AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

2022, Vol. 129, No. 2, 289–312
https://doi.org/10.1037/rev0000240

Reasoning About Properties: A Computational Theory

Sangeet Khemlani[1] and P. N. Johnson-Laird[2, 3]
[1] Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, Washington, DC, United States
[2] Department of Psychology, Princeton University
[3] Department of Psychology, New York University

We present a theory of how people reason about properties. Such inferences have been studied since Aristotle's invention of Western logic. But, no previous psychological theory gives an adequate account of them, and most theories do not go beyond syllogistic inferences, such as: *All the bankers are architects; Some of the chefs are bankers; What follows?* The present theory postulates that such assertions establish relations between properties, which mental models represent in corresponding relations between sets of entities. The theory combines the construction of models with innovative heuristics that scan them to draw conclusions. It explains the processes that can generate a conclusion from premises, decide if a given conclusion is necessary or possible, assess its probability, and evaluate the consistency of a set of assertions. A computer program implementing the theory embodies an intuitive system 1 and a deliberative system 2, and it copes with quantifiers such as *more than half the architects*. It fit data from over 200 different sorts of inference, including those about the properties of individuals, the properties of a set of individuals, and the properties of several such sets in syllogisms. Another innovation is that the program accounts for differences in reasoning from one individual to another, and from one group of individuals to another: Some tend to reason intuitively but some go beyond intuitions to search for alternative models. The theory extends to inferences about disjunctions of properties, about relations rather than properties, and about the properties of properties.

Keywords: mental models, properties, quantifiers, sets, syllogisms

> Every deduction is through three terms; and the one type is capable of proving that *A* belongs to *C* because it belongs to *B* and that to *C*, while the other is negative, having one proposition to the effect that one thing belongs to another and the other to the effect that something does not belong.
>
> —Aristotle, *Prior Analytics*, Book I, 19

Properties are at the core of human reasoning. Aristotle treated them as such in his *Prior Analytics*, where a property in modern terminology is a predicate that takes one argument. The medieval Schoolmen followed Aristotle, and the earliest psychologist to study reasoning did so too (Störring, 1908). Even Frege (1964/1893), who formulated the first comprehensive logic, got into trouble over properties. But no one has formulated a satisfactory theory of how individuals make inferences about them. Indeed, this problem is one of the great mysteries in cognitive science.

Consider this exemplary inference about a certain group of professors:

1. All of the professors are experts in psychology.
   One of the professors is Joan.
   Therefore, Joan is expert in psychology.

It refers to two properties—being a professor and being expert in psychology. It depends on two quantifiers: *all of the professors* and *one of the professors*. Yet, it is so simple and obvious that readers may wonder why a theory explaining it isn't equally simple and obvious. One reason is that it is easy to overlook inferences that appear to be of the same sort, but are not, such as:

2. All of the professors are experts in different disciplines.
   One of the professors is Joan.
   Therefore, Joan is expert in different disciplines.

The first inference is *valid*, that is, the conclusion is true in every case in which the premises are true (Jeffrey, 1981, p. 1), whereas the second inference is invalid, because each professor's expertise may be in a single discipline. Another impediment to a theory of these inferences is that introspection does not tell you much about how you reason. As a consequence, cognitive scientists have proposed at least a dozen theories of *syllogisms*, which Aristotle formulated, and

This article was published Online First September 23, 2021.
Sangeet Khemlani https://orcid.org/0000-0002-1158-553X
P. N. Johnson-Laird https://orcid.org/0000-0003-4461-1324
This research was supported by a funding from the Office of Naval Research to Sangeet Khemlani, and by National Science Foundation Grant No. SES 0844851 to P. N. Johnson-Laird to study deductive and probabilistic reasoning. We thank Gordon Briggs, Monica Bucciarelli, Ruth Byrne, Hannah Dames, Hua Gao, Sam Glucksberg, Adele Goldberg, Geoff Goodwin, Hillary Harner, Tony Harrison, Catrinel Haught, Laura Hiatt, Zach Horne, Laura Kelly, Joanna Korman, Markus Knauff, Niklas Kunze, Max Lotstein, Gorka Navarrete, Thad Polk, Marco Ragni, Nicolas Riesterer, and Greg Trafton, for their help and criticisms. Some of the experiments and data analysis in the article come from work presented at the Annual Meetings of the Cognitive Science Society in 2014 and 2016, and simulations on immediate inferences were described in Khemlani, Lotstein, and Johnson-Laird (2015).
Raw data, analysis scripts, figure scripts, computational modeling code, and simulated data sets, are available at https://osf.io/xtrp6/
Correspondence concerning this article should be addressed to Sangeet Khemlani, Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, 4555 Overlook Dr. SW, Washington, DC 20375, United States. Email: sunny.khemlani@nrl.navy.mil

which are inferences from two quantified premises about properties, for example:

3. Some of the professors are programmers.
   All of the programmers are psychologists.
   Therefore, some of the professors are psychologists.

This syllogism is of a sort that even 11-year olds can make (Johnson-Laird et al., 1986), but other sorts defeat almost everyone.

Inferences depending on quantifiers underlie logic, mathematics, science, and many everyday matters from the laws of the land to the rules of games. Our goal in the present article is to present a new theory of how *naive* individuals—those who have not studied logic or its cognate disciplines—interpret quantified properties to represent their meanings, envisage the situations to which these meanings refer, and use these models of situations to reason. The theory explains the three sorts of inference exemplified earlier (1 to 3), and many other sorts of inference too.

Any plausible theory of reasoning should have three immediate goals. It needs to explain:

1. how individuals can make valid inferences.
2. how inferences differ in difficulty.
3. how individuals differ in ability.

It should account both for the inferences that individuals draw (at the computational level) and for the mental processes underlying them (at the algorithmic level). A convincing theory should also—to use Chomsky (1965) terminology—be explanatorily adequate. That is, it should explain how individuals acquire the ability to reason. And it should extend to other sorts of reasoning. In terms of these goals, the theory that we present gives a more thorough explanation than its precursors.

The article has four parts. First, it summarizes previous psychological theories of reasoning about properties. Second, it describes a new theory based on mental models, which aims to overcome the defects of its predecessors. Third, it assesses corroboratory evidence in the theory's fit to data about various sorts of inference. Fourth, it discusses the theory's possible shortcomings and its extensions to other sorts of reasoning.

Psychological Theories of Quantified Properties

Psychologists began their study of reasoning with syllogisms (Störring, 1908), a set of inferences from 64 distinct pairs of premises that concern properties. Syllogisms are tractable, and researchers have studied them using broad sets of tasks, contents, and measures (see Khemlani & Johnson-Laird, 2012, for a meta-analysis). In early studies, participants had to evaluate given conclusions or select a conclusion from a set of options. The results yielded only accounts of causes of difficulty. In modern parlance, researchers proposed *heuristics* governing the initial formulation of conclusions. One such heuristic is the "atmosphere" created by the *mood* of the premises (e.g., Woodworth & Sells, 1935; Appendix A describes the concept of mood). A succinct formulation of the heuristic due to Begg and Denny (1969) echoes Scholastic principles: When at least one premise is negative, the conclusion is negative, otherwise it is affirmative; and when at least one premise contains "some," the conclusion does too, otherwise it is universal ("all" or "no"). The atmosphere effect is stronger for valid than for invalid conclusions (see, e.g., Madruga, 1984; Polk & Newell, 1988). Insofar as it exists, it does not occur when "only" is used in a quantifier, such as: "only the architects" (Johnson-Laird & Byrne, 1989).

Another potential cause of difficulty is the invalid conversion of premises, such as inferring from *All A are B* that *All B are A* (Chapman & Chapman, 1959). Such illicit conversions occur in immediate inferences from a premise to its converse (Wilkins, 1929), and they may reflect verbal processes (Revlin et al., 1980; Revlis, 1975). But, a feasible mechanism instead is that *All A are B* has a representation of the two sets as co-extensive. Instructions can block invalid conversions (Dickstein, 1975), and so too can premises conveying that *A* is a proper subset of *B*, such as, "All dogs are pets," because everyone knows that not all pets are dogs (Ceraso & Provitera, 1971).

In the 1970s, experiments for the first time called for participants to draw their own conclusions, and presented them with all 64 distinct pairs of syllogistic premise (e.g., Dickstein, 1978a, 1978b; Johnson-Laird, 1975). Participants were able to reason, as opposed to guess, and a marked effect of *figure* occurred, that is, the arrangement of the three terms in a syllogism. The most frequent sorts of conclusion, whether valid or invalid, over the four sorts of syllogistic figure were as follows, where *A*, *B*, and *C*, stand for terms such as professors, programmers, and psychologists (as in three above):

Figure 1.		*A–B*
		B–C
	∴	*A–C*
Figure 2.		*B–A*
		C–B
	∴	*C–A* more often than *A–C*
Figure 3.		*A–B*
		C–B
	∴	*A–C* with only a slight bias over *C–A*
Figure 4.		*B–A*
		B–C
	∴	*A–C* with only a slight bias over *C–A*

In these examples, *A* and *C* are end terms, which occur in the conclusions, and *B* is the middle term, which occurs in both premises, but not normally in the conclusion. Initial theories attributed figural effects to an inherent bias in semantic processing (Johnson-Laird & Steedman, 1978, p. 77), to the grammatical tendency to maintain a quantifier that was the subject of a premise as the subject of the conclusion (Chater & Oaksford, 1999), or to the "first in, first out" properties of working memory (Johnson-Laird & Bara, 1984a). Oberauer et al. showed that the effects are a consequence of an inherent directionality in the meaning of premises (Oberauer et al., 2005; Oberauer & Wilhelm, 2000). Hence, to interpret *All the A are B*, it is natural to represent the quantifier, "All the *A*" and then to add *B* to its interpretation.

After the results for all 64 pairs of syllogistic premises became available, psychologists started to speculate about the mental processes underlying reasoning. Theories rose fast. And we can divide them into seven broad categories:

- Those based on the first-order predicate calculus (e.g., Braine, 1998; Rips, 1994, 2002).

- Those based on Euler circles or the similar Gergonne diagrams in which circles represent sets, for example, for *Some A are B*, the circle for *A* overlaps in part with the circle for *B* (Erickson, 1974; Guyote & Sternberg, 1981; and see Politzer et al., 2006, for the history of such diagrams).

- Those based on Venn diagrams in which three overlapping circles stand for the terms in a syllogism, and are shaded or annotated to represent the two premises (e.g., Newell, 1981, used symbols to denote each diagram).

- Those postulating distinct processes for different individuals, either formal rules or Euler circles (Ford, 1995; Stenning & Yule, 1997).

- Those based on probabilistic validity and heuristics (Chater & Oaksford, 1999; Oaksford & Chater, 2007) or probabilistic sampling (Hattori, 2016; Tessler & Goodman, 2014).

- Those based on rules governing monotonically increasing and decreasing terms (Geurts, 2003; Politzer, 2007; see Appendix A for details).

- Those based on mental models (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991), or a verbal analogy of models but in which counterexamples play no part (Polk & Newell, 1995).

None of these accounts is adequate. Those based on first-order predicate calculus or diagrams cannot account for inferences using quantifiers outside their scope, such as:

4. More than half the artists are Cubists.
   More than half the artists are painters.
   Therefore, at least one Cubist is a painter.

Those based on probability logic cannot explain inferences of this sort:

5. Some of the artists are beekeepers.
   Some of the artists are not beekeepers.
   Could both of these assertions be true at the same time?

Probability logic makes no reference to truth (Adams, 1998), and it does not explain how to assess the consistency of the quantified assertions in (5). A meta-analysis of syllogistic reasoning showed that none of the theories, including the original account of mental models, is satisfactory (Khemlani & Johnson-Laird, 2012). We therefore turn to a new theory.

The New Model Theory of Inferences About Properties

Craik (1943) proposed that people make decisions by running simulations on small-scale models of the world. Models don't need to have any structural relation to what they represent. What matters is that given an input, model and reality both produce the same output, and Craik cited as an example Kelvin's tidal predictor in which a system of cogs models the tides, but has no structural resemblance to them. As for reasoning, Craik suggested that it depends on verbal rules. A more recent theory pushes Craik's idea to its natural destination: Mental models underlie reasoning too, and they are iconic in that their structure corresponds to the structure of what they represent—an idea first proposed for syllogisms (Johnson-Laird, 1975, 1983). Iconicity is an aspect of Peirce's (1931–1958, Vol. 4) theory of symbols and it has two great advantages for mental models. It allows them to underlie visual images, though models can also be abstract and impossible to visualize. And it allows them to yield inferences that emerge from their structure, such as:

6. The black ball is directly beyond the cue ball.
   The green ball is on the right of the cue ball, and there is a red ball between them.
   So, if I move so that the red ball is between me and the black ball, then the cue ball is to the left.

To discover axioms to deliver this inference calls for envisaging the spatial arrangement of the balls and the appropriate kinematic move, and so the model theory obviates their discovery and instead makes inferences from such models.

The new theory makes further innovations, and it focuses on quantified properties. We will show how it extends to reasoning about quantified relations, too. The theory embodies the following principles:

- *Natural language elicits meanings in the mind (intensional representations).* The theory posits that people parse the syntax of a sentence and use the meanings of its words to compose a representation of the meaning of the assertion, that is, its *intension*. This idea is an innovation in syllogistic theories, and it allows that a model can be modified and then compared with intensions to ensure that it remains consistent with a premise's meaning.

- *Meanings yield models of situations (extensional representations).* Intensional representations are used to construct or modify mental models. Hence, an assertion can initiate a new model—perhaps an alternative to an existing one, add information to a model, or be evaluated as true or false in a model. The construction of models takes into account relevant knowledge, which can modulate the literal meaning of an assertion (Johnson-Laird & Byrne, 2002).

