# A mental model theory of set membership

Sangeet Khemlani[1], Max Lotstein[2], and Phil Johnson-Laird[3,4]
sangeet.khemlani@nrl.navy.mil, mlotstein@gmail.com, phil@princeton.edu

[1]US Naval Research Laboratory, Washington, DC 20375 USA
[2]University of Freiburg, Freiburg, Germany
[3]Princeton University, Princeton NJ 08540 USA
[4]New York University, New York, NY 10003, USA

**Abstract**

Assertions of set membership, such as *Amy is an artist*, should not be confused with those of set inclusion, such as *All artists are bohemians*. Membership is not a transitive relation, whereas inclusion is. Cognitive scientists have neglected the topic, and so we developed a theory of inferences yielding conclusions about membership, e.g., *Amy is a bohemian*, and about non-membership, *Abbie is not an artist*. The theory is implemented in a computer program, *mReasoner,* and it is based on mental models. The theory predicts that inferences that depend on a search for alternative models should be more difficult than those that do not. An experiment corroborated this prediction. The program contains a parameter, σ, which determines the probability of searching for alternative models. A search showed that its optimal value of .58 yielded a simulation that matched the participant's accuracy in making inferences. We discuss the results as a step towards a unified theory of reasoning about sets.

**Keywords:** quantifiers, reasoning, sets, syllogisms.

## Introduction

Quantifiers raise problems for linguists in their syntax and semantics (e.g., Peters & Westerståhl, 2006; Steedman, 2012). And they raise problems for cognitive scientists in their mental representation and roles in inference (e.g., Johnson-Laird, 2006; Oaksford & Chater, 2007; Rips, 1994). Our goal is to elucidate quantifiers by considering a neglected topic: assertions of set membership. Consider, for example, these three assertions:

1. Viv is a judge.
2. Judges are lawyers.
3. Judges are appointed in different ways.

Assertion (1) states that an individual is a *member* of a set, and assertion (2) states that one set is *included* in another set. Assertion (3) is about a set, but it states, not that it is included in another set, but that it is a member of another set, i.e., the set of judges is a member of the set of those who are appointed in different ways. The difference matters because membership is not a transitive relation. Hence, the following inference is not valid:

4. Viv is a judge.
   Judges are appointed in different ways.
   Therefore, Viv is appointed in different ways.

In contrast, this inference is valid:

5. Viv is a judge.
   Judges are lawyers.
   Therefore, Viv is a lawyer.

In formal logic, the second premise is equivalent to: anyone who is a judge is a lawyer, i.e., inclusion is defined in terms of membership. The distinction between inferences (4) and (5) is therefore subtle, and psychological theories need to recognize the difference between them.

A powerful way in logic to represent the meaning of quantifiers, such as "all judges", is as sets of sets. This method was popularized by Montague (see the accounts in, e.g., Johnson-Laird, 1983; Partee, 1975; Peters & Westerståhl, 2006). But, an alternative representation treats quantified assertions as stating relations between sets (Boole, 1854, Ch. XV). More recently, psychologists have had the same idea (e.g., Ceraso & Provitera, 1971; Geurts, 2003; Johnson-Laird, 1970; Politzer, van der Henst, Luche, & Noveck, 2006). On this account, the assertion *all judges are lawyers* means that the set of judges is included in the set of lawyers. Likewise, the assertion *no judges are inmates* means that the intersection of the set of judges and the set of inmates is empty. The advantage of this treatment is that it readily extends to quantifiers that cannot be captured in standard logical accounts (such as Rips, 1994). The assertion *most judges are men*, contains a quantifier "most judges" that cannot be defined using the quantifiers of first-order predicate calculus (Barwise & Cooper, 1991). Its relational meaning is simple: the cardinality of the intersection of the set of judges and the set of males is greater than the cardinality of the set of judges that are not males (see Cohen & Nagel, 1934).

So, how do naïve individuals reason about set membership? The aim of the present paper is to answer this question. Our answer is based on the theory of mental models. We accordingly begin with an outline of the theory from which we derive one principal prediction about such inferences. We report an experiment that corroborates this prediction. We use a computer program, *mReasoner*, to simulate performance, and show that the simulation provides a satisfactory fit with the experimental results. Finally, we draw some general conclusions about the psychology of set membership.

