

The evaluation of the consistency of quantified assertions

**Marco Ragni, Sangeet Khemlani &
P. N. Johnson-Laird**

Memory & Cognition

ISSN 0090-502X

Mem Cogn

DOI 10.3758/s13421-013-0349-y



Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

The evaluation of the consistency of quantified assertions

Marco Ragni · Sangeet Khemlani · P. N. Johnson-Laird

© Psychonomic Society, Inc. 2013

Abstract Consistency is a hallmark of rationality, and this article reports three experiments examining how reasoners determine the consistency of quantified assertions about the properties of individuals—for example, *All of the actors are waiters*. The mental model theory postulates that reasoners determine consistency by trying to construct a model of all of the assertions in a set. As the theory predicts, consistency is easier to establish when many different sorts of individuals satisfy the assertions (Exp. 1), when the predicted initial mental model satisfies them (Exp. 2), and otherwise when the model satisfying them is not too distant from the initial model, according to Levenshtein's metric (Exp. 3).

Keywords Consistency · Inconsistency · Mental models · Logic · Reasoning

The importance of consistency goes back to Aristotle (*De Interpretatione*; see Edghill, 1928). He argued that pairs of assertions—such as *Every man is just* and *No men are just*—are contrary to one another. Contraries cannot both be true, although they can both be false, as in the case that only some of men are just. In contrast, in a pair of contradictory assertions, one is true and one is false. Three assertions, however, can be inconsistent

even though any pair of them are consistent, and indeed any number of assertions can be inconsistent even though any proper subset of them is consistent. Hence, *inconsistency* is a broader notion than *contradiction*. Inconsistency is also central to systems of logic in which the validity of an inference is established by showing that the negation of its conclusion conjoined with the premises yields an inconsistent set of assertions (e.g., Jeffrey, 1981). In other words, valid reasoning and the assessment of consistency are different aspects of deductive competence.

In logical terminology, a set of assertions is *satisfiable*—that is, consistent—if all of the assertions hold in at least one possible way (see Jeffrey, 1981, p. 8; Tarski, 1944). Any number of assertions can occur in a set, and the task of assessing satisfiability is computationally intractable; that is, with an increasing number of assertions, the task soon becomes impossible for any finite organism in a feasible amount of time (Garey & Johnson, 1979; Ragni, 2003). A corollary of intractability is that human reasoners (and digital computers) can cope with only small-scale problems. Our concern is the relative difficulties of different problems, because this helps to elucidate the mental processes underlying the evaluation of consistency.

A typical example of the problems that we have studied is as follows:

- (1) Some of the actors are waiters.
Some of the waiters are singers.
All of the singers are actors.
Can all three assertions be true at the same time?

A moment's thought should convince the reader that the assertions can all be true at the same time, and so the set is consistent. These sorts of assertion are monadic—that is, they refer to the properties of sets of entities—and they contain quantifiers, such as *some of the actors* and *all of the singers*. And they occur in syllogisms, which are deductions from two such premises to a conclusion—a form of inference that Aristotle was the first to study (see Barnes, 1999). In what follows, we describe a theory of how individuals assess consistency, and then describe

M. Ragni (✉)

Center for Cognitive Science, University of Freiburg,
Friedrichstr. 50, 79098 Freiburg, Germany
e-mail: ragni@cognition.uni-freiburg.de

S. Khemlani

Navy Center for Applied Research in Artificial Intelligence,
Naval Research Laboratory, 4555 Overlook Avenue SW,
Washington, DC 20375, USA
e-mail: skhemlani@gmail.com

P. N. Johnson-Laird

Department of Psychology, Princeton University, Princeton, USA
e-mail: phil@princeton.edu

P. N. Johnson-Laird

Department of Psychology, New York University, 6 Washington
Place, New York, NY 10003, USA

the results of experiments corroborating its three principal predictions. Finally, we discuss the implications of these results for alternative theories of reasoning.

The model theory of consistency

The theory of mental models—the “model theory” for short—gives an account of reasoning (e.g., Johnson-Laird, 2006; Johnson-Laird & Byrne, 1991), and we will formulate a theory of how individuals evaluate consistency on the basis of mental models. It postulates that individuals try to construct a single mental model of all of the assertions in a set. If they succeed, they evaluate the set as being consistent; otherwise, they evaluate it as being inconsistent. No one has hitherto studied the consistency of sets of quantified assertions, so our empirical studies concern the simplest possible instances—monadic assertions, which can be captured in first-order predicate logic, in which variables range over individual entities (see, e.g., Jeffrey, 1981). The model theory is based, not on logic, but on representations of sets, so if it fails with such assertions, it can be eliminated at once. Yet we emphasize that the theory has broader applications, which we will describe in the [General Discussion](#).

Individuals can use a description to construct a mental model of the corresponding situation. Consider assertions that describe the properties of sets of entities, which, as in all the cases that we will discuss, are sets of human beings—for example,

- (2) All of the actors are waiters.

The theory postulates two interleaved processes in its comprehension. One process composes the meaning of the assertion (its *intension*) out of the meanings of the words in the sentence according to the grammatical relations amongst them. Knowledge may modulate the composition and also help individuals determine the particular sets referred to in the assertion. The other process uses the intension to construct a mental model of the situation under description (the *extension* of the assertion). A mental model represents what is common to a set of possibilities in terms of the entities, their properties, and the relations amongst them (Barwise, 1993). In principle, each distinct sort of possibility requires a separate mental model. Because a subsequent assertion may call for the updating of the model, the mental representation of intensions is crucial to ensure that any revisions still yield a model of the description as a whole. The following diagram denotes a typical model of assertion (2) above:

- (3) actor waiter
 actor waiter
 actor waiter

Each row in the diagram denotes a separate entity, and the intension of the assertion allows for the presence of other sorts

of entity. Of course, we do not suppose that mental models consist of nouns, which we use to stand in for models of actual entities in the world.

Readers familiar with set theory will realize that the theory treats quantified assertions as describing relations among sets (see Johnson-Laird, 1983, p. 140). The preceding model accordingly represents the two sets as being coextensive, which is the preferred interpretation for such universal assertions (Bucciarelli & Johnson-Laird, 1999), and it does so iconically (in the sense of Peirce, 1931–1958, Vol. 4, para. 418 et seq.; that is, the set-theoretic structure of the model corresponds to the set-theoretic structure of the situation under description. Mental models are based on a principle of truth (e.g., Johnson-Laird & Byrne, 2002): They represent what is true, but not what is false, unless an assertion explicitly refers to falsity. This bias reduces the processing load on working memory. For example, the model above does not represent that it is false that there is an actor who is not a waiter, although assertion (2) has this meaning, too. As long as reasoners have access to an intensional representation of the assertion, they can modify the preferred model above in any way that satisfies the intension. Hence, they can modify it to represent the set of actors as being properly included within the set of waiters:

- (4) actor waiter
 actor waiter
 actor waiter
 waiter

According to the theory, reasoning is based on models. A conclusion that holds in all of the models of the premises is necessary, one that holds in most of these models is probable, and one that holds in at least one model is possible. When premises yield multiple models, the increase in the processing load on working memory makes inferences more difficult—they take longer and are more prone to error (e.g., Johnson-Laird & Byrne, 1991). Similarly, the failure to represent what is false can result in systematic fallacies (e.g., Johnson-Laird, 2006). Many findings have shown that reasoners often fail to consider all possible models of a set of premises, but instead consider only a small subset of them, and often just a single model (Bauer & Johnson-Laird, 1993; Vandierendonck, Dierckx, & De Vooght, 2004). Theorists have accordingly developed and corroborated an account of *preferred* mental models in certain domains, such as spatial reasoning (Jahn, Knauff, & Johnson-Laird, 2007; Knauff, Rauh, & Schlieder, 1995; Ragni, Knauff, & Nebel, 2005; Rauh et al., 2005; Rauh, Schlieder, & Knauff, 1997). In terms of a dual-process account of reasoning, preferred models are developed in System 1, which has no access to working memory (Johnson-Laird, 1983, chap. 6). The search for alternative models is recursive and calls for access to working memory. It is accordingly handled in System 2. Reasoners can progressively modify their preferred model (Ragni, Tsenden, &

Knauff, 2007; Rauh et al., 2005; Schleipen, Ragni, & Fangmeier, 2007), but they tend to make only a small number of changes (e.g., Ragni, Fangmeier, Webber, & Knauff, 2007).

A general method to assess the consistency of assertions is to try to build a model in which they all hold. If such a model exists, then they are consistent; otherwise, they are inconsistent. Evidence suggests that this method is one that individuals use in evaluating assertions based on sentential connectives, such as “if” and “or” (see, e.g., Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000).

The simplest monadic assertions in logic can be in one of four different “moods”:

- (5) All of the *a* are *b*.
- Some of the *a* are *b*.
- None of the *a* are *b*.
- Some of the *a* are not *b*.

where *a* and *b* denote sets of entities, such as actors and waiters. Table 1 shows the preferred mental models of each of these sorts of assertion (see Johnson-Laird, 2006), and the models are iconic representations of the relations between the sets, except that negation has to be represented symbolically. Preferred models are based on the assumption that each set has members, and so each model of a monadic assertion about A and B contains at least one token of A and at least one token of B and represents the appropriate relation between the two sets.

An assertion that refers to a set in an existing model calls for a procedure that updates the model. Consider, for instance, the following pair of assertions:

- (6) Some of the actors are waiters.
- Some of the waiters are singers.

The first assertion yields a model akin to the one shown in Table 1. The second assertion updates this model by

Table 1 The preferred mental models for four simple monadic assertions

Assertion	Initial Preferred Model
All of the <i>a</i> are <i>b</i>	<i>a b</i>
Some of the <i>a</i> are <i>b</i>	<i>a b</i>
	<i>a</i>
None of the <i>a</i> are <i>b</i>	<i>a ¬b</i>
	<i>b</i>
Some of the <i>a</i> are not <i>b</i>	<i>a ¬b</i>
	<i>a</i>
	<i>b</i>

The table shows only the different sorts of entities in each model rather than the number of instances of each token, of which there are a small but arbitrary number; “¬” is a symbol for negation; and the intensions of the assertions allow for certain other sorts of models

establishing coreferential relations between the two sets of waiters referred to in both assertions:

- (7) actor waiter singer
- actor

This model, as in Table 1, shows only the different sorts of entities in the model rather than the actual numbers of them, and we will adopt this convention in the rest of the article unless otherwise stated. A third assertion, such as

- (8) Some of the actors are singers,

holds in this model, so the three assertions are consistent. There are many other possible models of the first two assertions, but they are irrelevant for the consistency of the three assertions, given that the third assertion holds in model (7).

The preceding example illustrates the principle that initial models of affirmative assertions—their preferred models—tend to maximize the number of properties that entities have in common, and thereby reduce the load on working memory. This effect is a consequence of the assumption of coreference between the two sets of entities common to both assertions in (6). Model (7) represents explicitly only two sorts of entities. It could have represented various other sorts of entities, such as actors who are singers but not waiters, but such entities are ruled out by the principle underlying preferred models.

In contrast to the previous problem, consider this case:

- (9) Some of the actors are waiters.
- Some of the waiters are singers.
- None of the actors are singers.
- Could these three assertions all be true at the same time?

The first two assertions have the same preferred model (7) as before. The third assertion, however, does not hold in this model. Some reasoners may accordingly respond that the three assertions cannot all be true at the same time. However, reasoners who search for an alternative may find a model of the following sort:

- (10) actor waiter
- waiter singer
- actor

This model satisfies the three assertions in (9), so the correct response is that they can all be true at the same time.

The search for alternative models is constrained by the parameters in the intensions of assertions. In principle, the search could be random, with the rejection of any model that violates an intension: Any model that then remains is an alternative to the preferred model. However, the theory postulates instead that human reasoners use three operations to search for alternatives: the *addition* to models of tokens representing new entities or new properties, the *movement* of properties from one sort of entity to another, and the *breaking*

of entities with several properties into two sorts of entities. Model (10) results from the third operation—that is, breaking those who are actors, waiters, and singers into two sorts of entities: those who are actors and waiters, and those who are waiters and singers. When individuals made deductions using external models consisting of cutout shapes, they used these three operations (Bucciarelli & Johnson-Laird, 1999). They can be decomposed into two primitive operations: the addition and the deletion of tokens.

These manipulations of models can be assessed using Levenshtein's (1966) "edit" measure to determine the distance between an initial model and a new model. The closer the distance between two models, the greater their similarity. The metric was originally devised in order to compare strings of symbols: It computes the distance between them, using the number of edits needed to convert one string into the other; here, an "edit" consists of deleting a symbol, inserting a symbol, or substituting one symbol for another. In our case, the three operations that transform one model into another correspond to edits, and they can be decomposed into the insertion of a token representing a property into a model and the deletion of such a token from a model (see Fig. 1). For example, the operation of breaking an individual into two can be treated as a combination of both of these primitive operations: First, tokens are deleted from an individual, and then tokens are inserted into a model to form a new individual. In Fig. 1, we present an example for problem (9) above. The algorithm for computing the Levenshtein distance and its implementations in various programming languages are available at http://en.wikibooks.org/wiki/Algorithm_Implementation/Strings/Levenshtein_distance.

We have added a new component to a computer program, mReasoner, which carries out reasoning in various domains, for evaluating the consistency of monadic assertions. (The source code of the program is available at [http://](http://mentalmodels.princeton.edu/programs/mreasoner/)

mentalmodels.princeton.edu/programs/mreasoner/.) The theory and its computational implementation make three principal predictions about the evaluation of consistency. First, a direct consequence of model-based processing is that the greater the number of different sorts of entities that can occur according to a set of assertions, the easier it should be to decide that the set is consistent. If a set of assertions is consistent with any sort of entity defined in terms of the presence or absence of each property in the assertions, the task should be very easy, because any model is possible—for example, *some of the actors are waiters; some of the waiters are not singers; some of the singers are actors*. In this case, anything goes as long as at least one of the actors also waits and at least one of the singers also acts. However, if a set of assertions allows for only a smaller number of distinct sorts of entity, participants have to search harder to find a model containing them, and so it should be more difficult to decide that the set of assertions is consistent. Second, the search for an alternative model adds to the difficulty of the task, so problems for which the preferred model satisfies the set of assertions should be easier than problems for which only an alternative to the preferred model does. Third, the more similar a model of the assertions is to the preferred model, the easier it should be to evaluate the assertions as being consistent. Conversely, the more edits are necessary to reach the new model, the farther away it is from the original model in the space defined by the Levenshtein distance, and the more difficult it should be to evaluate the assertions as being consistent. Our experiments accordingly tested these three predictions in the domain of simple monadic assertions: The consistency of these assertions should be easier to establish when many different sorts of entities satisfy them, when the preferred model satisfies them, and otherwise when the model satisfying them is similar to the preferred model. A way to assess the latter prediction is by using Levenshtein's metric—for example, one or two changes should be easy for most participants, whereas more than two changes should be more difficult.