- *Each model represents a possibility.* Possibilities are the foundation of human reasoning (Johnson-Laird & Ragni, 2019), and a mental model is one of a small number of alternatives that humans can envisage in any situation. It represents what is common to an indefinite number of different realizations, which are sometimes known as "possible worlds." The former are in the head, not the latter, which are too vast to be contained there (Partee, 1979). Such mental models are tractable and they underlie probabilities too.

- *Each model represents a possibility that holds in default of knowledge to the contrary.* Everyday reasoning is defeasible (or "nonmonotonic" as in artificial intelligence, e.g., McDermott & Doyle, 1980) in that it can withdraw valid conclusions in the face of inconsistent facts (Johnson-Laird, 2006; Khemlani et al., 2018). But, in this case, unlike typical nonmonotonic logics, reasoners search for a resolution of the discrepancy (Khemlani & Johnson-Laird, 2011).

- *Models are iconic insofar as possible.* The model of a set of entities is iconic in that it contains a set of tokens representing them. But, models can also include abstract symbols, such as a symbol for negation, which is linked to the appropriate semantics. Models can be static, but they can also simulate the unfolding of a temporal sequence of events (see Khemlani et al., 2013).

- *Models yield the modal status of conclusions.* A conclusion that holds in all the models of the premises follows of necessity (an alethic modality), a conclusion that holds in at least some of these models is possible, a conclusion that holds in a proportion of equipossible models has a corresponding probability, and a conclusion that holds in none of the models of the premises is impossible given the premises.

- *There are two systems for reasoning: system 1 is intuitive and system 2 is deliberative.* Intuition has no access to working memory for intermediate results, and so it constructs a single *mental* model at a time from an intensional representation, updates it according to further premises, and scans it using heuristics in order to formulate a conclusion or to verify a given conclusion. Mental models represent only what is true according to the premises. The intuitive system may therefore yield an invalid conclusion. Deliberations have access to working memory, and construct a *fully explicit* model from an intensional representation. This model can be an alternative to the intuitive one, and so it can also serve as a counterexample to an intuitive conclusion. This notion of a dual system for reasoning is due to the late Peter Wason, and goes back to an algorithmic theory of how individuals select potential evidence to test quantified hypotheses in his selection task (Johnson-Laird & Wason, 1970), and psychologists have explored similar ideas for other cognitive processes since the 19th C. We now outline how this new theory works.

Quantified Properties: Meanings and Inferences

As its foundation in possibilities suggests, the model theory treats the meanings of quantified properties as referring to possible entities. For example, the assertion:

7. All shoplifters are prosecuted.

means that there are possible individuals who are shoplifters and prosecuted, but no possible individuals who are shoplifters and not prosecuted. Other sorts of individual are also possible, such as those who are not shoplifters and prosecuted, but the denial of (7) also refers to their possibility—it is a presupposition of both (7) and its denial. An assertion can also refer to individuals in a known set—one in normal discourse that is already represented in a model, as in:

8. All the shoplifters will be prosecuted.

The model theory's treatment of quantified properties is accordingly as relations between sets of individuals (see Table A1 in Appendix A). But, as we show later, differences exist between models and the sets of set theory.

As Aristotle remarked, two sorts of inference are the foundation of all reasoning about properties (see the epigraph to this article). The first sort of inference is exemplified in a known set of individuals who possess a property:

9. All of the architects are bankers.
   Pat is one of the architects.
   Therefore, Pat is a banker.

The terms in these premises "architects" and "bankers" refer to properties, and an iconic mental model represents the corresponding sets. The following diagram depicts such a model, which intuition constructs, and which for convenience uses words instead of representations of people:

```
architect    banker
architect    banker
             banker
```

Each row in the diagram denotes a different individual, and so the first row denotes someone who is both an architect and a banker, the second row denotes another such individual, and the third row denotes a banker leaving open whether or not this individual is an architect. Of course, there may be individuals who are neither architects nor bankers, but the mental model of (9) does not need to represent them. The second premise can be used to update the model by adding a token representing Pat, who can be either of the two architects in the existing model, for example:

```
architect    banker    Pat
architect    banker
             banker
```

It follows from the model that Pat is a banker, and no alternative model of the premises refutes this conclusion. The example illustrates how, given iconic models, monotonically decreasing terms require no special rules for reasoning (pace Geurts, 2003; see Appendix A). When an individual is a member of a set, the individual is a member of all the sets that include that set.

The second foundational inference concerns *not* possessing a property:

10. All of the architects are bankers.
    Viv is not a banker.

The first premise has the same sort of mental model as before. The addition of the information in the second premise yields a model, such as:

```
architect     banker
architect     banker
              banker
            ¬ banker     Viv
```

where "¬" is a symbol for negation in models. It follows from this mental model that Viv is not one of the architects in the known set, though she could be a member of another set of architects. Once again, no alternative model refutes this conclusion. The example illustrates how, given iconic models, monotonically increasing terms require no special rules for reasoning (pace Geurts, 2003): when an individual is not a member of a set that includes another set, the individual is not a member of this other set, either.

The next sort of inferences concerns immediate conclusions drawn from a single premise. The intuitive system constructs an iconic model for *All the architects are bankers*, such as:

```
architect   banker
architect   banker
            banker
```

It supports the immediate conclusion, which no alternative model refutes:

11. Some of the bankers are architects.

Suppose that the following assertion is true:

12. More than half of the architects are bankers.

Does it follow that more than half of the bankers are architects? The intuitive system is parsimonious, and so it is likely to represent the premise in the following mental model:

```
architect   banker
architect   banker
architect
```

In this case, more than half the bankers are architects—indeed, all of them are. So, the intuitive answer to the question is: Yes. The deliberative system, however, can search for an alternative model consistent with the meaning of the premise, for example:

```
architect   banker
architect   banker
architect
            banker
            banker
```

This model is counterexample to the intuitive conclusion, and so it does not follow that more than half the bankers are architects. Individuals, however, often make this fallacious inference (Power, 1984).

Given one premise asserting that a property holds for an entire set, such as, *All the architects are bankers*, a second premise can yield a valid inference. This other premise can be based on classical quantifiers, such as *All* ___, which is in the first-order predicate calculus, or on a quantifier such as *most* __, or *more than half* __, which calls for the second-order predicate calculus (see Appendix A). For instance, the following inference is valid:

13. Most of the designers are architects.
    All the architects are bankers.
    Therefore, most of the designers are bankers.

But, there are constraints that models make obvious. With the universal second premise here, any quantifier in the first premise that limits the number or numerical proportion of individuals having its property can yield an invalid inference, for example:

14. No more than two of the designers are architects.
    All the architects are bankers.
    Therefore, no more than two of the designers are bankers.

Deliberation can yield this model of the premises that is a counterexample to the conclusion:

```
designer   architect   banker
designer   architect   banker
designer               banker
designer               banker
```

We have entered the territory of syllogisms, which illustrate nearly all the components of the mReasoner program, and so we use them to illustrate the program in the next section.

A Computer Implementation of the Model Theory

The theory in the previous section is implemented in the computer program, mReasoner, written in Common Lisp, and its source code is at https://osf.io/xtrp6/. The program parses quantified sentences about properties to yield intensional representations of their meanings. It uses an intuitive system 1 that builds models based on these intensions, and a set of heuristics to formulate a conclusion that follows from their models. And it uses a deliberative system 2 to evaluate these conclusions: it searches for alternative conclusions, and if it finds one, it can weaken the conclusion so that it holds in the current set of models of the premises. Both systems can scan a model to carry out several different reasoning tasks—to generate a conclusion, to assess whether a conclusion is necessary, possible, or impossible, and to evaluate the consistency of a set of quantified properties. Figure 1 is a diagram of the main components of the two systems.

The intuitive system 1 has no access to working memory for intermediate results, and so it has only the power of a finite-state automaton (Hopcroft & Ullman, 1979). The deliberative system 2, in contrast, has access to a limited working memory in which it can store alternative models, at least until they exceed its processing capacity. We now illustrate how each component of the theory works for the testbed of syllogistic inferences in which reasoners draw their own conclusions from two quantified premises.

The Parser and the Representation of Meanings

To illustrate all the program's processes, we use the following exemplary pair of premises in syllogistic Figure 1:

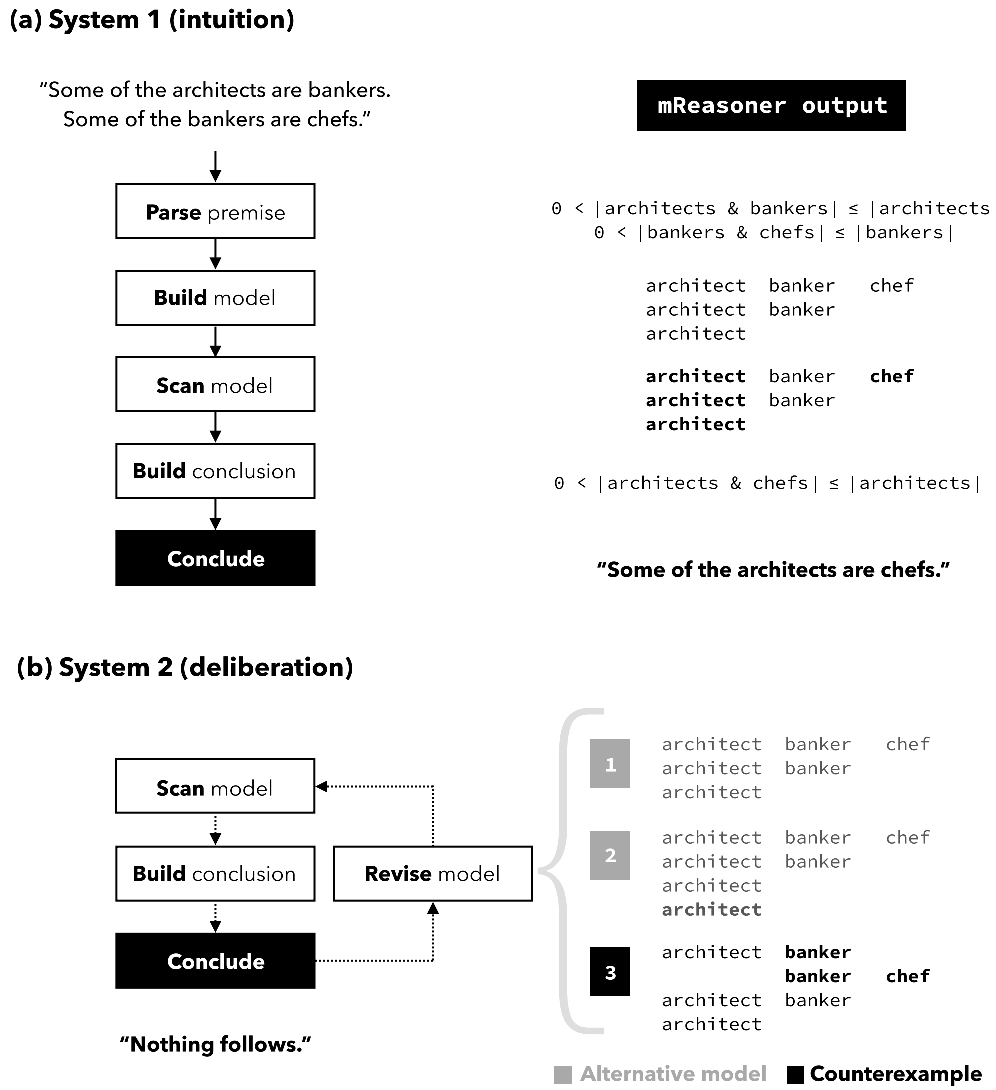
15. Some of the architects are bankers.
    Some of the bankers are chefs.

The parser constructs representations of their meanings—their intensions. Table 1 presents the intensional representations of the meanings of assertions about quantified properties. Their semantics is computable, plausible, and copes with quantifiers outside the first-order calculus. The intentional representations serve as blueprints for constructing and interpreting mental models.

As the parser analyses each sentence it constructs a representation of its intension. For the first premise in the preceding example, *Some of the architects are bankers*, it constructs the intension using representations akin to those in Table 1 in its lexicon of determiners, such as "all" and "some." It builds up the intension of the sentence using intensional rules that parallel the grammatical rules it uses in its parse of the sentence. The intension of the first premise is:

```
(:first-argument       architect
:second-argument       banker
:cardinality           ≥ 1
:relation              intersection
:boundary              | architect & banker | ≤ |
                         architect |
:polarity              t)
```

Figure 1
Schematic Diagrams of the Two Systems in the mReasoner Program



Note. Each process is depicted as a box, and an example of its output is on the right hand side of the diagram. Panel (a) shows system 1: It parses a premise to construct an intensional representation, which is then used to constrain the stochastic system that builds models. The result is a mental model, which is scanned using heuristics to yield another intension, which is transformed into a conclusion. Panel (b) shows system 2: It can alter the initial model from system 1 to try to falsify the initial conclusion. If it succeeds, system 2 can weaken the initial conclusion and deliberate about it again; when the conclusion cannot be weakened any further (as shown in the example), the system responds that no valid conclusion (NVC) follows.

The subject and object of the sentence yield the first and second arguments, and the values of the remaining variables are plugged in from the lexical entry for the determiner, "some." *Cardinality* specifies that the number of architects is one or more, *relation* specifies an intersection between the two sets, *boundary* that the number of tokens of artists and bankers is less than or equal to the number of architects, and *polarity* that the sentence is positive. The intensions of the two premises in (15) are summarized in Figure 1a.

