. Markman (Eds.), *Carmichael's handbook of child psychology, Vol. 3. Cognitive development* (4th ed., pp. 263–339). New York, NY: Wiley.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23,* 247–303. doi:10.1207/s15516709cog2303_1

Ceraso, J., & Provitera, A. (1971). Sources of error in syllogistic reasoning. *Cognitive Psychology, 2,* 400–410. doi:10.1016/0010-0285(71)90023-5

Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology, 58,* 220–226. doi:10.1037/h0041961

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38,* 191–258. doi:10.1006/cogp.1998.0696

Cherubini, P., Garnham, A., Oakhill, J., & Morley, E. (1998). Can any ostrich fly? Some new data on belief bias in syllogistic reasoning. *Cognition, 69,* 179–218. doi:10.1016/S0010-0277(98)00064-X

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Clark, H. H. (1974). Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12, pp. 1291–1428). The Hague, the Netherlands: Mouton.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10,* 101–129. doi:10.2307/3001666

Cohen, M. R., & Nagel, E. (1934). *An introduction to logic and scientific method.* London, England: Routledge & Kegan Paul.

Dias, M., Roazzi, A., & Harris, P. L. (2005). Reasoning from unfamiliar premises: A study with unschooled adults. *Psychological Science, 16,* 550–554. doi:10.1111/j.0956-7976.2005.01573.x

Dickstein, L. S. (1978). Error processes in syllogistic reasoning. *Memory & Cognition, 6,* 537–543. doi:10.3758/BF03198242

Edwards, A. W. F. (2004). *Cogwheels of the mind: The story of Venn diagrams.* Baltimore, MD: Johns Hopkins University Press.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. doi:10.1136/bmj.315.7109.629

Erickson, J. R. (1974). A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (Ed.), *Loyola symposium on cognition* (Vol. 2, pp. 305–330). Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454–459. doi:10.1016/j.tics.2003.08.012

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11,* 295–306. doi:10.3758/BF03196976

Evans, J. St. B. T., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1495–1513. doi:10.1037/0278-7393.25.6.1495

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition, 54,* 1–71. doi:10.1016/0010-0277(94)00625-U

Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General, 115,* 16–25. doi:10.1037/0096-3445.115.1.16

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

Garnham, A., & Oakhill, J. V. (2005). Accounting for belief bias in a mental model framework: Comment on Klauer, Musch, and Naumer (2000). *Psychological Review, 112,* 509–517. doi:10.1037/0033-295X.112.2.509

Geach, P. (1962). *Reference and generality: An examination of some medieval and modern theories.* Ithaca, NY: Cornell University Press.

Geiger, S. M., & Oberauer, K. (2010). Towards a reconciliation of mental model theory and probabilistic theories of conditionals. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 289–307). Oxford, England: Oxford University Press.

Geurts, B. (2003). Reasoning with quantifiers. *Cognition, 86,* 223–251. doi:10.1016/S0010-0277(02)00180-4

Geurts, B., & van der Slik, F. (2005). Monotonicity and processing load. *Journal of Semantics, 22,* 97–117. doi:10.1093/jos/ffh018

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669. doi:10.1037/0033-295X.103.4.650

Gilhooly, K. J. (2004). Working memory and reasoning. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 49–77). Cambridge, England: Cambridge University Press.

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology, 11,* 473–498. doi:10.1080/095414499382264

Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging, 9,* 356–371. doi:10.1037/0882-7974.9.3.356

Goel, V., Buchel, C., Frith, C., Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage, 12,* 504–514. doi:10.1006/nimg.2000.0636

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Vol. 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.

Guyote, M. J., & Sternberg, R. J. (1981). A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology, 13,* 461–525. doi:10.1016/0010-0285(81)90018-9

Hardman, D. K., & Payne, S. J. (1995). Problem difficulty and response format in syllogistic reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48A,* 945–975.