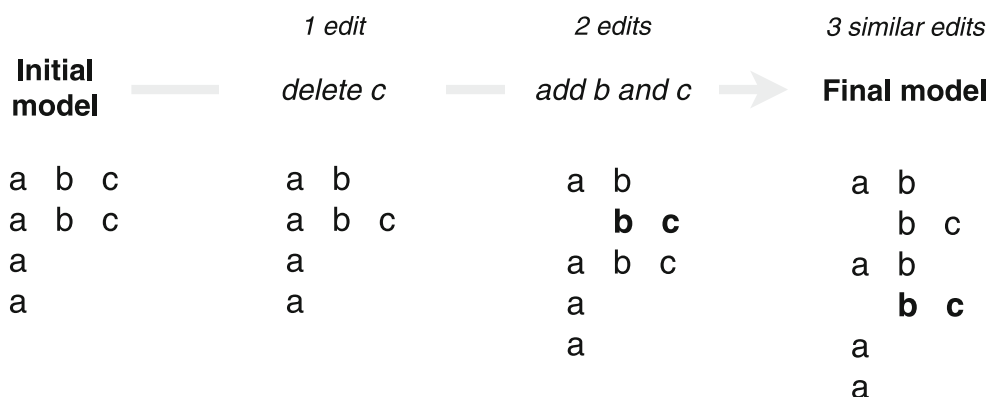


Fig. 1 Initial model of the first two assertions in problem (9). A sequence of three edits breaks each token of A B C into two individuals, so a total of six edits yields the final model that satisfies all three assertions. The Levenshtein distance between the initial and final models is accordingly 6

Experiment 1

The model theory predicts that the greater the number of different sorts of individuals that could exist given a set of assertions, the easier it should be to decide that the set is consistent. In the first experiment, we tested this prediction in a design in which the number of possible entities was independent of whether or not the preferred model satisfied the assertions, because all of the problems called for a search for an alternative to the preferred model.

Method

Participants We tested 25 logically naive participants (four men, 21 women; $M = 31.2$ years) on an online website (Amazon's Mechanical Turk, hereafter MTurk). They received a nominal fee for their participation. Previous experience had shown that the processes of reasoning are similar over different reasoners, and that robust results call for no more than 20–30 participants in an experiment (see, e.g., Johnson-Laird et al., 2000).

Design and materials The participants carried out 36 problems consisting of 22 consistent and 14 inconsistent problems (the six types of inconsistent problems were presented several times). For half of the problems, the arrangement of the terms (the *figure*) of the first two assertions was $a-b, b-c$; and for half of the problems, the figure of the first two assertions was $b-a, c-b$. The third assertion always had the figure $c-a$. These constraints made it easy to counterbalance the order of the assertions. Each consistent problem contained two affirmative assertions and one negative assertion, and the negative assertion occurred either first, second, or third in the set. Half of the negative assertions were of the form *Some of the a are not c* , and half of them were of the form *None of the a are c* .

The experiment included three sorts of consistent problem (see Table 2). First, six problems were based on assertions consistent with all eight possible individuals on the basis of the presence or absence of three independent properties—for example,

- (11) Some of the a are not b .
 Some of the b are c .
 Some of the c are a .

These assertions do not eliminate any individuals from the set of eight possible individuals defined in terms of the presence or absence of a , b , or c . Second, 12 problems were consistent with only six individuals out of the eight possibilities—for example,

- (12) None of the a are b .
 Some of the b are c .
 Some of the c are a .

These assertions eliminate individuals who are both a and b , whether or not they are c . Third, four problems were consistent with only five individuals out of the eight possibilities—for example,

- (13) None of the a are b .
 All of the b are c .
 Some of the c are a .

These assertions eliminate those individuals who are both a and b , and those who are b and not c —that is, they eliminate $ab\bar{c}$, $a\bar{b}c$, and $\bar{a}b\bar{c}$. In six problems, the assertions were inconsistent (see Table 3 in the Appendix). The problems were presented in a different random order to each participant.

Materials The contents of the problems concerned common occupations, such as *writers*, *waiters*, and *actors*. We devised a list of occupations, and each of these items was a common two-syllable plural noun. Nouns were selected from the list and assigned at random twice to each form of problem; half of the participants were tested with one allocation, and half of them were tested with the other allocation.

Procedure The experiment was presented online, and we took the usual precautions for such a procedure; for example, the program checked that participants were native speakers of English, and it allowed only one participant from a given computer. The instructions explained that the task was not a test of intelligence or personality, but concerned how people in general determine whether sets of assertions are consistent—that is, whether they can all be true at the same time. The participants were told that they would read sets of three assertions, and for each set they had to answer the question, Is it possible for all three assertions to be true at the same time? The first two assertions were presented simultaneously, and after a spacebar press, these two assertions disappeared and the third assertion and the question appeared. We used this procedure to record the total time that participants needed for reading the third assertion and deciding whether it was consistent with the previous assertions. The participants responded by pressing one of two buttons on the screen, one labeled “Yes” and one labeled “No.” They could take as much time as they needed, but they had to try to answer correctly. In this experiment and both of the subsequent ones, we used a flash implementation of mTurk, so the participants' response times were recorded on their computers and then sent to us. Hence, the response times do not depend on any transmission time. Before we carried out any of the experiments, we made extensive tests of this mTurk system to ensure that the response times were reliable (see also Paolacci, Chandler, &

Table 2 Forms of the three sorts of consistent problems in Experiment 1 and the percentages of correct responses to them