A model in which two out of three architects are bankers can represent that some of the architects are bankers, but also many other assertions, such as: exactly two of the architects are bankers. The previous model theory used conventions to deal with such limitations, but the present theory embodies a general solution: the number of individuals is arbitrary and can be changed, that is, individuals can be added or removed, but the model must satisfy the sentence's intension. So, a model is a possible extension given the

Table 1

Quantified Assertions, Their Set-Theoretic Intensions, and Their Intensional Representations in the Computational Implementation of the Model Theory of Reasoning

	Intensional representations		
Quantified assertion	Set-theoretic semantics	Constraint on number of tokens in the model that are *A*:	Constraint on number of tokens that are both *A* and *B*:
All *A* are *B*.	*A* are included in *B*.	>1	$\mid A \,\&\, B \mid = \mid A \mid$
Some *A* are *B*.	Intersection of *A* and *B* is not empty.	≥1	$0 < \mid A \,\&\, B \mid \leq \mid A \mid$
No *A* is a *B*.	Intersection of *A* and *B* is empty.	≥1	$\mid A \,\&\, B \mid = 0$
Some *A* are not *B*.	Set of *A* that are not *B* is not empty.	≥1	$0 < \mid A \,\&\, \neg B \mid \leq \mid A \mid$
At least three *A* are *B*.	Cardinality of the intersection ≥ 3.	≥1	$\mid A \,\&\, B \mid \geq 3$
Exactly three *A* are *B*.	Cardinality of the intersection = 3.	≥1	$\mid A \,\&\, B \mid = 3$
Neither *A* is a *B*.	Cardinality of *A* is two, and intersection of *A* and *B* is empty.	2	$\mid A \,\&\, \neg B \mid = 2$
Most *A* are *B*.	Cardinality of intersection > ½ of cardinality of *A*s.	≥2	$½ \mid A \mid < \mid A \,\&\, B \mid < \mid A \mid$
More than half the *A* are *B*.	Cardinality of intersection > ½ of cardinality of *A*s.	≥2	$½ \mid A \mid < \mid A \,\&\, B \mid < \mid A \mid$
The *A* is a *B*.	There is one *A*, which is a *B*.	1	$\mid A \,\&\, B \mid = 1$

Note. $\mid A \mid$ denotes the number of tokens representing individuals with the property of *A* and $\mid A \,\&\, B \mid$ denotes the number of tokens representing individuals with the properties of *A* and *B*. See Appendix A for a further discussion of set-theory and the logic of properties.

premise—an example of the situation to which it refers. Any premise has many possible extensions. In fact, reasoners make systematic interpretations for the cardinalities of the subject and object in quantified assertions (Moxey & Sanford, 2000, p. 242).

Intuitive System 1: The Construction of Models

Reasoners do not always construct the same model from the same premises. Construction is stochastic, but biased to create small simple models rather than large complex ones. So, mReasoner builds models following the constraints of two parameters:

Parameter 1: Size. The `size` parameter controls the maximum number of tokens a model contains. A sample is drawn from a left-truncated Poisson distribution that `size` governs. It can be set to approximate any real number greater than 0. If, say, it is set to 3.2, then the system draws samples of natural numbers, for example, 2, 3, 4, and 5, that have an expected mean of 3.2. Once a sample is drawn (e.g., 4), the system builds a model of four individuals. We constrain `size` < 6.0 in order to yield small models.

Parameter 2: Atypicality. The `atypicality` parameter governs the model's contents. It sets the probability of constructing a model of a typical set of individuals for a premise (`atypicality` = 0 guarantees it) or one with the full set of possible individuals for a premise (`atypicality` = 1 guarantees it). These sets were established from previous experiments (e.g., Bucciarelli & Johnson-Laird, 1999).

The premise, *some of the architects are bankers*, has a typical model of the following sort:

```
architect   banker
architect   banker
architect
```

But, the complete set of possibilities includes bankers who are not architects. Our simulation studies tend to converge on values of `atypicality` of less than .5, because reasoners tend to build typical models. The consequence of the parameter settings is a model of both the premises of (15) shown in Figure 1a.

Intuitive System 1: Heuristics and the Formulation of Conclusions

Heuristics have played a role in syllogistic reasoning since Aristotle, and theories continue to postulate them (Chater & Oaksford, 1999; Hattori, 2016; cf. Khemlani, 2021). The model theory embodies their use in system 1 to explain how individuals formulate the mood and figure of conclusions. The precursors to these heuristics are the speculation of Johnson-Laird and Steedman (1978, p. 91 et seq.), the observations of Johnson-Laird and Bara (1984a, p. 47), the program for syllogisms in Johnson-Laird and Byrne (1991), and the theories of Stenning and Yule (1997) and Chater and Oaksford (1999). However, the present heuristics differ from all the preceding, ones including the atmosphere and the matching heuristics, which determine mood only (Wetherick & Gilhooly, 1995).

Two tacit heuristics scan models to formulate conclusions. One determines the mood of the conclusion. The mental model for a negative premise represents the negations of properties, and individuals acquire the heuristic that any necessary conclusion about these entities must be negative too, but otherwise it is affirmative. This heuristic occurs in other sorts of reasoning too (e.g., Evans et al., 1996). Likewise, individuals learn that intensions for a premise containing a *particular* quantifier, *Some of the X,* yield models in which a subset of *X* has the property in the premise's predicate, and so any necessary conclusion must be particular, but otherwise it is universal, as in *All of the X* or *None of the X*. The heuristic applies to other determiners, such as *a few*, and *more than half of*.

The figural heuristic determines the order of the two terms in conclusions. They follow the order in which end terms are introduced into a model: if a quantified end term is represented in a model before the other end term, then the conclusion should be in that order too, but when one end term is just as likely as the other to be represented first, both orders of the end terms should occur with roughly equal frequencies. The heuristic is therefore implemented in this way:

> Premises in figure A–B, B–C yield the conclusion: A–C.
> Premises in figure B–A, C–B yield the conclusion: C–A.
> For the other two figures, mReasoner makes a random choice between: A–C or C–A.

The mood heuristic is simple: It depends only on the polarity of the premises and whether a particular quantifier occurs in them. The figural heuristic is not as simple: It depends on the relation between terms in the subject and predicate of each premise. Hence, the heuristics yield the following rank order in the salience of the different sorts of conclusion, that is, the readiness with which they should come to mind when reasoners formulate conclusions:

O: Some _ are not _	(Negative and particular)
E: No _ is a _	(Negative and universal)
I: Some _ are _	(Affirmative and particular)
A: All _ are _	(Affirmative and universal)

The consequence is that given the model of the exemplary premises, the heuristics yield the conclusion (in Figure 1a):

16. Some of the architects are chefs.

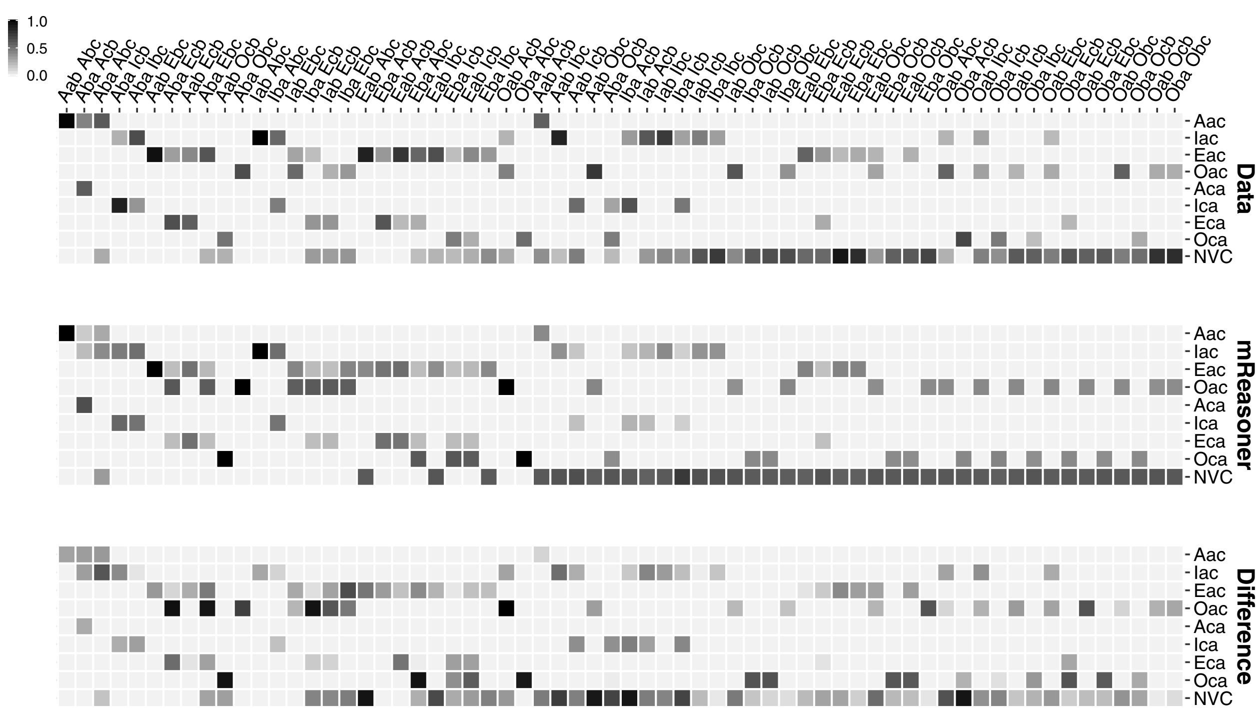
Most individuals make this inference (see Figure 2 below for the conclusions that participants drew for each of the 64 syllogistic premises). Table 2 illustrates the effects of the figural and mood heuristics for four sorts of syllogisms.

Unlike earlier heuristics, which use premises to derive constraints on conclusions, the present heuristics scan models to yield both the mood of the conclusion and the order of its terms. The reason for this procedure is that the heuristics are acquired from experience in building models that survive a search for counterexamples. The process could be a form of sampling from a distribution of possible conclusions (Phillips et al., 2019). The mood heuristic is similar to probability heuristics (Chater & Oaksford, 1999) except that, unlike the latter, it implies that O conclusions should be frequent—a prediction that the data corroborate (see Figure 2). Likewise, the figural heuristic predicts the figure of conclusions, which previous heuristics do not (e.g., Chater & Oaksford, 1999; Stenning & Yule, 1997; Wetherick & Gilhooly, 1995).

In mReasoner, no free parameters govern the generation of conclusions. But, any intuitive conclusion must hold in the model that system 1 constructs (see Figure 1a). This constraint also separates the theory from previous accounts (e.g., Chater & Oaksford, 1999; Hattori, 2016). A corollary is that syllogisms for which intuitive conclusions are correct should be easy—participants should draw them first and most often, whereas those syllogisms for which deliberation is called for to reach a correct

Figure 2

The Proportions of Responses From for 64 Syllogisms in Six Studies in Khemlani and Johnson-Laird (2012)'s Meta-Analysis



Note. Proportions range from 0 (white) to 100 (black). Top panel: the experimental data; middle panel: mReasoner's best-fitting simulation ($r = .80$); bottom panel: the differences between them. Abbreviations are as in previous figures. Each of the 64 pairs of premises is in a column, and each of the nine possible responses is in a row. The leftmost 27 columns denote syllogisms with valid conclusions and the rightmost 37 columns denote syllogisms with "no valid conclusion (NVC)."

Table 2
The Results of the Mood and Figural Heuristics for Four Illustrative Syllogisms

Syllogism: First premise	Syllogism: Second premise	Polarity	Particularity	Resulting initial conclusion
All of the *A* are *B* (affirmative)	All of the *B* are *C* (affirmative)	affirmative	universal	All of the *A* are *C*.
Some of the *B* are *A* (particular)	All of the *C* are *B* (affirmative)	affirmative	particular	Some of the *C* are *A*.
No *A* is a *B* (negative)	All of the *C* are *B* (affirmative)	negative	universal	No *A* are *C*.
No *B* is a *A* (negative)	Some of the *B* are not *C* (negative)	negative	particular	Some of the *A* are not *C*.

Note. The mood heuristic yields a conclusion depending on the polarity of the premises (affirmative or negative) and the nature of its determiners (particular or universal). The figural heuristic yields the order of the terms in the conclusion from the order in which they are represented in a model.

conclusion should take longer and yield more errors. The conclusion in (16), for instance, is erroneous—it does not follow of necessity from its premises. To correct the error, reasoners need to engage system 2.

Deliberations in System 2

System 2 corroborates, modifies, or rejects an initial conclusion to reach a correct outcome in principle. Given an intuitive conclusion, it seeks a counterexample, that is, a model of the premises in which the conclusion is false. If it finds one, it can recursively attempt to frame an alternative conclusion, test it, and so on, or, on failing to find such a conclusion, it can respond that nothing follows from the premises. The system uses three ways to construct alternative models:

- the addition of new tokens of individuals to a model;
- the breaking up of the properties of an existing individual into two sorts of individual;
- the swapping of properties from one individual to another.

Table 3 illustrates the effects of each of these ways. They are based on how individuals worked with external models of premises made from cutout shapes (Bucciarelli & Johnson-Laird, 1999). The algorithm that implements the three operations executes them contemporaneously, because experiments have not yet established the order in which reasoners use them. Two probabilistic parameters govern them:

Parameter 3: Search. The `search` parameter determines the probability of a search for counterexamples, where 0 determines that no search occurs, 1.0 determines that it does occur, and intermediate values determine its probability. Hence, high values of `search` should match the performance of highly proficient reasoners.