## The model theory of set membership

The mental model theory – the "model" theory, for short – applies to reasoning of many sorts, including reasoning based on quantifiers and on sentential connectives, such as *if*, *or*, and *and* (Johnson-Laird & Byrne, 1991) and reasoning about the subjective probability of unique events (Khemlani, Lotstein, & Johnson-Laird, 2012). Three main principles underlie the theory (Johnson-Laird, 2006). First, individuals use a representation of the meanings of assertions and their knowledge to construct mental models of each distinct sort of possibility to which the assertions refer. Second, mental models are iconic insofar as that is possible, i.e., the structure of a model corresponds to the structure of what it represents (see Peirce, 1931-1958, especially Vol. 4). Hence, a set is represented by a set of mental tokens in a model. Third, more models mean more work: reasoners are likely to rely on their initial model for most inferences, and if a particular inference requires them to consider alternative models, it will be difficult. Reasoners accordingly use the meaning of premises – representations of their *intensions* – and their knowledge to construct mental models – representations of their *extensions* – of premises. Depending on whether a putative conclusion holds in all, most, or some of these models, they draw a conclusion that it is necessary, probable, or possible.

According to the model theory, logical properties such as transitivity are emergent properties from models (e.g., Byrne & Johnson-Laird, 1989; Goodwin & Johnson-Laird, 2005; Huttenlocher, 1968). Consider these contrasting examples:

6. All the players are tall.
7. All the players are equal in height.
8. All the players are tall to varying degrees up to the tallest.

Given the further premise:

9. Ann is one of the players.

It follows from (6) that *Ann is tall*. But, what follows from (7) is not that *Ann is equal in height*, which doesn't make sense, but rather that:

10. Ann is equal in height to the other players.

Likewise, what follows from (8) is not that Ann is tall to varying degrees, but rather that:

11. Ann is tall and possibly the tallest of the players.

The challenge to theories based on formal rules of inference or probabilistic considerations is to account for these inferences. In contrast, they emerge from mental models of the quantified premises. For example, a model of (8) represents iconically the set of players as varying in height up to the tallest. When (9) is used to update the model, *Ann* is added to a representation of an individual player. This player could be the one that is tallest, but needn't be. Hence, only the modal conclusion in (11) follows. This sort of machinery seems to be a prerequisite either for establishing the logical form of premises (pace Rips, 1994) or their appropriate probabilistic analysis (pace Oaksford & Chater, 2007).

Our investigation begins with two sorts of inference that are fundamental. The first sort concerns set membership, as in the following premises:

12. Ansel is an artist.
    All artists are benefactors.

They elicit the following sort of model in which each row represents a different individual:

```
Ansel    artist   benefactor
         artist   benefactor
         artist   benefactor
```

It yields the valid conclusion: *Ansel is a benefactor*.

The second sort of inference concerns non-membership of a set, such as:

13. Igor is not a benefactor.
    All artists are benefactors.

They elicit the model:

```
Igor            ¬ benefactor
         artist   benefactor
         artist   benefactor
         artist   benefactor
```

where '¬' denotes negation, i.e., the individual is not a benefactor. The corresponding conclusion, *Igor is not an artist*, is also valid.

According to the theory, inferences such as (12) and (13) should be easy, because the correct response can be inferred from the initial mental model of the premises. The theory distinguishes such one-model inferences from multiple-model inferences such as:

14. Faye is not an artist.
    All artists are benefactors.

The premises yield the following model:

```
Faye   ¬ artist
         artist   benefactor
         artist   benefactor
         artist   benefactor
```

This model yields the conclusion that *Faye is not a benefactor*. However, the conclusion is invalid, because the quantified assertion allows that benefactors may not be artists. Reasoners can modify their initial model (Bucciarelli

& Johnson-Laird, 1999; De Neys, Schaeken, & d'Ydewalle, 2003; Johnson-Laird & Hasson, 2003; Neth & Johnson-Laird, 1999) to yield a counterexample to the previous conclusion:

```
Faye    ¬ artist   benefactor
          artist   benefactor
          artist   benefactor
          artist   benefactor
```

Such a modification calls for the additional resources of system 2, and so many reasoners should rely on their initial model. The theory therefore predicts that one-model inferences should be easier than multiple-model inferences. We now describe a study designed to test this hypothesis.

## Experiment

The experiment tested the prediction of the model theory of set-membership inferences. A typical trial in the experiment was:

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

The experiment examined eight different sorts of inference. The model theory classified four of the inferences as one-model inferences and the other four as multiple-model inferences.