Problems Consistent With Eight Sorts of Entity		Problems Consistent With Six Sorts of Entity		Problems Consistent With Five Sorts of Entity		
Some of the <i>b</i> are not <i>a</i> Some of the <i>c</i> are <i>b</i> Some of the <i>a</i> are <i>c</i>	92	Some of <i>b</i> are not <i>a</i> Some of the <i>c</i> are <i>b</i> All of the <i>a</i> are <i>c</i>	62	None of the <i>b</i> are <i>a</i> Some of the <i>c</i> are <i>b</i> Some of the <i>a</i> are <i>c</i>	77	Inconsistent problem (see Table 3)
Some of the <i>b</i> are <i>a</i> Some of the <i>c</i> are <i>b</i> Some of the <i>a</i> are not <i>c</i>	96	Some of the <i>b</i> are <i>a</i> All of the <i>c</i> are <i>b</i> Some of <i>a</i> are not <i>c</i>	65	Some of the <i>b</i> are <i>a</i> Some of the <i>c</i> are <i>b</i> None of the <i>a</i> are <i>c</i>	73	Inconsistent problem (see Table 3)
Some of the <i>a</i> are <i>b</i> Some of the <i>b</i> are not <i>c</i> Some of the <i>c</i> are <i>a</i>	100	Some of the <i>a</i> are <i>b</i> Some of <i>b</i> are not <i>c</i> All of the <i>c</i> are <i>a</i>	77	Some of the <i>a</i> are <i>b</i> None of the <i>b</i> are <i>c</i> Some of the <i>c</i> are <i>a</i>	73	All of the <i>a</i> are <i>b</i> 35 None of the <i>b</i> are <i>c</i> Some of the <i>c</i> are <i>a</i>
Some of the <i>a</i> are not <i>b</i> Some of the <i>b</i> are <i>c</i> Some of the <i>c</i> are <i>a</i>	96	Some of the <i>a</i> are not <i>b</i> All of the <i>b</i> are <i>c</i> Some of the <i>c</i> are <i>a</i>	77	None of the <i>a</i> are <i>b</i> Some of the <i>b</i> are <i>c</i> Some of the <i>c</i> are <i>a</i>	73	None of the <i>a</i> are <i>b</i> 35 All of the <i>b</i> are <i>c</i> Some of the <i>c</i> are <i>a</i>
Some of the <i>a</i> are <i>b</i> Some of the <i>b</i> are <i>c</i> Some of the <i>c</i> are <i>a</i>	96	All of the <i>a</i> are <i>b</i> Some of the <i>b</i> are <i>c</i> Some of the <i>c</i> are not <i>a</i>	92	Some of the <i>a</i> are <i>b</i> Some of the <i>b</i> are <i>c</i> None of the <i>c</i> are <i>a</i>	73	All of the <i>a</i> are <i>b</i> 42 Some of the <i>b</i> are <i>c</i> None of the <i>c</i> are <i>a</i>
Some of the <i>b</i> are <i>a</i> Some of the <i>c</i> are not <i>b</i> Some of the <i>a</i> are <i>c</i>	100	All of the <i>a</i> are <i>b</i> Some of <i>c</i> are not <i>b</i> Some of the <i>a</i> are <i>c</i>	77	Some of the <i>b</i> are <i>a</i> None of the <i>c</i> are <i>b</i> Some of the <i>a</i> are <i>c</i>	62	All of the <i>b</i> are <i>a</i> 38 None of the <i>c</i> are <i>b</i> Some of the <i>a</i> are <i>c</i>

Where no consistent problem exists, the cell states, “inconsistent problem”

Ipeirotis, 2010, for additional reliability analyses of the mTurk platform).

Results and discussion

We excluded the data from six participants, who evidently guessed their responses and did not perform reliably better than chance. Table 2 presents the percentages of correct responses to each of the three sorts of consistent problems, and Fig. 2 presents the accuracy of the participants’ evaluations of the consistent problems and the corresponding response times as a function of

the number of distinct entities that the assertions allowed. The trends were reliable for both accuracy (Page’s *L* test, $z = 5.30$, $p < .0000001$, one tail, as are all reported tests of predictions) and response times (Page’s *L* test, $z = 2.31$, $p = .01$). We used nonparametric statistical tests in all of our analyses in order to obviate problems of distribution, and also because they would allow us to test the reliability of the predicted ordinal trends. They test solely for a stochastic increase from one condition to another, so they are less powerful than parametric tests, such as analyses of variance, and therefore are less likely to lead to an incorrect rejection of the null hypothesis (a Type I error). The order of the assertions within each sort of problem had no reliable effect on accuracy (73 % correct when the negative assertion was first, 77 % when it was second, 70 % when it was third; Friedman nonparametric analysis of variance, $\chi^2 = 1.32$, $p > .5$). Accuracy did, however, depend on the sort of negation: consistent assertions that included *Some of the ___ are not ___* were evaluated correctly for 86 % of problems, whereas consistent assertions that included *None of the ___ are ___* were evaluated correctly for only 58 % of problems (Wilcoxon test, $z = 4.04$, $p < .001$, two-tailed test).

This factor, however, was not independent of the number of entities consistent with the assertions: *Some of the ___ are not ___* does not eliminate any entities, whereas *None of the ___ are ___* eliminates two entities. The two sorts of assertions did occur in problems consistent with six entities, which yielded comparable levels of accuracy: 75 % correct for *Some of the ___ are not ___*, and 72 % correct for *None of the ___ are ___*. In summary, the

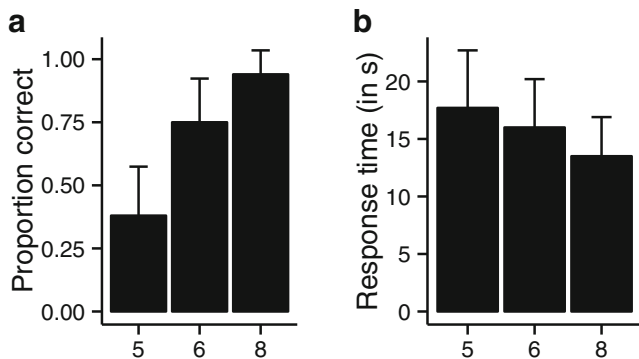


Fig. 2 Proportions of correct responses (a) and their median response times (b) in Experiment 1, as a function of the numbers of distinct entities consistent with the sets of assertions

results corroborated the prediction that the greater the number of possible entities consistent with a set of assertions, the easier it was for individuals to assess the set as being consistent: They were faster to respond, and they made fewer errors.

Experiment 2

The aim of this experiment was to test the model theory's prediction that individuals tend to build certain sorts of preferred model, which they rely on in evaluating consistency. The participants accordingly carried out the following sorts of problems:

- (14) Some of the actors are waiters.
Some of the waiters are singers.
Some of the actors are singers.
Is it possible for all three assertions to be true at the same time?

In this case, as we explained earlier, the third assertion holds in the preferred mental model of the first two assertions. But the next problem violates this constraint, as we also showed earlier:

- (15) Some of the actors are waiters.
Some of the waiters are singers.
None of the actors are singers.

In the experiment, we examined whether problems in which the third assertion was consistent with the preferred model of the first two assertions were easier than consistent problems in which the third assertion was not consistent with the preferred models. All of the problems were evaluated using the mReasoner software, which also computed the transformation of the preferred model into alternatives, where necessary, and the Levenshtein distance from the preferred to the alternative model.

Method

Participants We tested 25 logically naive participants (nine men, 16 women; $M = 26.0$ years) from the same population as before.

Design, materials, and procedure The participants acted as their own controls and evaluated sets of three assertions. All of the problems had the following figure: $a-b$, $b-c$, $a-c$. Six of them were consistent, because the third assertion held in the preferred model of the first two assertions (the preferred problems). Six of the problems were consistent but had a third assertion that did not hold in the preferred mental model of the first two assertions, and thus called for the construction of an

alternative model (the nonpreferred problems). Six of the problems were inconsistent, because the third assertion did not hold in any model of the first two assertions (see Table 3 in the Appendix). The first two assertions in each matched pair of preferred and nonpreferred problems were identical. In three of the six pairs, the preferred problems had affirmative third assertions and the nonpreferred problems had negative third assertions, but in three further pairs, these polarities were reversed (see Table 4 in the Appendix). The problems were presented in a different random order to each participant. The contents of the problems and the procedure were identical to those of the previous experiments.