Parameter 4: Revising conclusions. The `reconclude` parameter is a nested parameter, that is, given a search for counterexamples that succeeds, it determines the probability of trying a weaker conclusion. If `reconclude` = 0, system 2 reports that nothing follows after discovering a counterexample; if it equals 1, it revises the conclusion that was falsified by the counterexample, and it can then engage in a search for counterexamples to the new conclusion. Intermediate values of `reconclude` determine its probability. The weakest conclusions system 2 can draw are existential ones.

Given the initial conclusion in our example (16) *Some of the architects are chefs*, the second process in Table 3 (and in Figure 1) breaks the individual having all three properties in the initial model:

Table 3
The Three Ways of Searching for Alternative Models Used in System 2: A Syllogism, Its Heuristic Conclusion, and a Modified Model That Refutes It (With Changes Highlighted in Bold)

Description	Syllogism	Heuristic conclusion	Models: Initial model	Models: Modified model
Add an individual to a model.	All of the *B* are *A*. All of the *B* are *C*.	All of the *A* are *C*.	B A C B A C	B A C B A C **A**
Break an individual with multiple properties into two separate individuals.	Some of the *A* are *B*. Some of the *B* are *C*.	Some of the *A* are *C*.	A B C A B A	A **B** **B** **C** A B A
Move a property from one individual to another.	No *A* is a *B*. No *B* is a *C*.	No *A* is a *C*. No *C* is a *A*.	A ¬B B ¬C C	A ¬B **C** B ¬C

Note. The models show only the different sorts of individual, where "¬B" denotes not B. Table modified (with permission) from Khemlani and Johnson-Laird (2013).

```
architect   banker   chef
```

into two separate individuals:

```
architect   banker
            banker   chef
```

This result is the final model in the left-hand column of Table 1b, and it refutes the conclusion. No weaker conclusion follows from both the original model and this new model, and so nothing definite follows from both these premises.

The parameters govern the interaction between the intuitive and deliberative processes of mReasoner in carrying out various inferential tasks. The program can generate its own conclusions, and evaluate given conclusions. The intuitive system assesses that a conclusion is possible if it holds in its initial model. When it does not hold, the deliberative system may find an alternative model in which it does hold. Likewise, the program can assess whether a given conclusion is necessary, or even that it is probable depending on the proportion of models in which it holds (Johnson-Laird et al., 1999; Khemlani, Lotstein, & Johnson-Laird, 2015). It can assess the consistency of a set of assertions: if, and only if, it finds a model of them they are consistent—a task beyond the probability heuristics model (Chater & Oaksford, 1999; Chater, personal communication, 5-27-2010). In sum, mReasoner explains a variety of reasoning tasks, and its algorithm is parsimonious (cf. Cassimatis et al., 2008). But, it also predicts the difficulty within a task over different sorts of problem. The next section addresses the extent to which it succeeds.

Empirical Assessments of the Theory

We used mReasoner to simulate a broad swathe of data sets from studies of inferences about properties. We chose these data sets for two reasons: first, each of them is a complete benchmark for one sort of inference, such as syllogistic reasoning (cf. Khemlani & Johnson-Laird, 2012). An experiment that tests only 10 of 64 syllogistic premises is not a complete benchmark, because it is bound to be a biased sample. Second, where possible, data sets came from tasks in which participants formulated their own conclusions in contrast to tasks in which they evaluated given conclusions or selected conclusions from a set of options. These latter tasks can lead reasoners to conclusions that they would never draw for themselves. Immediate inferences are an exception: it is hard to prompt reasoners to generate an inference from a single premise, and so the available data sets are from evaluative tasks. We describe 11 principal simulations ranging from simple inferences to syllogisms. Table 4 at the end of this section summarizes each simulation.

Simulation 1: Inferences About the Properties of Individuals

A simple sort of reasoning about properties concerns whether or not an individual has a property, for example:

17. Rachel is a soldier.
    All of the soldiers are bakers.
    What, if anything, follows?

mReasoner constructs a model of the premises, such as:

```
Rachel   soldier   baker
         soldier   baker
         soldier   baker
```

The heuristics use the model to create an immediate conclusion: Rachel is a baker. No alternative model of the premises refutes it. Likewise, the following premises:

18. Bran is not a baker.
    All of the soldiers are bakers.

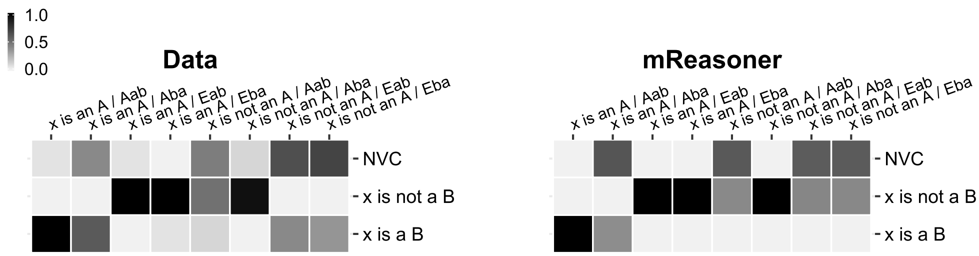
Table 4

The Experimental Data sets in mReasoner's 11 Simulations, With the Number of Inferences in a Data set; Its Task; the Values of the System's Four Parameters (Abbreviated) for the Best Fit; and Two Measures of the Goodness of the Fit (r and the root mean squared error [RMSE])

Data set	# of inferences	Task	Values of mReasoner's parameters				Goodness of fit	
			`size`	`atyp.`	`search`	`reconc.`	r	RMSE
Simulation 1: Inferences about individual properties								
Khemlani et al. (2014)	8	Generate conclusion	4.5	.20	.60	.20	.92	.15
Simulations 2–6: Immediate inferences								
Newstead and Griggs (1983, Experiment 1)	32	Evaluate necessity	4.0	.30	.50	—	.75	.21
Newstead and Griggs (1983, Experiment 2)	32	Evaluate necessity	4.0	.30	.40	—	.78	.23
Khemlani, Lotstein, et al. (2015, Experiment 1)	32	Evaluate possibility	3.8	.20	.60	—	.62	.14
Khemlani, Lotstein, et al. (2015, Experiment 2)	32	Evaluate possibility	3.8	.40	.40	—	.85	.13
Khemlani, Lotstein, et al. (2015, Experiment 3)	32	Evaluate consistency	3.5	.60	.70	—	.63	.16
Simulation 7: The meta-analysis of syllogistic reasoning								
Khemlani and Johnson-Laird (2012)	64	Generate conclusion	3.0	.40	.60	1.00	.80	.13
Simulation 8: Simulating individual reasoners								
Johnson-Laird and Steedman (1978, worst fit participant)	64	Generate conclusion	2.0	.00	.60	.00	.57	.35
Johnson-Laird and Steedman (1978, best fit participant)	64	Generate conclusion	2.0	.00	1.00	.60	.86	.07
Simulations 9–11: Simulating post hoc groups of reasoners								
Johnson-Laird and Steedman (1978, group 1)	64	Generate conclusion	2.0	.00	.60	.60	.78	.14
Johnson-Laird and Steedman (1978, group 2)	64	Generate conclusion	3.0	.60	.80	.60	.84	.14
Johnson-Laird and Steedman (1978, group 3)	64	Generate conclusion	4.5	.40	.80	.60	.92	.10

Figure 3

The Proportions of Responses in Khemlani et al.'s (2014) Experiment on Inferences About the Properties of Individuals (Left Panel) and the Predicted Responses From mReasoner's Simulation 1 (Right Panel)



Note. Each inference is labeled with its premise (e.g., *x is an A*) paired with its abbreviated quantified premise, such that Aab = *All of the a are b*, Eab = *None of the a is a b*, and its conclusion where NVC = *no valid conclusion* (see Appendix A for the origins of these abbreviations). The darker the color in each cell the greater is the proportion of corresponding conclusions. The cells for which the theory fails are the two on the bottom right of the data.

yield a model, such as:

```
Bran                  ¬ baker
          soldier      baker
          soldier      baker
          soldier      baker
```

It yields the heuristic conclusion: Bran is not a soldier, and no alternative model refutes it. Figure 3 shows the data and the best-fitting simulation of mReasoner for the eight different inferences about properties of individuals (from Khemlani et al., 2014, see Appendix B). The program had a close fit to the data ($r = .92$, RMSE = .15), and so the theory accounts well for simple inferences about properties.

The last two columns in Figure 3 of the plots of the data and the simulation reveal a noticeable failure of the theory. For premises of this sort:

19. Doran is not a soldier.
    None of the soldiers are bakers.

and of the sort:

20. Doran is not a soldier.
    None of the bakers are soldiers.

people often drew the conclusion:

21. Doran is a baker.

The inference is an error. However, mReasoner predicts a different error: Doran is not a baker—an inference participants never made. A tentative post hoc explanation is that individuals may have constructed a model that yields the error that occurred:

```
baker    ¬ soldier    Doran
baker    ¬ soldier
           soldier
           soldier
```

Instead, mReasoner constructs the following model:

```
¬ baker   ¬ soldier    Doran
  baker   ¬ soldier
  baker   ¬ soldier
            soldier
            soldier
```

It fits the heuristics, which call for a negative conclusion. One potential solution is that given two negative premises, the heuristics allow an affirmative conclusion. Such errors do occur in syllogistic reasoning too (see Table 4 in Bucciarelli & Johnson-Laird, 1999). Apart from these two disparities, mReasoner tends to predict the responses that reasoners make, and not to predict responses that they do not make.

Simulations 2–6: Immediate Inferences

An immediate inference from a quantified premise to a quantified conclusion is:

22. None of the architects is a designer.
    Therefore, none of the designers is an architect.

In two experiments, participants assessed whether such conclusions were necessary (Newstead & Griggs, 1983), and in three experiments, they assessed whether such conclusions were possible and whether the two assertions were consistent with each other (Khemlani, Lotstein, Trafton, & Johnson-Laird, 2015). These studies used classical quantifiers and those such as, *Most of the architects*, which call for the second-order predicate calculus (see Appendix B for the details of the simulations). Figure 4 shows the data and the best-fitting simulations for the five separate studies. The fits between each simulation and its data set are apparent in the degree to which the top row of cells matches the bottom row of cells for each experiment; and the fits were robust ($rs > .62$, RMSEs $<.23$). So, mReasoner can emulate immediate inferences about necessary and possible conclusions, including those that use quantifiers of the sort: *Most of the A*.

Simulation 7: Syllogistic Inferences

Syllogisms vary much more in difficulty than the inferences in the previous simulations. As mReasoner predicts, most reasoners make intuitive inferences using system 1, but a few go further and use system 2 to deliberate. Figure 2 presents the data from six studies in Khemlani and Johnson-Laird (2012) meta-analysis of the 64 sorts of syllogistic inference. Its top panel shows the data, its middle panel shows mReasoner's performance, and its bottom panel shows the differences between the two. A high correlation occurred between the simulation and the data (r = .80, RMSE = .13). As Figure 2 shows, mReasoner predicts the frequent inferences that reasoners make (black cells) and tends not to predict those they refrain from making (white cells). It also accounts for the figural effect, and, unlike recent computational accounts (Hattori, 2016; Tessler & Goodman, 2014), it can respond "no valid conclusion (NVC)." Figure 2 also shows that the theory predicts these responses slightly more often for invalid syllogisms than their occurrence in experiments. One reason may be again that some individuals consider an alternative model, but draw a conclusion that overlooks the intuitive model. This post hoc explanation implies that individuals may differ in systematic ways, and so the next simulation examines whether mReasoner can account for such differences.

Simulation 8: Syllogistic Performance of Individuals

Some people are better at syllogistic reasoning than others (Bucciarelli & Johnson-Laird, 1999; Johnson-Laird, 1983, p. 117 et seq.). Good reasoners appear to consider more alternatives than poor reasoners do (Galotti et al., 1986). Ability depends, in part, on the processing capacity of working memory to maintain these separate possibilities: It was the only factor out of several mediating reasoning that correlated with syllogistic performance (Bara et al., 1995). However, reasoners can compensate for its limitations by spending more time on the task (Baron et al., 1986). Yet, no previous theory gives a systematic account of how individuals differ in syllogistic ability.

We used mReasoner to simulate each of 20 participants' syllogistic inferences in an early experiment (Johnson-Laird & Steedman, 1978), because their individual data were available, and because they each reasoned from all 64 sorts of pairs of premise. They were more accurate than participants in any other study in the meta-analysis, and they responded "NVC" more often than in any other study (45% vs. 30%, Wilcoxon test, z = 6.19, p < .001). Appendix B summarizes the methodology. Table 5 provides the parameter settings and statistics on the fit of the best and worst simulations for each of the 20 participants. The mean correlation between the best fitting simulation and the data was .70, and it ranged from .57 to .86 over the participants, whereas the mean correlation between the worst fitting simulation and the data was .24. Individual reasoners varied considerably in the conclusions they drew; nevertheless, all but four of the best-fitting simulations achieved a correlation of .60 or higher. Hence, the simulation accounted for about 50% of their variance. The optimal parameter values obtained from the analysis provide some insight into the participants' reasoning. And, the data show no significant concordance between parameter settings (Kendall's W = .28, p = .33). In sum, mReasoner successfully simulates individuals' patterns of inference, but to make sense of these patterns calls for the discovery of separate groups of participants who reason in similar ways.

Simulations 9–11: Simulating Differences Among Separate groups of Reasoners

A hierarchical cluster analysis (see Appendix B) revealed that an optimal separation of the participants was into three main groups, and mReasoner's best-fitting simulations correlated with them with rs = .78, .83, and .91, respectively. Table 4 above reports the parameter settings of the three simulations and their goodness-of-fit metrics. Overall, a close fit occurred between the simulations and the data (rs > .98, RMSEs < .14).