Hertwig, R., Ortmann, A., & Gigerenzer, G. (1997). Deductive competence: A desert devoid of content and context. *Current Psychology of Cognition, 16,* 102–107.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558. doi:10.1002/sim.1186

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology, 38,* 883–894. doi:10.1037/0012-1649.38.6.883

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child.* London, England: Routledge & Kegan Paul.

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

Johnson-Laird, P. N. (1970). The interpretation of quantified sentences. In G. B. Flores D'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics* (pp. 347–372). Amsterdam, the Netherlands: North-Holland.

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, MA: Harvard University Press.

Johnson-Laird, P. N. (2002). Peirce, logic diagrams, and the elementary operations of reasoning. *Thinking & Reasoning, 8,* 69–95. doi:10.1080/13546780143000099

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

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 18243–18250. doi:10.1073/pnas.1012933107

Johnson-Laird, P. N., & Bara, B. G. (1984a). Logical expertise as a cause of error: A reply to Boolos. *Cognition, 17,* 183–184. doi:10.1016/0010-0277(84)90019-2

*Johnson-Laird, P. N., & Bara, B. G. (1984b). Syllogistic inference. *Cognition, 16,* 1–61. doi:10.1016/0010-0277(84)90035-0

Johnson-Laird, P. N., & Byrne, R. M. J. (1989). *Only* reasoning. *Journal of Memory and Language, 28,* 313–330. doi:10.1016/0749-596X(89)90036-3

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

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory & Cognition, 31,* 1105–1113. doi:10.3758/BF03196131

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science, 288,* 531–532. doi:10.1126/science.288.5465.531

Johnson-Laird, P. N., Mancini, F., & Gangemi, A. (2006). A hyper emotion theory of psychological illnesses. *Psychological Review, 113,* 822–841. doi:10.1037/0033-295X.113.4.822

Johnson-Laird, P. N., Oakhill, J., & Bull, D. (1986). Children's syllogistic reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38A,* 35–58.

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition, 71,* 191–229. doi:10.1016/S0010-0277(99)00015-3

*Johnson-Laird, P. N., & Steedman, M. J. (1978). The psychology of syllogisms. *Cognitive Psychology, 10,* 64–99. doi:10.1016/0010-0285(78)90019-1

Kahneman, D. (2011). *Thinking, fast and slow.* New York, NY: Farrar, Straus & Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY: Cambridge University Press.

Khemlani, S., & Johnson-Laird, P. N. (2009). Disjunctive illusory inferences and how to eliminate them. *Memory & Cognition, 37,* 615–623. doi:10.3758/MC.37.5.615

Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2008). Syllogistic reasoning with generic premises: The generic overgeneralization effect. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 619–624). Austin, TX: Cognitive Science Society.

Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2012). *A unified theory of syllogistic reasoning.* Manuscript submitted for publication.

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

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107,* 852–884. doi:10.1037/0033-295X.107.4.852

Kneale, W., & Kneale, M. (1962). *The development of logic.* Oxford, England: Oxford University Press.

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. doi:10.1016/j.brainres.2008.07.128

Kunze, N., Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2010). Illusions of consistency in quantified assertions. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2028–2032). Austin, TX: Cognitive Science Society.

Lee, N. Y. L., Goodwin, G. P., & Johnson-Laird, P. N. (2008). The psychological problem of Sudoku. *Thinking & Reasoning, 14,* 342–364. doi:10.1080/13546780802236308

Makinson, D. (1969). Remarks on the concept of distribution in traditional logic. *Noûs, 3,* 103–108. doi:10.2307/2216161

Mazzocco, A., Legrenzi, P., & Roncato, S. (1974). Syllogistic inference: The failure of the atmosphere effect and the conversion hypothesis. *Italian Journal of Psychology, 2,* 157–172.

Montague, R. (1974). *Formal philosophy: Selected papers.* New Haven, CT: Yale University Press.

Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology, 14,* 237–255. doi:10.1002/(SICI)1099-0720(200005/06)14:3<237::AID-ACP641>3.0.CO;2-R

Moxey, L. M., Sanford, A. J., & Dawydiak, E. J. (2001). Denials as controllers of negative quantifier focus. *Journal of Memory and Language, 44,* 427–442. doi:10.1006/jmla.2000.2736

Neimark, E. D., & Chapman, R. H. (1975). Development of the comprehension of logical quantifiers. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults* (pp. 135–151). Hillsdale, NJ: Erlbaum.

Neth, H., & Johnson-Laird, P. N. (1999). The search for counterexamples in human reasoning. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (p. 806). Austin, TX: Cognitive Science Society.

Newell, A. (1973). You can't play 20 questions with nature and win. In

W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York, NY: Academic Press.

Newell, A. (1981). Reasoning, problem solving and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), *Attention and Performance VIII* (pp. 693–718). Hillsdale, NJ: Erlbaum.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language, 28,* 78–91. doi:10.1016/0749-596X(89)90029-6

Newstead, S. E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language, 34,* 644–664. doi:10.1006/jmla.1995.1029

Newstead, S. E. (2003). Can natural language semantics explain syllogistic reasoning? *Cognition, 90,* 193–199. doi:10.1016/S0010-0277(03)00117-3

Newstead, S. E., & Griggs, R. A. (1983). Drawing inferences from quantified statements: A study of the square of opposition. *Journal of Verbal Learning and Verbal Behavior, 22,* 535–546. doi:10.1016/S0022-5371(83)90328-6

Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory & Cognition, 27,* 344–354. doi:10.3758/BF03211418

Newstead, S. E., Pollard, P., Evans, J. St. B. T., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition, 45,* 257–284. doi:10.1016/0010-0277(92)90019-E

Oakhill, J., & Garnham, A. (1993). On theories of belief bias in syllogistic reasoning. *Cognition, 46,* 87–92. doi:10.1016/0010-0277(93)90023-O

Oakhill, J. V., & Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology: Human Experimental Psychology, 37A,* 553–569.

Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition, 31,* 117–140. doi:10.1016/0010-0277(89)90020-6

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381–391. doi:10.1037/0033-295X.103.2.381

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5,* 349–357. doi:10.1016/S1364-6613(00)01699-5

Oaksford, M., & Chater, N. (2010). Conditional inference and constraint satisfaction: Reconciling mental models and the probabilistic approach. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 309–333). Oxford, England: Oxford University Press.

Oberauer, K., Hörnig, R., Weidenfeld, A., & Wilhelm, O. (2005). Effects of directionality in deductive reasoning: II. Premise integration and conclusion evaluation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 58A,* 1225–1247. doi:10.1080/02724980443000566

Oberauer, K., & Wilhelm, O. (2000). Effects of directionality in deductive reasoning: I. The comprehension of single relational premises. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1702–1712. doi:10.1037/0278-7393.26.6.1702

Partee, B. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry, 6,* 203–300.

Partee, B. H. (1979). Semantics—Mathematics or psychology? In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 1–14). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-642-67458-7_1

Pascal, B. (1966). *Pensées* (A. J. Krailsheimer, Trans.). Harmondsworth, England: Penguin. (Original work published 1670)

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

Peters, S., & Westerståhl, D. (2006). *Quantifiers in language and logic.* Oxford, England: Oxford University Press.

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. (2007). The psychological reality of classical quantifier entailment properties. *Journal of Semantics, 24,* 331–343. doi:10.1093/jos/ffm012

Politzer, G., Van der Henst, J. B., Luche, C. D., & Noveck, I. A. (2006). The interpretation of classically quantified sentences: A set-theoretic approach. *Cognitive Science, 30,* 691–723. doi:10.1207/s15516709cog0000_75

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102,* 533–566. doi:10.1037/0033-295X.102.3.533

Ragni, M. (2003). An arrangement calculus, its complexity and algorithmic properties. *Lecture Notes in Computer Science, 2821,* 580–590. doi:10.1007/978-3-540-39451-8_42

R Development Core Team. (2011). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Revlin, R., Leirer, V., Yopp, H., & Yopp, R. (1980). The belief-bias effect in formal reasoning: The influence of knowledge on logic. *Memory & Cognition, 8,* 584–592. doi:10.3758/BF03213778

Revlis, R. (1975). Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior, 14,* 180–195. doi:10.1016/S0022-5371(75)80064-8

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

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.