Results

Overall, the six preferred problems yielded 95 % correct evaluations, but the six nonpreferred problems yielded only 65 % correct evaluations; this predicted difference was reliable (Wilcoxon test, $z = 2.10$, $p = .017$). Table 4 in the Appendix presents the percentages of correct evaluations for each of the individual problems. The difference was reliable in a by-materials analysis of the problems (Wilcoxon test, $z = 1.90$, $p = .029$). An analogous difference was found in the latencies to respond correctly, with medians of 13.16 s for preferred problems and 15.54 s for nonpreferred problems (Wilcoxon test, $z = 2.13$, $p = .016$). The first two assertions in each matched pair were identical, and the polarities of the third assertions were counterbalanced between affirmatives and negatives, and so this factor could not readily account for our results.

Experiment 3

In this experiment, we tested the prediction that the difficulty of consistent problems depends on the Levenshtein (1966) distance from the initial preferred model to the model satisfying the assertions. We illustrate the metric with three examples. Consider, first, a problem of this sort:

- (16) None of the a are b .
All of the b are c .
None of the a are c .

The first two assertions yield the preferred model (again we present only the different types):

- (17) $a \neg b$
 a
 $b \quad c$

The third assertion holds in this model, so no operations are needed to transform the model, and the Levenshtein distance equals 0. A second sort of problem has the same initial two assertions, but the third assertion is

(18) Some of the a are c .

This assertion does not hold in the preferred model above. An alternative model of the first two assertions satisfies the third assertion:

(19) $a \neg b \quad c$
 a
 $b \quad c$

A single operation of adding the property c to the first individual transforms the preferred model into this new model, so the Levenshtein distance equals 1 (and, for the full model, they can have several types; see the example above, which calls for two such operations, creating a Levenshtein distance of 2). A third sort of problem is based on the same two initial assertions, but the third assertion is

(20) All of the a are c .

In this case, a minimum of two edits, which each add c , are needed to transform the preferred model into one satisfying the third assertion:

(21) $a \neg b \quad c$
 $a \quad c$
 $b \quad c$

Hence, the Levenshtein distance to the complete preferred model equals 2. The experiment examined problems with distances of 0, 2, 3, and more. In contrast to Experiment 2, the numbers of entities consistent with the assertions were equally distributed across the different sorts of problem. For example, problem (16) above is consistent with four possible entities, problem (18) with five entities, and problem (20) with four entities. In most other problems, there is no difference in the numbers of entities satisfying the assertions, regardless of their Levenshtein distances. Overall, the theory predicts a general trend in the difficulty of constructing models of assertions: Problems in which the preferred model satisfies the assertions (distance of 0) should be easier than problems calling for model manipulations leading to higher Levenshtein distances, which should correlate inversely with the percentages of correct responses.

One other factor, which we have hitherto not mentioned, is relevant to this experiment: the role of Gricean implicatures. Grice (1989) argued that an assertion of the form *Some of the a are not b* suggests in normal conversation that *Some of the a are b* ; otherwise, a felicitous speaker would have asserted, *None of a are b* . Newstead (1995) found that such

implicatures did not play a reliable role in syllogistic reasoning, but we examined their potential effects on the evaluation of consistency.

Method

Participants We tested 26 participants (nine men, 17 women; $M = 31.7$ years) from the same population as before.

Design, materials, and procedure In this experiment, we examined all possible problems based on three assertions concerning three sets of individuals, in which the figure was held constant to $a-b$, $b-c$, $a-c$, and in which at least three of the third assertions were consistent with the first two assertions (see Table 5 in the Appendix). Of these problems, 44 were consistent and 16 were inconsistent. Within the consistent problems, according to mReasoner's classification, 17 problems called for zero operations to satisfy the third assertion, 12 called for two operations on the preferred model to satisfy the third assertion, and 15 called for three or more operations on the preferred model to satisfy the third assertion. The contents, their allocations to the problems, and the procedure were similar to those in the previous experiments. The problems were presented in a different random order to each participant.

Results and discussion

We excluded the data of five participants because they performed below chance. The percentages of correct evaluations of the consistent problems depended on the Levenshtein distance from the preferred model to the model

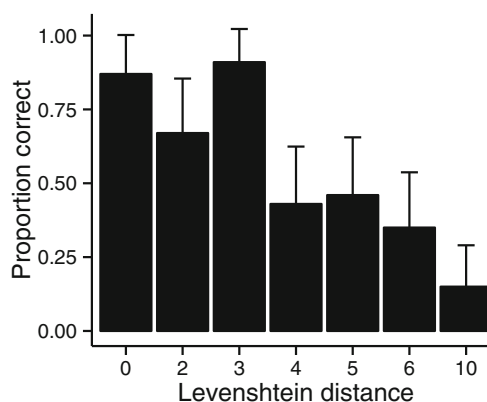


Fig. 3 Proportions of correct evaluations as a function of Levenshtein distance in Experiment 3

required to satisfy the assertions. In Fig. 3, we present the proportions of correct evaluations as a function of Levenshtein distance.

A highly significant inverse correlation occurred between the Levenshtein distances from the preferred model not incorporating Gricean implicatures (Grice, 1989) to the required model and the percentages of correct responses (Spearman's $\rho = -.57$, one-tailed $p < .00005$), and the correlation was even higher for mental models based on Gricean implicatures (Spearman's $\rho = -.68$, two-tailed $p < .000001$). This result suggests that implicatures may play more of a role in the evaluation of consistency than in reasoning (Newstead, 1995), but future research will need to make a direct comparison between the two tasks. One unexpected result was that the latencies of correct responses tended to correlate, albeit not significantly, negatively with the Levenshtein distance (Spearman's $\rho = -.25$, $p > .09$). Overall, the results show that the Levenshtein distance from a preferred model to one needed to satisfy a set of assertions predicts the difficulty of evaluating consistent assertions.

General discussion

We have proposed a theory of how naive individuals—those who know no logic—assess the consistency of sets of monadic assertions about the properties of sets of individuals, such as

- (22) Some of the actors are waiters.
 Some of the waiters are singers.
 Some of the actors are singers.
 Can all three assertions be true at the same time?

The theory postulates that they search for a model in which all of the assertions are true, starting with a preferred model, and if necessary looking for an alternative model. If they find such a model, they judge that the assertions could all be true at the same time. Our three experiments corroborated this theory. The experiments showed, first, that if an alternative model is necessary to satisfy the assertions, the greater the number of different sorts of individual that can occur in this alternative model, the easier it is to evaluate the assertions as being consistent (Exp. 1). Problems in which the preferred mental model of the initial assertions satisfies the third assertion are reliably easier than problems in which it is necessary to find an alternative model to satisfy the third assertion (Exp. 2). And the smaller the number of operations required to transform the preferred mental model into one that satisfies the third assertion—that is,

the closer the Levenshtein (1966) distance between these models—the easier it is to evaluate the assertions as being consistent (Exp. 3).

In a subsequent study, we have eliminated a potential confound in the experiments. Only affirmative assertions have a single sort of individual that satisfies them in a preferred model, whereas negative assertions call for more than one sort of individual in a preferred model. However, an additional study showed that the number of negated assertions had no reliable effect on accuracy or latency, and that the most difficult problem contained only one negative assertion. If the distance from a preferred model to a model satisfying a set of assertions is large, the latter model does not necessarily contain a greater number of entities. And as the theory predicts in such cases, problems that call for only one operation to yield an alternative to the preferred model tend to be easier than those that call for multiple operations. It would be interesting to pit number of entities and numbers of operations against one another, but a systematic comparison would be impossible in the domain of monadic assertions.