The parameter settings revealed the characteristic reasoning of the three groups of participants. Those in the group of poorest reasoners built relatively small models: The optimal parameter setting of the size of their models (`size`) was 2.0: A model representing only two individuals, who were in typical models (`atypicality` = 0), and so these participants did not explore the problem space in full. They had only a slight tendency to search for alternative models (`search` = 0.4) and on finding one to weaken their intuitive conclusion (`reconclude` = .6). In short, they were intuitive reasoners. In contrast, those in the group of intermediate reasoners constructed medium-size models (`size` = 3.0), which tended to be atypical (`atypicality` = 0.6), but otherwise they were similar to the next group. This group contained the best reasoners, and they constructed large models (`size` = 4.5), which were sometimes atypical (`atypicality` = .4), and they tended to search for alternative models (`search` = 0.8) and to use them to weaken their intuitive conclusions (`reconclude` = .6). In short, they were deliberative reasoners. Hence, mReasoner discovered major differences in the way groups of individuals of different ability made syllogistic inferences.

Summary

Table 4 summarizes all 11 simulations. As it shows, the optimal values of parameters differ across the simulations. For instance, in some simulations the optimal value of `size` was greater than 4.0, but in others `size` was 2.0, yielding smaller models. These differences may reflect inherent differences among the experiments—in their quantifiers and inferences, in the ability of the participants, and so on. The results also corroborated a tenet of the model theory: Reasoning calls for only a small number of models of the premises. The computation of dozens of models should fare no better. Human reasoners are parsimonious.

General Discussion

Inferences about properties are a large part of everyday reasoning, and our simulations imply that the model theory gives a plausible account of what people compute, and how they compute it, in making such deductions. The theory outperforms its rivals that have computer implementations (Appendix C). And theories lacking such an implementation are hard to fit to data. It can even be unclear what conclusions they predict for each of the 64 sorts of syllogistic premises.

What the theory maintains from its precursors are the general nature of mental models, which tend to be schematic, parsimonious, and iconic; the interpretation of quantified assertions as relations between sets, which accommodates quantifiers such as "most artists" that cannot be captured in first-order logic; two systems of reasoning, which allow deliberation to refute conclusions based on intuitions; and the greater ease of inferences that depend on only one intuitive model. The theory introduces three principal innovations. First, it is founded on possibilities, and can draw conclusions about what is possible, probable, and necessary,

Table 5
The Values of mReasoner's Four Parameters in Simulation 8 for the Best-Fitting Simulations of Each of the 20 Participants, and the Pearson Correlations for the Best and Worst Fits of Their Data, Ranked in Descending Order by r_{best}

Participant	Values of the parameters and *r* values for the best and worst fits					
	`size`	`atypicality`	`search`	`reconclude`	r_{best}	r_{worst}
19	2.0	0.0	1.0	0.6	**.86**	**.07**
5	2.0	0.0	1.0	0.8	.86	.09
6	2.0	0.0	0.8	0.6	.82	.13
8	4.0	1.0	0.8	0.6	.82	.20
9	3.0	0.6	0.8	0.6	.81	.19
11	2.0	0.2	0.8	0.8	.80	.16
14	3.5	0.6	0.6	0.0	.74	.26
13	3.5	1.0	0.6	0.6	.73	.30
10	2.0	0.0	0.8	0.6	.72	.22
2	2.0	0.0	0.6	0.6	.68	.27
1	2.5	0.0	0.6	1.0	.68	.38
15	4.5	0.2	0.8	0.6	.66	.23
17	3.0	0.2	0.6	0.4	.66	.30
12	4.0	0.6	0.6	1.0	.65	.34
20	3.5	0.6	0.8	0.6	.64	.14
7	3.0	1.0	0.8	0.4	.63	.24
16	2.5	0.0	0.4	1.0	.58	.30
3	2.0	0.2	0.6	0.6	.58	.34
4	2.0	0.4	0.6	1.0	.57	.28
18	2.0	0.0	0.6	0.0	**.57**	**.35**

Note. The highlighted rows show the best and worst fitting simulations. The `size` parameter controlled the size of the models in the simulations; the `atypicality` parameter controlled the tendency to build typical models; the `search` parameter controlled the tendency to engage in a deliberative search for counterexamples; and the `reconclude` parameter controlled the ability to revise conclusions recursively.

and evaluate the consistency of assertions (see Johnson-Laird & Ragni, 2019, for an analogous treatment of sentential connectives). Second, it relies on intensions in order to ensure that any changes to a model remain faithful to the meanings of premises. Third, it uses heuristics, which reflect the mood and figure of premises, to formulate its initial conclusions. These heuristics supersede previous accounts (e.g., Begg & Denny, 1969; Chater & Oaksford, 1999; Wetherick & Gilhooly, 1995). They yield both the mood and the order of the terms in conclusions, and in principle individuals can acquire them from their experiences in reasoning. In sum, the current theory has several advantages over its rivals (see Appendix C for details).

The mReasoner implementation of the theory also embodies novelties. Unlike previous simulations, which use dozens of free parameters (e.g., Hattori, 2016; Polk & Newell, 1995), it relies on only four. They correspond to sensible psychological variables yielding testable predictions. They constrain the number of entities in an initial model, the degree to which it represents typical individuals or the full gamut of possibilities, the likelihood that the deliberative system searches for an alternative model, and, if it finds one, the likelihood that it seeks a weaker conclusion. The resulting fit to the data from over 200 different sorts of inference is good, and it includes inferences about the properties of individuals (see Figure 3), immediate inferences from quantified assertions (see Figure 4), and syllogistic inferences from pairs of quantified assertions (see Figure 2). And for the first time in studies of syllogisms, it yields a fit to the results from individual reasoners (see Table 5), and a plausible account of how they fall into three groups, ranging from the worst reasoners, who rely on intuition, through to the best reasoners, who try to deliberate.

One of the advantages of fitting computational theories to experimental results is that it makes their flaws salient. One of them was in mReasoner's predictions for inferences about the absence of a property (see Figure 3), and we suggested a possible explanation. Participants build a different initial model and do not consider alternatives to it. Likewise, in its fit to the 64 syllogistic inferences, it predicts too many responses of "NVC," many for premises that have no sensible valid conclusion, but that elicit erroneous conclusions from reasoners (see Figure 2).

A good theory should extend to other sorts of inference, and the model theory has several such extensions. Its semantics for quantifiers solves the well-known "paradox" of confirmation (Hempel, 1945). Consider a hypothesis such as:

23. All black holes have massive gravity

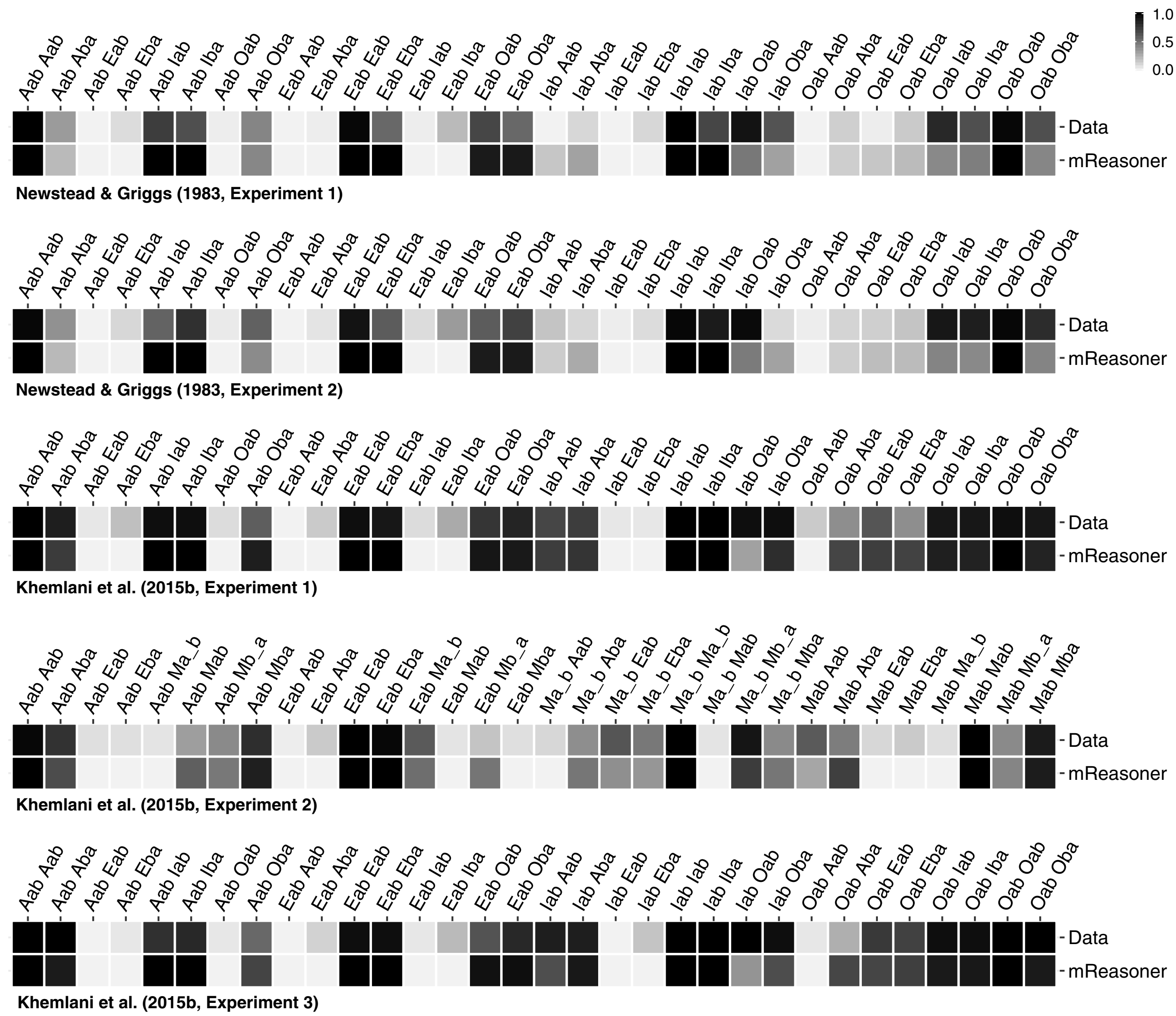
In first-order logic, it is equivalent to its contrapositive:

24. All things that do not have massive gravity are not black holes.

The existence of a teddy bear corroborates (24), because it does not have massive gravity and it's not a black hole; and therefore it also corroborates its logical equivalent (23) about black holes. The burden of Hempel's paper is that it would be paradoxical for a teddy bear to lend weight to a cosmological hypothesis. General claims such as (23) concern four sorts of entity: those that are, or are not, black holes, in conjunction with those that have, or don't have, massive gravity. In the model theory, the verification of (23) rests on only two of them: the possibility of black holes that have a massive

Figure 4

Simulations 2–6 of Two Studies From Newstead and Griggs (1983) and Three Studies From Khemlani, Lotstein, et al. (2015)



Note. The top rows show the proportions (from 0 to 1) of affirmative evaluations of immediate inferences, and bottom rows the mReasoner's best-fitting predictions. Each immediate inference is abbreviated as a premise and its conclusion, for example, an inference labeled Aab Aba denotes the following: *All of the A are B. Does it follow that all of the B are A?* But, for Khemlani et al.'s Experiment 1, the abbreviation denotes: *All of the A are B. Is it possible that all of the B are A?* Mab denotes: *Most of the a are b*, and *Ma_b = Most of the a are not b*.

gravity, and the impossibility of black holes that don't have a massive gravity. The other two sorts of entity—those that are not black holes and have or do not have massive gravity—are also possible according to (23), because it presupposes their existence. They are therefore irrelevant to its verification, because they are also possible given its negation:

25. Not all black holes have massive gravity.

In sum, teddy bears are not pertinent to verification of (23), and the paradox is resolved. Likewise, as Nicod (1950, p. 219) claimed, the two cases of black holes are crucial for the induction of general hypotheses. A Bayesian solution to the paradox also exists (e.g., Howson & Urbach, 1993, Ch. 7). The model explanation is simpler, and even naive individuals grasp it when they select evidence to test general hypotheses (see Ragni et al., 2018).

Sentential connectives, such as *if*, *or*, and *and*, can interrelate quantified clauses, as in:

26. One of these claims is true and one of them is false: Some of the architects are bankers or all of the architects are bankers.

The intuitive system 1 constructs two models of the alternative possibilities, for example:

```
First possibility              Second possibility
architect      banker          architect     banker
architect                      architect     banker
               banker                        banker
```

Given the question:

27. Is it possible that all of the architects are bankers?

most people answer "yes" (85% in an experiment, Yang & Johnson-Laird, 2000). It is a compelling but erroneous response—an "illusory" inference. It occurs because the intuitive models above represent what is true, and overlook that when one clause in the disjunction is true, the other clause is false. System 2 constructs a correct model of the disjunction, for example:

```
architect     banker
architect   ¬ banker
```

This model holds when the first disjunct in (26) is true and the second disjunct is false. When the first disjunct is false, none of the architects is a banker, which is inconsistent with the truth of the second disjunct, and so no model with any content results in this case. A simple control problem poses a question about the converse conclusion:

28. Is it possible that all of the bankers are architects?

Nearly everyone responds in the affirmative (95%), and as the model above shows the answer is correct. Similar illusions occur if the task is to assess whether or not pairs of quantified assertions are consistent (Kunze et al., 2010)—a task that is beyond the scope of the alternative theories.