*Roberts, M. J., Newstead, S. E., & Griggs, R. A. (2001). Quantifier interpretation and syllogistic reasoning. *Thinking & Reasoning, 7,* 173–204. doi:10.1080/13546780143000008

Sells, S. B. (1936). The atmosphere effect: An experimental study of reasoning. *Archives of Psychology, 29*(Whole No. 200).

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

Shin, S.-J. (1992). A semantic analysis of inference involving Venn diagrams. In N. H. Narayanan (Ed.), *AAAI spring symposium on reasoning with diagrammatic representations* (pp. 85–90). Stanford, CA: Stanford University.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22. doi:10.1037/0033-2909.119.1.3

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Stenning, K., & Cox, R. (2006). Reconnecting interpretation to reasoning through individual differences. *Quarterly Journal of Experimental Psychology, 59,* 1454–1483. doi:10.1080/17470210500198759

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science.* Cambridge, MA: MIT Press.

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology, 34,* 109–159. doi:10.1006/cogp.1997.0665

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

Strawson, P. F. (1952). *Introduction to logical theory.* New York, NY: Wiley.

Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (2011). Acquisition of generic

noun phrases in Chinese: Learning about lions without an "-s." *Journal of Child Language, 39,* 130–161. doi:10.1017/S0305000910000735
Tennant, N. (1986). The withering away of formal semantics. *Mind and Language, 1,* 302–318. doi:10.1111/j.1468-0017.1986.tb00328.x
Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. I. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review, 10,* 184–189. doi:10.3758/BF03196483
Tsujimoto, S., Genovesio, A., & Wise, S. P. (2011). Frontal pole cortex: Encoding ends at the end of the endbrain. *Trends in Cognitive Sciences, 15,* 169–176. doi:10.1016/j.tics.2011.02.001
Van der Henst, J.-B., Yang, Y., & Johnson-Laird, P. N. (2002). Strategies in sentential reasoning. *Cognitive Science, 26,* 425–468. doi:10.1207/s15516709cog2604_2
Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning, 11,* 239–278. doi:10.1080/13546780442000178
Viechtbauer, W. (2010). Conducting meta-analyses in R with the *metafor* package. *Journal of Statistical Software, 36,* 1–48.
Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.
Wetherick, N. E., & Gilhooly, K. J. (1995). "Atmosphere," matching, and logic in syllogistic reasoning. *Current Psychology, 14,* 169–178. doi:10.1007/BF02686906
Wilkins, M. C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology, 16*(Whole No. 102).
Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal reasoning. *Journal of Experimental Psychology, 18,* 451–460. doi:10.1037/h0060520
Yang, Y., & Johnson-Laird, P. N. (2000a). How to eliminate illusions in quantified reasoning, *Memory & Cognition, 28,* 1050–1059. doi:10.3758/BF03209353
Yang, Y., & Johnson-Laird, P. N. (2000b). Illusory inferences in quantified reasoning: How to make the impossible seem possible, and vice versa. *Memory & Cognition, 28,* 452–465. doi:10.3758/BF03198560
Zielinski, T. A., Goodwin, G. P., & Halford, G. S. (2010). Complexity of categorical syllogisms: A comparison of two metrics. *European Journal of Cognitive Psychology, 22,* 391–421. doi:10.1080/09541440902830509

Received January 6, 2011
Revision received October 20, 2011
Accepted November 29, 2011 ■