At present, the model theory provides the only explanation of how human reasoners evaluate consistency. However, a close relation exists between deductive reasoning and the evaluation of consistency. The relation is obvious in logic, because one method to establish that a deduction is valid is to show that the negation of its conclusion together with the premises yields an inconsistent set of assertions (Jeffrey, 1981). In view of this close relation, other theories of reasoning might be extended to cope with consistency. In particular, a theory of syllogistic reasoning might be extended to explain our results. Such theories include those based on formal rules of inference akin to those in logic (e.g., Rips, 1994) and one based on probabilistic heuristics (e.g., Oaksford & Chater, 2007). In fact, no existing theory of syllogistic reasoning gives a good enough account of the experimental results with syllogisms (see a recent meta-analysis in Khemlani & Johnson-Laird, 2012b). We are therefore in a tricky position. On the one hand, we cannot discount the possibility that adaptations of these theories might explain the evaluation of consistency. Such a theory, for instance, might combine mental models with probabilistic heuristics (see Geiger & Oberauer, 2010; Oaksford & Chater, 2010). On the other hand, we have no explicit proposals or predictions from any alternative theories of consistency. It is easy enough to see how, in principle, a theory based on, say, Euler circles (e.g., that of Erickson, 1974) could explain the evaluation of consistency, but the real challenge to such a theory is to predict the differences in difficulty from one sort of problem to another. In particular, the theory needs to account for the present results

concerning the number of alternative entities consistent with the premises, preferred versus alternative models, and the Levenshtein distance between them. Instead of our trying to second-guess extensions of these alternative theories, we have opted to point out general challenges to such extensions.

The problem for theories based on formal rules is illustrated by example (22) above. It is easy for participants to evaluate its assertions as being consistent, but the corresponding syllogistic deduction is not valid:

- (23) Some of the actors are waiters.
Some of the waiters are singers.
Therefore, some of the actors are singers.

Thus, the validity of a corresponding deduction is not a universal test of consistency: The latter is a wider notion than the former. One potential method for using deductive proofs to assess consistency depends on four steps: (1) Choose any assertion from the set of assertions to be evaluated, such as the final one; (2) form its negation; (3) try to prove that this negative assertion follows from the remaining assertions in the set; and (4) if such a proof exists, then the set of assertions is inconsistent, but otherwise it is consistent (see, e.g., Johnson-Laird et al., 2000). This procedure is correct, but implausible psychologically. Step 1 is easy, but Step 2 is fraught with difficulty, especially in the domain of monadic assertions. For example, many naive individuals consider the negation of *some of the actors are waiters* to be *some of the actors are not waiters*, but these two assertions are not even contraries (see, e.g., Khemlani, Orenes, & Johnson-Laird, 2013). The correct negation is *none of the actors is a waiter*, because the two assertions need to contradict one another. Indeed, naive reasoners are most unlikely even to discover the four-step procedure, and evidence counts against it. The procedure predicts that a correct evaluation of a set of assertions as being inconsistent should be faster than a correct evaluation of a set of assertions as being consistent, because the former depends on a single proof, whereas the latter calls for an exhaustive search of all possible proofs and a failure to find one. The evidence, at least for assertions based on sentential connectives, such as “if” and “or,” refutes this prediction (Johnson-Laird et al., 2000). In fact, proponents of psychological theories of reasoning based on formal rules of inference, such as Rips (1994), have never advocated the four-step procedure.

The present theory extends to problems beyond those in our experiments, and it deals with three other sorts of quantified assertions. The first sort refers to monadic assertions expressible with the quantifiers of standard logic (the first-order predicate calculus) but using numerical quantifiers, such as *at least three actors* and *no more than four waiters*. The intensions of assertions ensure

that the numbers of tokens in mental models represent the numbers in assertions, and large numbers, which are beyond the human capacity for iconic representations, are used to tag models with explicit numerals.

The second sort of assertion to which the theory applies is monadic but concerns proportions, such as *more than half of the actors are waiters*. These quantifiers cannot be defined using the quantifiers of the first-order predicate calculus (Barwise & Cooper, 1981). They call for a logic in which variables can also range over sets of entities—that is, the second-order predicate calculus. Mental models, however, represent these quantifiers as relations between sets (Johnson-Laird, 1983, p. 140; Khemlani & Johnson-Laird, 2012a). Consider, for example, this set of assertions:

- (24) More than half of the artists are bakers.
More than half of the bakers are chemists.
More than half of the artists are chemists.

The first two assertions elicit the following sort of model, and here we show the numbers of entities because their proportions are crucial:

- (25) artist baker chemist
 artist baker chemist
 artist baker chemist
 artist baker
 artist

The third assertion holds in this model, so the set is consistent. In contrast, consider the problem

- (26) More than half of the artists are bakers.
More than half of the bakers are chemists.
All of the artists are chemists.

The first two assertions elicit the same model (25) as before, so the assertions seem inconsistent, because the third assertion does not hold in this model. But, an alternative model satisfying all three assertions does exist, and it can be constructed by adding tokens representing chemists to each token representing an artist. Reasoners who succeed in constructing this model should respond correctly that the set of assertions is consistent. These representations have been implemented in the mReasoner program.

The third sort of assertion to which the theory applies contain multiple quantifiers, such as *Some of the actors are in the same room as all of the waiters*. Some of these assertions are expressible in the quantifiers of first-order logic, but others are not. The extension to the case of multiple quantifiers is illustrated by this example (cf. Johnson-Laird, Byrne, & Tabossi, 1989):

- (27) None of the Avon letters is in the same place as any of the Bethel letters.
All of the Bethel letters are in the same place as all of the Cam letters.

None of the Avon letters is in the same place as any of the Cam letters.

The operations for building models from quantifiers consist of a loop in which each cycle adds a token of the corresponding set to the model—a process that applies even to monadic assertions. With multiple quantifiers in a sentence, a loop for one quantifier may be embedded within a loop for another quantifier, and the order of embedding depends largely, but not entirely, on their surface order in the sentence. (The order of embedding in fact corresponds to the logical “scope” of quantifiers.) For example, the first assertion in (27) adds a token representing an Avon letter to a model, and then carries out the loop for the second quantifier, establishing the appropriate spatial relations between the tokens. It therefore yields the following partial model:

(28) | Avon | Bethel Bethel Bethel Bethel |,

where the vertical lines demarcate separate places, and “Avon” and “Bethel” denote the different sorts of letters. The loop for the first quantifier now adds the next token, and the loop for the second quantifier is already satisfied. The process continues until it yields the model

(29) | Avon Avon Avon | Bethel Bethel Bethel Bethel |.

The second assertion in (27) updates the model in a similar way:

(30) | Avon Avon Avon | Bethel Bethel Bethel Bethel Cam Cam Cam |

The third assertion holds in this model, so the set of assertions is consistent. Other sets of assertions can have preferred models that do not satisfy a third assertion, but a search for an alternative model yields a model satisfying all of the assertions.