Many quantifiers in everyday life cannot be expressed in the first-order predicate calculus. They include those based on determiners, such as *few, more than half*, and *most*, and they challenge many theories of quantifiers (Geurts et al., 2010). Unless the cardinality of the relevant sets is fixed, their inferential properties reflect the intersection of two sets of individuals, for example:

29. Most architects are bakers.
    All bakers are cyclists.
    Therefore, most architects are cyclists.

When the relevant cardinality is fixed, other sorts of inference may be invalid. The following example (from Kroger et al., 2008) illustrates this point:

30. 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 writers in the room is all three: a jogger, a writer, and a dancer?

Individuals who seek to minimize the overlap between the different sets are liable to assume that the inference is valid, but a counterexample exists in which the five students have these properties:

```
writer   jogger
writer   jogger
writer            dancer
         jogger   dancer
                  dancer
```

A model of such a possibility satisfies the premises but refutes the conclusion. In an fMRI study, Kroger et al. discovered that the attempt to generate such counterexamples activates the frontal pole of the right cerebral hemisphere, which is known to mediate other conflicts such as the Stroop effect.

Certain inferences depend on inferences about relations that cannot be treated as properties, for example:

31. Everyone loves anyone who loves someone else.
    Diana loves Charles.
    Therefore, everyone loves Diana.

Most people can make this inference (Cherubini & Johnson-Laird, 2004; see also Goodwin & Johnson-Laird, 2005; Johnson-Laird et al., 1989), but not the further valid inference:

32. Therefore, everyone loves everyone.

A previous program implementing the model theory can build models of the premises in (31). It calls for a model of the relation between Diana and Charles to be updated using the quantified premise to yield the conclusion in (31). But, naive individuals don't seem to realize that the quantified premise can be used to update the model again in order to yield the conclusion in (32).

The model theory predicts a striking pattern of reasoning from quantifiers that is inconsistent with the predicate calculus. The following inference is intuitively valid in the model theory:

33. Few of the customers had lobster or else steak.
    Therefore, few of the customers had lobster.

Reasoners tend to accept the conclusion (Johnson-Laird et al., 2021). The premise yields a model of this sort:

```
customer   lobster
customer              steak
customer
customer
```

The conclusion holds in this model. In contrast, the same inference but with the quantifier, "Most of the customers" is intuitively invalid in the model theory, and tends to be rejected. The crux is that the quantifier in the premise should hold for the subset to which the conclusion refers. Reasoners do not need to grasp this principle, because the status of inference is an emergent property of their models.

Finally, properties can have properties. The property of being an idea is an idea too. So, the set of ideas is a member of itself, just as, say, a bibliography of *all* bibliographies should list itself. In contrast, the property of being a Scot is not itself a Scot, but corresponds to a set of individuals. Logicians think about such matters, and the way they first did so mirrors errors of inference in daily life. Frege's (1964/1893) logic contained a devastating paradox, because it assumed that any property corresponds to a set. As Russell (1902) pointed out in a letter to him, which we paraphrase:

34. Suppose *w* is a property that is not a property of itself. The set corresponding to *w* is the set of all sets that are not members of themselves. Is *w* a member of itself? From each answer the opposite follows. Therefore, *w* is not a property.

Most bibliographies do not list themselves. So, what about a bibliography of all bibliographies that do not list themselves—should it list itself? From each answer the opposite follows. So, no such bibliography can exist. Russell's paradox is not so easy to dismiss.

As Frege replied, it shook the foundations of arithmetic. After some work, logicians axiomatized set theory in a way that is free of the contradiction. The model theory suggests a simple prophylactic: if a set such as *A* is to be formed from sets such as *B*, then the members of these latter sets need to be established first (Johnson-Laird, 1983, p. 429). The question of whether *A* includes itself cannot occur, because all of its members must be determined first.

Conclusions

No previous theory gave a satisfactory account of reasoning about properties (see, e.g., Khemlani & Johnson-Laird, 2012, for a meta-analysis). The discovery of their inadequacies was progress of a sort: If none of these theories could have been refuted, then a cognitive science of reasoning would have been impossible. The model theory accounts for what is computed in reasoning about properties and for how it is computed. The meanings of quantified assertions about properties are relations between sets. The representations of these intensions constrain models. Intuitive processes (in system 1) use them to construct initial mental models.

Deliberative processes (in system 2) can search for alternative models and can form new conclusions to accommodate them. Theories rise faster than they fall in cognitive science, and one brake on this unfortunate tendency is to implement theories in computer programs, such as mReasoner's simulation of the model theory. It reveals where the theory is wrong. In fact, it yields a good fit to the valid and invalid inferences that occur in experimental studies of various sorts of inference, from possession of a property to syllogisms. Reasoners err because they fail to consider atypical members of a set, fail to search for alternative models of premises, or, if they do find one, fail to weaken their initial conclusion. The program fits the performance of individuals, and shows that they fall into three main groups: those who stick to their intuitions, those who deliberate to attain the best performance, and those who vacillate between the two. The theory and its implementation cope with all sorts of quantifier, including "more than half the architects." And consequences include a range of phenomena from systematic fallacies—illusory inferences—to the selection of evidence to test general hypotheses. These phenomena are beyond the explanatory power of theories based on the first-order predicate calculus, Euler circles, Venn diagrams, or probabilistic logic. The model theory may not be the last word on reasoning about properties, but it is perhaps the most accurate word now available.

References

Adams, E. W. (1998). *A primer of probability logic*. Center for the Study of Language and Information.

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *The American Journal of Psychology*, *108*(2), 157–193. https://doi.org/10.2307/1423127

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

Baron, J., Badgio, P., & Gaskins, I. W. (1986). Cognitive style and its improvement: A normative approach. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 173–220). Erlbaum.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. In J. Kulas, J. H. Fetzer, & T. L. Rankin (Eds.), *Philosophy, language, and artificialIntelligence. Studies in cognitive systems* (vol 2). Springer. https://doi.org/10.1007/978-94-009-2727-8_10

Begg, I., & Denny, J. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. *Journal of Experimental Psychology*, *81*(2), 351–354. https://doi.org/10.1037/h0027770

Boole, G. (1854). *An investigation of the laws of thought*. Macmillan.

Braine, M. D. S. (1998). *Steps toward a mental-predicate logic*. Psychology Press.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*(3), 247–303. https://doi.org/10.1207/s15516709cog2303_1

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Sage Publications.

Cantor, G. (1895). Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, *46*, 481–512. https://doi.org/10.1007/BF02124929

Cassimatis, N. L., Bello, P., & Langley, P. (2008). Ability, breadth, and parsimony in computational models of higher-order cognition. *Cognitive Science*, *32*(8), 1304–1322. https://doi.org/10.1080/03640210802455175

Ceraso, J., & Provitera, A. (1971). Sources of error in syllogistic reasoning. *Cognitive Psychology*, *2*(4), 400–410. https://doi.org/10.1016/0010-0285(71)90023-5

Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, *58*(3), 220–226. https://doi.org/10.1037/h0041961

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, *38*(2), 191–258. https://doi.org/10.1006/cogp.1998.0696

Cherubini, P., & Johnson-Laird, P. N. (2004). Does everyone love everyone? The psychology of iterative reasoning. *Thinking & Reasoning*, *10*(1), 31–53. https://doi.org/10.1080/13546780342000025

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

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

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

Dickstein, L. S. (1975). Effects of instructions and premise order on errors in syllogistic reasoning. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 376–384. https://doi.org/10.1037/0278-7393.1.4.376

Dickstein, L. S. (1978a). The effect of figure on syllogistic reasoning. *Memory & Cognition*, *6*(1), 76–83. https://doi.org/10.3758/BF03197431

Dickstein, L. S. (1978b). Error processes in syllogistic reasoning. *Memory & Cognition*, *6*(5), 537–543. https://doi.org/10.3758/BF03198242

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). Lawrence Erlbaum.

Evans, J. St. B. T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language*, *35*(3), 392–409. https://doi.org/10.1006/jmla.1996.0022

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

Frege, G. (1964). *The basic laws of arithmetic*. (M. Furth, Ed. & Trans.), University of California Press. (Originally published, 1893).

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*(1), 16–25. https://doi.org/10.1037/0096-3445.115.1.16

Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, *86*(3), 223–251. https://doi.org/10.1016/S0010-0277(02)00180-4

Geurts, B., Katsos, N., Cummins, C., Moons, J., & Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, *25*(1), 130–148. https://doi.org/10.1080/01690960902955010

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*(2), 468–493. https://doi.org/10.1037/0033-295X.112.2.468

Guyote, M. J., & Sternberg, R. J. (1981). A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology*, *13*(4), 461–525. https://doi.org/10.1016/0010-0285(81)90018-9

Hartigan, J. A. (1975). *Clustering algorithms*. Wiley.

Hattori, M. (2016). Probabilistic representation in syllogistic reasoning: A theory to integrate mental models and heuristics. *Cognition*, *157*, 296–320. https://doi.org/10.1016/j.cognition.2016.09.009

Hempel, C. G. (1945). Studies in the logic of confirmation, Parts I & II. *Mind*, *LIV*, 1–26. https://doi.org/10.1093/mind/LIV.213.1

Hopcroft, J. E., & Ullman, J. D. (1979). *Formal languages and their relation to automata*. Addison-Wesley.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Open Court.

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

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

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

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

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

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

Johnson-Laird, P. N., Byrne, R. M., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, *96*(4), 658–673. https://doi.org/10.1037/0033-295X.96.4.658

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

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

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*(4), 646–678. https://doi.org/10.1037/0033-295X.109.4.646

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*(1), 62–88. https://doi.org/10.1037/0033-295X.106.1.62

Johnson-Laird, P. N., Oakhill, J., & Bull, D. (1986). Children's syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *38*(1), 35–58. https://doi.org/10.1080/14640748608401584

Johnson-Laird, P. N., Quelhas, A. C., & Rasga, C. (2021). *Mental models, free choice, and other paradoxical disjunctive inferences* [Under submission].

Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, *193*, Article 103950. https://doi.org/10.1016/j.cognition.2019.04.019

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

Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, *1*(2), 134–148. https://doi.org/10.1016/0010-0285(70)90009-5

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

Khemlani, S. (2021). Psychological theories of syllogistic reasoning. In M. Knauff & W. Spohn (Eds.), *Handbook of rationality (in press)*. MIT Press.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*(3), 427–457. https://doi.org/10.1037/a0026841

Khemlani, S., & Johnson-Laird, P. N. (2013). The processes of inference. *Argument & Computation*, *4*, 4–20. https://doi.org/10.1080/19462166.2012.674060

Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2014). A mental model theory of set membership. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 2489–2495). Cognitive Science Society.

Khemlani, S., Lotstein, M., Trafton, J. G., & Johnson-Laird, P. N. (2015). Immediate inferences from quantified assertions. *Quarterly Journal of Experimental Psychology*, *68*, 2073–2096. https://doi.org/10.1080/17470218.2015.1007151

Khemlani, S. S., Byrne, R. M. J., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science*, *42*(6), 1887–1924. https://doi.org/10.1111/cogs.12634

Khemlani, S. S., & Johnson-Laird, P. N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, *64*(11), 2276–2288. https://doi.org/10.1080/17470218.2011.592593

Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. N. (2015). Naive probability: Model-based estimates of unique events. *Cognitive Science*, *39*(6), 1216–1258. https://doi.org/10.1111/cogs.12193

Khemlani, S. S., Mackiewicz, R., Bucciarelli, M., & Johnson-Laird, P. N. (2013). Kinematic mental simulations in abduction and deduction. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *110*(42), 16766–16771. https://doi.org/10.1073/pnas.1316275110

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. https://doi.org/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). Cognitive Science Society.

Madruga, J. A. G. (1984). Procesos de error en el razonamiento silogístico: doble procesamiento y estrategia de verificación. In M. Carretero & J. A. G. Madruga (Eds.), *Lecturas de psicologia del pensamiento*. Alianza.

McDermott, D., & Doyle, J. (1980). Non-Monotonic logic I. *Artificial Intelligence*, *13*(1–2), 41–72. https://doi.org/10.1016/0004-3702(80)90012-0

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

Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, *44*(2), 12–36. https://doi.org/10.4064/fm-44-1-12-36

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

Newell, A. (1981). Reasoning, problem solving and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), *Attention and performance* (Vol. 8, pp. 693–718). Lawrence Erlbaum.

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*(5), 535–546. https://doi.org/10.1016/S0022-5371(83)90328-6

Nicod, J. (1950). *Foundations of geometry and induction*. The Humanities Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524496.001.0001

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*, *58*(7), 1225–1247. https://doi.org/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*(6), 1702–1712. https://doi.org/10.1037/0278-7393.26.6.1702

Partee, B. H. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry*, *6*(2), 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). Springer-Verlag. https://doi.org/10.1007/978-3-642-67458-7_1

Peirce, C. S. (1931–1958). In C. Hartshorne, P. Weiss, & A. Burks (Eds.), *Collected papers of Charles Sanders Peirce* (Vols. 1–8). Harvard University Press.

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

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007

Politzer, G. (2007). The psychological reality of classical quantifier entailment properties. *Journal of Semantics*, *24*(4), 331–343. https://doi.org/10.1093/jos/ffm012

Politzer, G., Van der Henst, J. B., Delle Luche, C., & Noveck, I. A. (2006). The interpretation of classically quantified sentences: A set-theoretic approach. *Cognitive Science*, *30*(4), 691–723. https://doi.org/10.1207/s15516709cog0000_75

Polk, T. A., & Newell, A. (1988). Modelling human syllogistic reasoning in Soar. *Proceedings of the Cognitive Science Society–1988* (pp. 181–187). Erlbaum.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, *102*(3), 533–566. https://doi.org/10.1037/0033-295X.102.3.533

Power, R. (1984). *Inverted deductions with numerical quantifiers* (Report 104). Institute of Psychology, Padua University.

Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: A theory of selection tasks. *Psychological Bulletin*, *144*(8), 779–796. https://doi.org/10.1037/bul0000146

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*(6), 584–592. https://doi.org/10.3758/BF03213778

Revlis, R. (1975). Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, *14*(2), 180–195. https://doi.org/10.1016/S0022-5371(75)80064-8

Rips, L. J. (1994). *The psychology of proof*. MIT Press. https://doi.org/10.7551/mitpress/5680.001.0001

Rips, L. J. (2002). Reasoning. In D. Medin (Ed.), Stevens' Handbook of Experimental Psychology, Vol. 2: *Memory and Cognitive Processes* (3rd Edn. pp. 317–362). Wiley.

Russell, B. A. W. (1902). Letter to Frege. In J. van Heijenoort (Ed.) [English translation in van Heijenoort, J. (Ed.) (1967)]. *From Frege to Godel*, (pp. 124–125) Harvard University Press.

Russell, B. A. W. (1905). On denoting. *Mind*, *XIV*, 479–493. https://doi.org/10.1093/mind/XIV.4.479

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology*, *34*(2), 109–159. https://doi.org/10.1006/cogp.1997.0665

Störring, G. (1908). Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv fur die Gesamte Psychologie*, *11*, 1–27.

Strawson, P. F. (1950). On referring. *Mind*, *59*, 320–344. https://doi.org/10.1093/mind/LIX.235.320

Tessler, M., & Goodman, N. (2014). Some arguments are probably valid: Syllogistic reasoning as communication. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the annual meeting of the cognitive science society* (Vol. 36, No. 36). Cognitive Science Society.

Wetherick, N. E., & Gilhooly, K. J. (1995). "Atmosphere", matching, and logic in syllogistic reasoning. *Current Psychology*, *14*(3), 169–178. https://doi.org/10.1007/BF02686906

Wilkins, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives de Psychologie*, *102*, Article 83.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, *18*(4), 451–460. https://doi.org/10.1037/h0060520

Yang, Y., & Johnson-Laird, P. N. (2000). How to eliminate illusions in quantified reasoning. *Memory & Cognition*, *28*(6), 1050–1059. https://doi.org/10.3758/BF03209353

Appendix A

A Primer on the Logic of Properties

The logic of properties goes back to Aristotle's invention of logic and to his account of syllogisms (see his *Prior Analytics*, Barnes, 1984), which dominated Western logic for over two millennia. Scholastic logicians extended his analysis and identified four *moods* of syllogistic assertions, which we present here with examples and their traditional abbreviations (based on *a*ff*i*rmo and n*e*g*o*):

i.

All A are B:	All Scots are Celts.	(Aab)
Some A are B:	Some Scots are Celts.	(Iab)
No A are B:	No Scots are Celts.	(Eab)
Some A are not B:	Some Scots are not Celts.	(Oab)

When we use Scholastic abbreviations such as Aab, we switch to lowercase letters to refer to the terms in a syllogism, such as Scots and Celts. The order of the two terms in each premise can be swapped round (e.g., *A–B*, or *B–A*), and so syllogisms can be arranged in one of four *figures*, which we number according to the following scheme (see Johnson-Laird & Bara, 1984a), which is closer to Aristotle's system than to that of the Scholastics:

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

Each premise can be in one of four moods, and so there are 16 moods in each figure, and therefore 64 possible pairs of premises. A conclusion can be in one of the four moods and in one of two orders: *A–C* or *C–A*, or it can be that "nothing follows," that is, no definite conclusion interrelates *A* and *C*. Hence, there are nine possible responses to an orthodox syllogism, and in principle 576 complete syllogisms. Medieval logicians defined a term that refers to all entities that it denotes as "distributed" (see, e.g., Cohen & Nagel, 1934, p. 37–38), and so in *All B are C*, *B* is distributed but *C* is undistributed, because the assertion says nothing about whether or not all *C* are *B*. The Scholastics identified valid syllogisms using such principles as:

(*Appendices continue*)

- The “middle” term that occurs in both premises (*B* in the figures above) must be distributed in at least one of them.
- No term can be distributed in the conclusion unless it is distributed in the premises.
- Nothing follows if both premises contain the *particular* determiner, “some.”
- Nothing follows if both premises are negative.
- If both premises are affirmative then the conclusion must be affirmative; otherwise, it is negative.

A modern analog to distribution is that in, *All B are C*, the term *B* is *monotonically decreasing* because anything that is included in *B* is itself included in *C*, whereas the term *C* is *monotonically increasing* because anything that *C* implies is in turn implied by *B* (Barwise & Cooper, 1981). The idea is perhaps clearer in the following example in which the arrows denote implications based on knowledge:

humans
↓
ii. All mammals are equipped with hearts.
↓
nutrient circulators

So, it follows that all humans are equipped with hearts, and therefore that all mammals are equipped with nutrient circulators.

Scholastic logicians analyzed immediate inferences from one premise to a conclusion, as did Aristotle, for example:

iii. Some debtors are neurotics.
Therefore, some neurotics are debtors.

They likewise assumed that all three terms, *A*, *B*, and *C*, in syllogisms refer to entities that exist—an assumption that the modern predicate logic does not make for the quantifiers *All A* and *No A*. And they assumed that syllogisms treat the predicate of an assertion as referring to a property. So, an inference such as:

iv. Some of the writers were guests.
All the guests tipped a porter.
Therefore, some of the writers tipped a porter.

is valid, but its analysis as a syllogism calls for “tipped a porter” to be treated as though it referred to a property, not a relation that holds between guests and porters. The British Broadcasting Company once aired a monastic debate conducted using Scholastic syllogisms. The monks were adept at transforming arguments into syllogisms, and one of their maneuvers was to translate relations into properties.

Syllogisms cannot elucidate inferences that depend on quantified relations, such as:

v. Pat respects anyone who respects Viv.
Viv respects herself.
Therefore, Pat respects Viv.

The modern predicate calculus handles such inferences and to a limited degree so do naive reasoners (see the General Discussion), and it includes syllogisms as a tiny fragment (see Jeffrey, 1981, p. 115–116). It uses two sorts of quantifier, which are interdefinable using negation: the universal quantifier, *any x*, and the existential quantifier, *some x*, where *x* is variable whose value can be any entity in the relevant situation. But, many quantifiers in natural language cannot be represented in this *first-order* predicate calculus. As Barwise and Cooper (1981) proved, quantifiers such as “more than half the artists” call for the *second-order* predicate calculus in which variables range over properties as well as individuals. Likewise, the rules for forming quantifiers in natural language are recursive, and so they can increase in their complexity, for example:

vi. The dogs that barked also wagged their tails.
The six dogs that barked also wagged their tails.
More than half of the six dogs that barked also wagged their tails.
Some of the more than half of the six dogs that barked also wagged their tails.

Experiments have studied only a handful of quantifiers, which are the focus of this article. One important contrast, however, is the difference between these two assertions:

vii.
a. All trespassers will be prosecuted.
b. All the trespassers will be prosecuted.

The first assertion makes no claim that any trespassers exist, but the definite description in the second assertion, *the trespassers*, is often analyzed as either asserting their existence (Russell, 1905) or presupposing it (Strawson, 1950). Some definite descriptions do neither, for example:

viii. If any trespasser enters the property then the trespasser will be prosecuted.

Its definite description in the *then*-clause neither asserts nor presupposes the existence of a trespasser. It merely refers back to a possible situation in which there is a trespasser. This referential relation seems to be what definite descriptions really mean. An entity or set of them has been mooted, and the definite description is an anaphor—it refers back to them. So, if the existence of trespassers has already been established, then (vii b) presupposes their existence. But, it fails to do so in a description such as:

ix. Some trespassers may enter the property today.
If they do then all of the trespassers will be prosecuted.

Experimenters sometimes instruct participants that entities do exist corresponding to the terms in syllogisms, which then use the definite article to make sure that this point is clear (see Johnson-Laird & Bara, 1984b).

One final development is essential to the model theory. In the second half of the 19th century, mathematicians developed what is now known as *naive* set theory (e.g., Boole, 1854; Cantor, 1895). The theory embodies many assumptions likely to underlie the ways in which naive individuals think about properties. It assumes that corresponding to each property, there is a set of entities having the property. For example, the earth is round, and so the earth is a member of the set of round entities. This relation of *set-membership* is the foundation of set theory. As Cantor wrote, a set is defined by its members, that is, the elements that make up the set. A single individual or a set can be a member of a set, so the assertion in inference (a) in the main text, *One of the professors is Joan*, asserts

(*Appendices continue*)

Table A1
The Meanings of Quantified Assertions About Properties as Relations Between Sets, Presented in Their Set-Theoretic Notations and Their Informal Descriptions

Quantified assertions	Set-theoretic notations	Informal descriptions
a is a *B*	$a \in B$	*a* is a member of *B*.
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*.	$A - B \neq \varnothing$	*A* that are not *B* exist.
Most *A* are *B*.	$\mid A \cap B \mid > \mid A - B \mid$	Cardinality of intersection of *A* and *B* is greater than that of *A* that are not *B*.
More than half of *A* are *B*.	$\mid A \cap B \mid > \mid A \mid / 2$	Cardinality of intersection of *A* and *B* is greater than that of half of that of *A*.

that Joan is a member of a particular set of professors; and, likewise, the assertion in (b), *All of the professors are experts in different disciplines*, has an interpretation that a particular set of professors is a member of the set of experts in different disciplines. As the inference in (b) illustrates, set-membership is not a transitive relation: if set *A* is a member of set *B*, and *B* is a member of set *C*, it does not necessarily follow that *A* is a member of *C*.

All set-theoretic notions depend on set-membership. For instance, if every member of one set, *A*, is a member of another set, *B*, then set *A* is included in set *B*, though there may be members of *B* who are not members of *A*. This relation of inclusion is transitive: if *A* is included in *B*, and *B* is included in *C*, then it follows that *A* is included in *C*. And, a central principle of set theory is:

x. If *x* is a member of set *A*, and set *A* is included in set *B*, it follows that *x* is a member of set *B*.

We now have in place all the elements for a logical catastrophe—a paradox—which we describe in the General Discussion, because its origins illuminate how individuals think about the properties of properties.

Modern axiomatic set theory is free from the paradox. It underlies Montague (1974) universal treatment of languages, which is based on "generalized" quantifiers (Mostowski, 1957). They allow Montague to treat all noun phrases, such as "Fred," "the present King of Albania, "some pets," "more than half of the dogs," in a uniform way in both their syntax and semantics. Each noun-phrase refers to a set of sets, so "Fred" refers to the set of all sets of which Fred is a member. A sentence such as: "Fred is left-handed," is true provided that the set of people who are left-handed is a member of the set to which "Fred" refers (see, e.g., Johnson-Laird, 1983, Ch. 8; Partee, 1975; Peters & Westerståhl, 2006). No plausible psychological theory, however, can rely on Montague grammar, because it calls for intractable computations, for example, "some dogs" calls for the representation of all sets containing some dogs. The concept of monotonically increasing and decreasing terms, which we described earlier, was first formulated using this semantics.

From the standpoint of psychology, an alternative set-theoretic interpretation to Montague's is much more plausible as an account of properties: Quantified assertions establish relations between sets corresponding to properties (Boole, 1854; Cohen & Nagel, 1934, p. 92). For example:

xi. More than half of the students are athletes

states a relation between the set of students and the set of athletes, which in modern notation, has the following meaning:

xii. $\mid \text{students} \cap \text{athletes} \mid > ½ \mid \text{students} \mid$

where "∩" denotes the intersection of two sets, and "| A |" denotes the cardinality of the set A. It can be paraphrased as the cardinality of those who are in the set of students and the set of athletes is greater than half the cardinality of the relevant set of students. Table A1 summarizes the set-theoretic representations of various quantified assertions, presenting the modern notation for each assertion and its informal paraphrase.

Appendix B

Methodology for Simulations 1–11

Simulation 1

Simulation 1 describes a study using mReasoner to fit data on how participants carried out set-membership reasoning. Khemlani et al. (2014) described a study in which participants received eight sorts of inference based on the following schema:

x is [not] an *A*.

[All/None] of the *A* are *B*/*B* are *A*.

What, if anything, follows?

Participants had to formulate their own conclusions. They tended to draw one of three sorts of conclusion: x is a B, x is not a B, or no valid conclusion (NVC).

Method and Procedure

To simulate set-membership inferences, mReasoner generated data sets by systematically varying the settings of its four parameters (Busemeyer & Diederich, 2010). The parameter settings were quantized to span their ranges as follows:

(*Appendices continue*)

`size`: 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0

`atypicality`: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0

`search`: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0

`reconclude`: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0

Hence, the system can be used to generate $7 \times 6 \times 6 \times 6 = 1{,}512$ separate simulated data sets. The system carried out the eight problems 100 times for each of the 1,512 parameter settings. In order to locate the best fitting parameter setting, a grid search was used to compute two goodness-of-fit metrics—Pearson correlation (*r*) and root mean squared error (RMSE) between the aggregated data and the proportions of responses in each simulated data set across the eight set-membership problems. When the best-fitting parameter setting was established, the parameters were fixed and mReasoner carried out the eight set-membership problems 1,000 times each. The proportions of responses were aggregated for the eight problems and the three possible responses to yield 24 datapoints.

Simulations 2–6

mReasoner simulated participants' immediate inferences concerning quantified premises in order to fit the results from five separate experiments.