The search operations are once again identical to those used to find models to refute putative conclusions in deducing a conclusion from premises, and they are outlined in Johnson-Laird et al. (1989). The intensions of the assertions contain various parameters that govern the building of models and the search for alternative models. One parameter is a default value for the number of tokens in a model representing the cardinality of a set, such as *a* in the assertions above. Another parameter is a default value for the number of these tokens that are in the set, *b*. These parameters allow for numerical quantifiers, such as *at least three a are b*, and for proportional quantifiers, such as *more than half of the a are b*. The values of these parameters can be changed in the light of subsequent information, if necessary, provided that the new values do not violate the constraints represented in the intension.

The model theory offers a straightforward explanation of both valid deduction (see, e.g., Johnson-Laird, 2006, for a review) and the evaluation of consistency. Individuals who can construct a mental model of a set of assertions evaluate the set as being consistent; otherwise, they evaluate the assertions as being inconsistent. The model theory, as we have shown here, applies not just to the monadic assertions that our experiments have examined, but also to those quantifiers, such as “most of the artists,” that cannot be expressed using the quantifiers of first-order logic, and also to multiply-quantified assertions, such as *none of the Avon letters is in the same place as any of the Bury letters*. Our empirical studies were of the simplest quantifiers, so these other sorts of quantifiers stand in need of empirical examination, too.

In conclusion, reasoning about the consistency of quantified assertions can be explained in terms of models of different sorts of individuals. When a model contains all possible sorts of individuals, consistency is easier to infer than when it contains only a restricted set of individuals. When an assertion is satisfied by a preferred model of the initial assertions, consistency is easier to infer than when constructing a new model is necessary, and the difficulty of this construction depends in turn on the distance of the new model from the preferred model.

Author note This research was supported in part by Grant No. SES 0844851 to the third author from the National Science Foundation, to study deductive and probabilistic reasoning; in part by a National Research Council Research Associateship to the second author; and in part by a grant to the first author from the Deutsche Forschungsgemeinschaft (German Research Foundation) as part of project SPP 1516. The authors are grateful to Hua Gao, Catrinel Haight, Max Lotstein, Tobias Sonntag, and two anonymous reviewers for their help and advice.

Appendix

Table 3 The grammatical forms of the inconsistent problems used in Experiments 1 and 2, and included in Experiment 3

Assertions	Assertions
All of the <i>a</i> are <i>b</i>	Some of the <i>a</i> are <i>b</i>
All of the <i>b</i> are <i>c</i>	None of the <i>b</i> are <i>c</i>
None of the <i>a</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>
All of the <i>a</i> are <i>b</i>	All of the <i>a</i> are <i>b</i>
None of the <i>b</i> are <i>c</i>	All of the <i>b</i> are <i>c</i>
All of the <i>a</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>
Some of the <i>a</i> are <i>b</i>	All of the <i>a</i> are <i>b</i>
All of the <i>b</i> are <i>c</i>	None of the <i>b</i> are <i>c</i>
None of the <i>a</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>

Table 4 The matched preferred and non-preferred forms of consistent problems, the different sorts of individual in their mental models, and the percentages of correct evaluations for each of them in Experiment 2

Preferred Model Problems	Preferred Mental Model				Nonpreferred model Problems	Alternative Model			
All of the <i>a</i> are <i>b</i>	<i>a</i>	<i>b</i>	<i>c</i>	96	All of the <i>a</i> are <i>b</i>	<i>a</i>	<i>b</i>		52
Some of the <i>b</i> are <i>c</i>	<i>a</i>	<i>b</i>			Some of the <i>b</i> are <i>c</i>		<i>b</i>	<i>c</i>	
Some of the <i>a</i> are <i>c</i>					None of the <i>a</i> are <i>c</i>				
All of the <i>a</i> are <i>b</i>	<i>a</i>	<i>b</i>	$\neg c$	96	All of the <i>a</i> are <i>b</i>	<i>a</i>	<i>b</i>	$\neg c$	92
Some of the <i>b</i> are not <i>c</i> *	<i>a</i>	<i>b</i>			Some of the <i>b</i> are not <i>c</i> *	<i>a</i>	<i>b</i>	<i>c</i>	
Some of the <i>a</i> are not <i>c</i>			<i>c</i>		Some of the <i>a</i> are <i>c</i>			<i>c</i>	
Some of the <i>a</i> are <i>b</i>	<i>a</i>	<i>b</i>	<i>c</i>	100	Some of the <i>a</i> are <i>b</i>	<i>a</i>	<i>b</i>		72
Some of the <i>b</i> are <i>c</i>	<i>a</i>				Some of the <i>b</i> are <i>c</i>		<i>b</i>	<i>c</i>	
Some of the <i>a</i> are <i>c</i>					None of the <i>a</i> are <i>c</i>		<i>a</i>		
None of the <i>a</i> are <i>b</i>	<i>a</i>	$\neg b$		84	None of the <i>a</i> are <i>b</i>	<i>a</i>	$\neg b$	<i>c</i>	60
Some of the <i>b</i> are <i>c</i>		<i>b</i>	<i>c</i>		Some of the <i>b</i> are <i>c</i>		<i>b</i>	<i>c</i>	
None of the <i>a</i> are <i>c</i>					All of the <i>a</i> are <i>c</i>				
Some of the <i>a</i> are not <i>b</i>	<i>a</i>	$\neg b$	<i>c</i>	100	Some of the <i>a</i> are not <i>b</i> *	<i>a</i>	$\neg b$		40
All of the <i>b</i> are <i>c</i>	<i>a</i>		<i>c</i>		All of the <i>b</i> are <i>c</i>	<i>a</i>			
Some of the <i>a</i> are <i>c</i>		<i>b</i>	$\neg c$		None of the <i>a</i> are <i>c</i>		<i>b</i>	<i>c</i>	
Some of the <i>a</i> are not <i>b</i>	<i>a</i>	$\neg b$		96	Some of the <i>a</i> are not <i>b</i>	<i>a</i>	$\neg b$	<i>c</i>	72
Some of the <i>b</i> are not <i>c</i> *	<i>a</i>				Some of the <i>b</i> are not <i>c</i>	<i>a</i>		<i>c</i>	
Some of the <i>a</i> are not <i>c</i>		<i>b</i>	$\neg c$		All of the <i>a</i> are <i>c</i>		<i>b</i>	$\neg c$	
			<i>c</i>					<i>c</i>	

The models were generated by mReasoner; we present only the types. * The model can be interpreted with and without the Gricean implicature (leading sometimes to slightly different models; see the text)

Table 5 The forms of the consistent problems, the Levenshtein distance and the percentages of correct responses to each sort of problem in Experiment 3

Levenshtein Distance	Premise 1	Premise 2	Premise 3	Percentage Correct	Latencies in ms for Correct Answers (Medians)
0	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	81	6,808
0	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	95	8,252
0	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	Some of the <i>a</i> are not <i>c</i>	95	8,507
0	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	100	5,908
0	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	95	5,859
0	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	None of the <i>a</i> are <i>c</i>	57	11,384
0	None of the <i>a</i> are <i>b</i>	All of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	81	8,941
0	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	86	7,880
0	None of the <i>a</i> are <i>b</i>	None of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	86	8,457
0	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	None of the <i>a</i> are <i>c</i>	81	6,203
0	Some of the <i>a</i> are not <i>b</i>	None of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	57	11,384
0	Some of the <i>a</i> are not <i>b</i>	All of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	100	10,509
0	Some of the <i>a</i> are not <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	95	9,298
0	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	Some of the <i>a</i> are <i>c</i>	90	6,170
0	Some of the <i>a</i> are not <i>b</i>	None of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	95	5,946
0	Some of the <i>a</i> are not <i>b</i>	All of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	81	7,426
0	Some of the <i>a</i> are not <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	100	5,922

Table 5 (continued)

Levenshtein Distance	Premise 1	Premise 2	Premise 3	Percentage Correct	Latencies in ms for Correct Answers (Medians)
2	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	None of the <i>a</i> are <i>c</i>	38	13,486
2	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	Some of the <i>a</i> are not <i>c</i>	95	5,962
2	Some of the <i>a</i> are not <i>b</i>	None of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	76	8,078
2	None of the <i>a</i> are <i>b</i>	All of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	76	8,941
2	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	95	10,761
2	None of the <i>a</i> are <i>b</i>	None of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are not <i>c</i>	90	8,030
2	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	Some of the <i>a</i> are not <i>c</i>	100	8,143
2	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	33	6,259
2	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	29	8,127
2	None of the <i>a</i> are <i>b</i>	All of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	33	8,168
2	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	67	7,669
2	None of the <i>a</i> are <i>b</i>	None of the <i>b</i> are <i>c</i>	Some of the <i>a</i> are <i>c</i>	76	9,721
3	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	Some of the <i>a</i> are <i>c</i>	95	3,985
3	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	Some of the <i>a</i> are <i>c</i>	86	8,327
4	Some of the <i>a</i> are not <i>b</i>	All of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	33	8,168
4	None of the <i>a</i> are <i>b</i>	All of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	24	7,248
4	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	57	8,906
4	None of the <i>a</i> are <i>b</i>	None of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	71	7,066
4	Some of the <i>a</i> are not <i>b</i>	All of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	29	7,899
5	Some of the <i>a</i> are not <i>b</i>	Some of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	38	9,577
5	None of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	All of the <i>a</i> are <i>c</i>	62	9,129
5	Some of the <i>a</i> are not <i>b</i>	Some of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	38	9,577
6	Some of the <i>a</i> are not <i>b</i>	None of the <i>b</i> are <i>c</i>	All of the <i>a</i> are <i>c</i>	19	8,939
6	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	33	8,153
6	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are <i>c</i>	None of the <i>a</i> are <i>c</i>	52	12,625
10	Some of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	All of the <i>a</i> are <i>c</i>	19	5,925
10	All of the <i>a</i> are <i>b</i>	Some of the <i>b</i> are not <i>c</i>	All of the <i>a</i> are <i>c</i>	10	7,832

References

- Barnes, J. (1999). *Complete works of Aristotle (2 vols.)*. Princeton, NJ: Princeton University Press.
- Barwise, J. (1993). Everyday reasoning and logical inference. *The Behavioral and Brain Sciences*, 16, 337–338.
- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4, 159–219.
- Bauer, M., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*, 4, 372–378.
- Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, 23, 247–303.
- Edghill, E. M. (1928). Categoriae and De Interpretatione. In W. D. Ross (Ed.), *The works of Aristotle (24, b5)*. Oxford, UK: Oxford University Press, Clarendon Press.
- Erickson, J. R. (1974). A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (Ed.), *Loyola symposium on cognition (vol. 2)* (pp. 305–330). Hillsdale, NJ: Erlbaum.
- Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. New York, NY: W. H. Freeman.
- Geiger, S. M., & Oberauer, K. (2010). Towards a reconciliation of mental model theory and probabilistic theories of conditionals. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 289–307). Oxford: Oxford University Press.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, 35, 2075–2087.
- Jeffrey, R. (1981). *Formal logic: Its scope and limits* (2nd ed.). New York, NY: McGraw-Hill.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N. (2006). *How we reason*. New York, NY: Oxford University Press.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduct ion*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646–678.
- Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, 96, 658–673.

- Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science*, *288*, 531–532.
- Khemlani, S., & Johnson-Laird, P. N. (2012a). Processes of inference. *Argument and Computation*, *4*, 1–20.
- Khemlani, S., & Johnson-Laird, P. N. (2012b). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*, 427–457.
- Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2013). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, *24*, 541–559.
- Knauff, M., Rauh, R., & Schlieder, C. (1995). Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In *Proceedings of the 17th annual conference of the cognitive science society* (pp. 200–205). Mahwah, NJ: Erlbaum.
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics – Doklady*, *10*, 707–710. Originally published in Russian in 1965, in *Doklady Akademii Nauk SSSR*, *163*, 845–848.
- Newstead, S. E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language*, *34*, 644–664.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.
- Oaksford, M., & Chater, N. (2010). Conditional inference and constraint satisfaction: Reconciling mental models and the probabilistic approach. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 309–333). Oxford, UK: Oxford University Press.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments using Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411–419.
- Peirce, C. S. (1931–1958). In C. Hartshorne, P. Weiss, & A. Burks (Eds.) *Collected papers of Charles Sanders Peirce* (8 vols.). Cambridge, MA: Harvard University Press.
- Ragni, M. (2003). An arrangement calculus, its complexity and algorithmic properties. In A. Günter, R. Kruse, & B. Neumann (Eds.), *Advances in Artificial Intelligence, 26th Annual German Conference on AI* (pp. 580–590). Berlin, Germany: Springer.
- Ragni, M., Fangmeier, T., Webber, L., & Knauff, M. (2007a). Preferred mental models: How and why they are so important in human reasoning with spatial relations. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, & T. Barkowsky (Eds.), *Spatial cognition V: Reasoning, action, interaction* (pp. 175–190). Berlin, Germany: Springer.
- Ragni, M., Knauff, M., & Nebel, B. (2005). A computational model for spatial reasoning with mental models. In B. Bara, B. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1064–1070). Mahwah, NJ: Erlbaum.
- Ragni, M., Tsenden, B., & Knauff, M. (2007b). Cross-cultural similarities in topological reasoning. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), *Cosit (vol. 4736)* (pp. 32–46). Berlin, Germany: Springer.
- Rauh, R., Hagen, C., Knauff, M., Kuß, T., Schlieder, C., & Strube, G. (2005). Preferred and alternative mental models in spatial reasoning. *Spatial Cognition and Computation*, *5*, 239–269.
- Rauh, R., Schlieder, C., & Knauff, M. (1997). Präferierte mentale Modelle beim räumlich-relationalen Schließen: Empirie und kognitive Modellierung. *Kognitionswissenschaft*, *6*, 21–34.
- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.
- Schleipen, S., Ragni, M., & Fangmeier, T. (2007). Negation in spatial reasoning: A computational approach. In J. Hertzberg, M. Beetz, & R. Englert (Eds.), *Advances in Artificial Intelligence—30th Annual German Conference on AI* (pp. 175–189). Berlin, Germany: Springer.
- Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, *4*, 13–47.
- Vandierendonck, A., Dierckx, V., & De Vooght, G. (2004). Mental model construction in linear reasoning: Evidence for the construction of initial annotated models. *Quarterly Journal of Experimental Psychology*, *57A*, 1369–1391.