Data sets

The data were from (Newstead & Griggs, 1983, Experiments 1 and 2) in which participants assessed whether a conclusion followed of necessity for 32 inferences of the following sorts:

[All/Some/Some_not/None] of the *A* are *B*.

Does it follow that [all/some/some_not/none] of the [*A* are *B*/*B* are *A*]?

Khemlani, Lotstein, et al. (2015, Experiment 1) carried out a similar study of 32 immediate inferences except that the question concerned the possibility of a conclusion:

Is it possible that [all/some/some_not/none] of the [*A*s are *B*s/*B*s are *A*s]?

Khemlani, Lotstein, et al. (2015, Experiment 2) examined inferences of the following sorts:

[All/Most/Most_not/None] of the *A* are *B*.

Is it possible that [all/most/most_not/none] of the [*A* are *B*/*B* are *A*]?

which included the second-order quantifier "most of the A." Khemlani, Lotstein, et al. (2015, Experiment 3) examined participants' assessments of the consistency of 32 classic pairs of assertions:

[All/Some/Some_not/None] of the *A* are *B*.

[All/some/some_not/none] of the [*A* are *B*/*B* are *A*].

Can both of these statements be true at the same time?

The computational implementation was fit to each of the 32 problems across all five of the studies described above.

Method and Procedure

None of the experiments required participants to formulate conclusions, and so mReasoner made no use of its heuristic processes (see Figure 3) to simulate the results. Likewise, the parameter governing the weakening of conclusions (`reconclude`) was irrelevant. The program computes that a conclusion is necessary if it holds in all of the models that it constructs, and that it is possible if it holds in at least one model. It computes that a set of assertions is consistent if it constructs a model in which they hold.

Simulations 2–6 followed procedures similar to those for Simulation 1. However, because mReasoner carried out fewer inferences for each simulation and one of its parameters was disabled, it was possible to increase the granularity of the values of the three parameters in the search for optimal fits. The values were as follows:

`size`: 2.5, 2.8, 3.0, 3.3, 3.5, 4.0, 4.3, 4.5, 4.8, 5.0

`atypicality`: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0

`search`: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0

They yield 1,000 distinct settings of the three parameters, and for each setting, it carried out 32 immediate inferences 100 times. An automated grid-search discovered the parameter settings of the best-fitting simulations for each of the five experiments. When a best-fitting parameter setting was found for a particular experiment, it was used to simulate the 32 inferences 1,000 times.

Simulation 7

For syllogistic reasoning, mReasoner simulated inferences for all 64 sorts of syllogistic premises.

Method and Procedure

mReasoner varied the settings for its four parameters. For each unique parameter setting, the system generated a data set in which it carried out 64 syllogisms 100 times. For each syllogism, the system could yield one of nine separate responses. Any theory has to account for the conclusions that reasoners draw and the conclusions that they don't draw, and so it needs to explain 576 separate datapoints (see Appendix A). For a tractable survey of the space of possible inferences, the parameter settings were quantized to span ranges identical to those in simulation 1 in order to yield 1,512 simulated data sets. A grid search located best-fitting parameter values to fit the aggregated data from the meta-analysis reported in Khemlani and Johnson-Laird (2012): The 576 simulated datapoints were compared with the actual data reported in the meta-analysis. The best-fitting values were used for mReasoner to carry out the 64 syllogisms 1,000 times each. The proportions of responses from this final simulated data set were aggregated across the 64 syllogisms and the nine possible responses to yield 576 datapoints.

Qualitative Analyses

In addition to the quantitative goodness-of-fit analyses reported in Table 4 in the main text, we carried out further analyses. The model theory predicts that those syllogisms for which the heuristics deliver a valid conclusion should be easier than those for which they deliver

(*Appendices continue*)

an invalid conclusion. The data from the meta-analysis corroborated the prediction: Reasoners produced correct conclusions 67% of the time for those inferences that mReasoner's heuristics delivered correct responses, but only 17% of the time when they did not (Mann–Whitney test, $z = 4.08$, $p < .0001$, in a by-materials analysis). For some syllogisms, mReasoner produces multiple conclusions, but for other syllogisms, the system produces only one response. It follows that the greater the diversity of predicted conclusions for a given pair of premises, the greater should be the difficulty for participants to formulate a correct conclusion. The results yielded the following trend over syllogisms for which mReasoner predicted

one predicted response:	62% correct;
two predicted responses:	46% correct;
three predicted responses:	30% correct;
four predicted responses:	21% correct.

The trend was reliable in a by-materials analysis (Jonckheere trend test, $z = 3.72$, $p < .0001$).

A qualitative theory, such as the atmosphere hypothesis described in the main text, enumerates the conclusions that reasoners are likely to draw, whereas a quantitative theory also predicts their frequencies, for example, 92% of the time. Khemlani and Johnson-Laird (2012) showed that quantitative and qualitative theories can be compared using a cutoff based on how often a particular conclusion (of the nine possibilities) could be generated by chance. An accurate theory should predict those responses that reasoners make (e.g., "hits") and it should not predict those responses reasoners do not make (e.g., "correct rejections"), and any theory, qualitative or quantitative, can be analyzed by comparing hits against correct rejections. Figure B1 shows how eight theories of syllogistic reasoning fare when this qualitative analysis is applied to their predictions: mReasoner is more accurate than verbal models (Polk & Newell, 1995; conversions of premises Chapman & Chapman, 1959), previous model theory (Johnson-Laird & Byrne, 1991), the atmosphere hypothesis (Begg & Denny, 1969), first-order predicate calculus in the Psychology of Proof (PSYCOP) program (Rips, 1994), probabilistic heuristics model (Chater & Oaksford, 1999), and the matching heuristic (Wetherick & Gilhooly, 1995).

Simulation 8

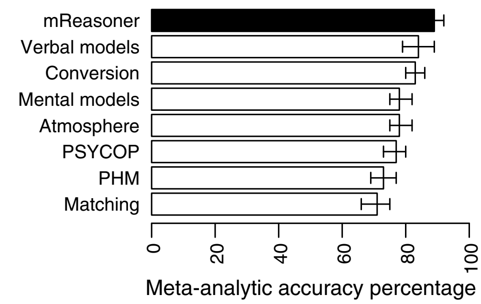
Separate simulations were carried out of each of the 20 participants' responses to the 64 syllogisms (Johnson-Laird & Steedman, 1978).

Method and Procedure

The same analysis was used as in the prior two simulations. The system generated a simulated data set for 1,512 unique settings of the four parameters, and for each setting, it carried out 64 syllogisms 100 times. An automated analysis discovered the parameter settings of the best-fitting simulations for each of the 20 participants. The proportions of responses were aggregated as described in Simulations 7 and 8–10.

Figure B1

Accuracies of Eight Theories of Syllogistic Reasoning Based on a Method to Assess Quantitative and Qualitative Theories (see Khemlani & Johnson-Laird, 2012)



Simulations 9–11

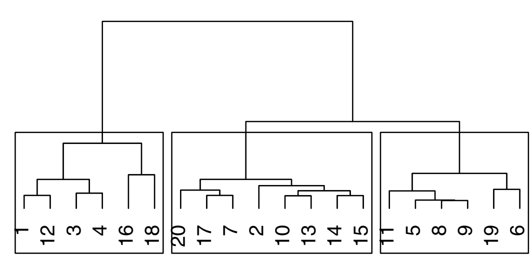
To separate syllogistic reasoners (in Johnson-Laird & Steedman, 1978 study) into sensible groups, we carried out an exploratory cluster analysis (Hartigan, 1975). It discovered similarities in participants' patterns of reasoning. The raw data from each participant were subjected to the partitioning around medoids clustering algorithm (Kaufman & Rousseeuw, 2009). It revealed that the optimal number of subsamples in the data was three. We used this estimate to constrain a hierarchical cluster analysis on the full range of participants' responses for the 64 syllogisms (see Hartigan, 1975). Figure B2 shows how the cluster analysis grouped the 20 participants. A separate simulation analysis was conducted for each of the three subsamples.

Method and Procedure

To simulate the individual differences between the different subsamples discovered in Johnson-Laird and Steedman (1978) experiment, mReasoner generated simulated data sets by

Figure B2

Dendrogram of a Hierarchical Cluster Analysis Performed on Data Provided by 20 Participants' Tendency to Yield 9 Syllogistic Reasoning Responses Pooling Across 64 Syllogisms From the Data (Johnson-Laird & Steedman, 1978)

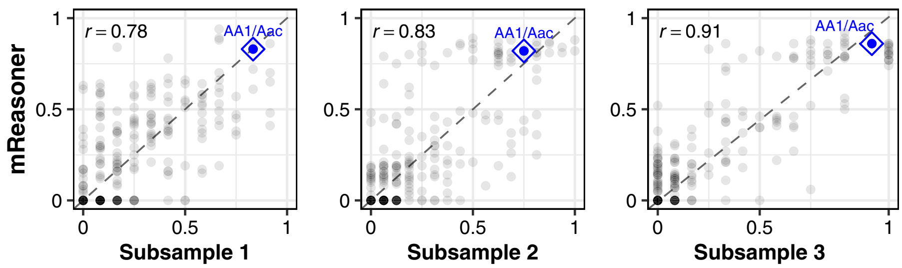


Note. Each leaf in the tree reports a participant's unique identifying number. The analysis separated the participants' behavior into three subsamples.

(*Appendices continue*)

Figure B3

Scatterplots of Three Simulations: 8 (Left Panel), 9 (Middle Panel), and 10 (Right Panel) Correlated With the Data From Subsamples 1, 2, and 3 (in Johnson-Laird & Steedman, 1978)



Note. Each dot represents a pair of syllogistic premises and the proportion of times participants drew a particular conclusion (from the nine possibilities) and the proportion of times that mReasoner predicted that response. So, each panel has 576 datapoints corresponding to the 64 syllogisms and the nine possible conclusions reasoners can draw. Hence, a perfect fit of the model to the data would be a straight line along the diagonal. To help readers made sense of the figure, it highlights one datum in blue for all three subsets of performance: Participants' tendency to conclude Aac from syllogism abbreviated AA1, that is, Aab Abc. See the online article for the color version of this figure.

systematically varying the parameter settings of its four parameters as described in Simulations 1 and 7. The system generated 1,512 different initial simulations, and a grid search located best-fitting parameter settings for each of the three subsamples. Given the best-fitting parameter settings, mReasoner carried out the 64 syllogisms 1,000 times each for the three subsamples. The proportions of responses were aggregated as described in Simulation 7.

Results

Figure B3 presents the data aggregated across the three subsamples yielded by the cluster analysis, along with mReasoner's best-fitting simulations of those data, and Table 4 reports the their goodness-of-fit metrics. There was a close fit between the simulation and the data (*r*s > .78, RMSEs < .14). The optimal parameter settings for the three subsamples are in Table B1.

As Figure B3 shows, reasoners' responses in subsample 1 appeared to vary more than in any other subsample. The parameter settings for the subsamples explain their general characteristics. Subsample 1 built small models of only two wholly typical individuals, their likelihood of searching for counterexamples was only slightly better than chance. So, reasoners in subsample 1 were intuitive and the least accurate. Subsample 2 tended to build larger models, which quite often represented atypical individuals, searched for counterexamples, and in case they found one tended to weaken their initial conclusions. They deliberated and were of intermediate ability. Subsample 3 was deliberative too, and differed from the previous group in that they built models that were 50% larger. This factor helped them to be the most accurate group of all.

Table B1

The Optimal Parameter Settings for Three Subsamples of Participants in a Syllogistic Study of 64 Sorts of Syllogism (Johnson-Laird & Steedman, 1978)

Subsample	size	atypicality	search	reconclude
Subsample 1	2.0	.00	.60	.60
Subsample 2	3.0	.60	.80	.60
Subsample 3	4.5	.40	.80	.60

(*Appendices continue*)

Appendix C

Advantages and Disadvantages of Computational Theories of Syllogistic Reasoning

Table C1

A Summary of Advantages and Disadvantages of Existing Computer Implementations of Theories of Syllogistic Reasoning Based on Mental Models Compared to the Unified Account. Other Theories Exist but Have No Implementations in the Public Domain

Desideratum	Johnson-Laird & Byrne (1991)	Polk & Newell (1995)	Hattori (2016)	Tessler & Goodman (2014)	Model theory: mReasoner
Meta-analytic prediction accuracy (see note)	78%	84%	n/a	n/a	**89%**
Can model difficulty between syllogisms	+	—	—	—	**+**
Yields quantitative predictions	—	+	+	+	**+**
Generates conclusions	+	+	—	—	**+**
Evaluates given conclusions	+	+	+	+	**+**
Can infer that no valid conclusion (NVC) follows	+	+	—	—	**+**
Can infer if a conclusion is possible	—	—	—	—	**+**
Can infer if a conclusion is necessary	+	—	+	+	**+**
Can assess consistency	—	—	—	—	**+**
Copes with nonstandard quantifiers (e.g., “most”)	—	—	+	+	**+**
Copes with numerical quantifiers (e.g., “both”)	—	—	—	—	**+**
Can model individual differences	—	—	—	—	**+**
Can model set membership inferences	—	—	—	—	**+**
Can model immediate inferences	—	—	—	—	**+**

Note. “+” indicates that the theory accounts for the desideratum, and the bolded column shows the features of the theory proposed in this article. The numbers reported for the meta-analytic prediction accuracy were computed using the same method as the meta-analysis reported in Khemlani and Johnson-Laird (2012)—see Simulation 7 above. In order to compare different theories' qualitative predictions, the method concerns what is predicted more often than chance rather than exact quantitative prediction.

Received May 19, 2020
Revision received April 18, 2021
Accepted April 18, 2021 ■