## Method

*Participants.* Twenty-one participants completed the study on Mechanical Turk, an online platform hosted by Amazon.com that distributes experimental tasks to volunteers. Participants received monetary compensation for taking part in the study, none of them reported having had any training in logic, and they were all on their account native speakers of English.

*Design and materials.* Each inference contained a premise about an individual (e.g., *Rachel is a baker*) and a quantified premise about all the members of a set (e.g., *All of the soldiers are bakers*). The study manipulated three variables: 1. whether the premise about the individual asserted membership or non-membership of a set; 2. whether the quantified premise asserted inclusion ("All") or non-inclusion ("None") of one set within another; and 3. whether the set referred to in the premise about the individual matched the first or the second term in the quantified premise (see Table 1 below for the 8 different sorts of inference). The third variable yields two figures, as follows:

| *Figure 1* | *Figure 2* |
|---|---|
| X is an A. | X is an A. |
| All of the *As are Bs*. | All of the *Bs are As*. |

As Table 1 below shows, half of the inferences were one-model inferences, and half of the inferences were multiple-model inferences. The exploration of all permutations of the premises yielded an inevitable confound, such that all of the multiple-model problems were also those for which there existed no valid, non-trivial conclusion. The eight inferences were presented twice with different contents, and so participants carried out 16 inferences in total. The order of the premises was counterbalanced so that the membership premise was presented first for half the inferences and second for the other half. The contents were based on common names and common nouns referring to vocations. We devised a list of 16 pairs of such vocations, which we assigned at random to the inferences. The inferences were presented to each participant in a different random order.

*Procedure.* The study was administered using an interface written in Ajax. The instructions stated that the task was to decide "what conclusions, if any, must also be true" given the truth of a pair of premises presented on the screen. On each trial, participants read the premises, and, when ready, they pressed a button marked "Next", which replaced the premises with a question, "What, if anything, follows?" They responded by typing their answer out on a text box provided on the screen and then clicked a button to advance to the next inference. An independent coder classified participants' typed responses as falling into one of three categories (given the schematic described earlier):

i. x is a B.
ii. x is not a B.
iii. No valid conclusion.

| **Set membership premise** | **Monadic premise** | **Figure** | **Predicted response to "What follows?"** | **Correct response** | **Inference type** | **Correct (%)** |
|---|---|---|---|---|---|---|
| x is an A | All of the A are B | 1 | x is a B | x is a B | One-model | 95% |
| x is an A | All of the B are A | 2 | x is a B | No valid conclusion | Multiple-model | 47% |
| x is an A | None of the A are B | 1 | x is not a B | x is not a B | One-model | 93% |
| x is an A | None of the B are A | 2 | x is not a B | x is not a B | One-model | 88% |
| x is not an A | All of the A are B | 1 | x is not a B | No valid conclusion | Multiple-model | 50% |
| x is not an A | All of the B are A | 2 | x is not a B | x is not a B | One-model | 90% |
| x is not an A | None of the A are B | 1 | x is not a B | No valid conclusion | Multiple-model | 83% |
| x is not an A | None of the B are A | 2 | x is not a B | No valid conclusion | Multiple-model | 66% |

**Table 1.** The proportion of correct responses to the eight different inferences in the experiment.

## Results and discussion

The vast majority (99%) of participants' responses fell into the three categories of responses as described in the previous section. The remaining responses were dropped from subsequent analyses. A small portion of the responses (9.5%) contained qualifications, e.g., a modal operator of the form, *x is possibly a* B. These responses occurred only in the case of multiple model inferences. The model theory accordingly accounts for them: individuals construct an initial model that supports a conclusion such as *x is a B*; they envisage an alternative model, and it refutes the conclusion; and so they weaken the conclusion to a modal claim, *x is possibly a B*. An analogous phenomenon occurs in syllogistic reasoning (e.g., Bucciarelli & Johnson-Laird, 1999). We note that theories that do not rely on representations of possibilities fail to offer any ready account of this phenomenon. Nevertheless, we omitted modal responses from our analyses because it is unclear whether they should count as easy or difficult responses.

Table 1 presents the percentage of correct responses for the eight different inferences. The experiment corroborated the theory's main prediction: one-model inferences were reliably easier than multiple-model inferences (92% vs. 51%, Wilcoxon test, $z = 3.21$, $p = .001$, Cliff's $\delta = .56$). Inferences from the quantifier *none of the X* were easier than those from *all of the X* (83% vs. 72%, Wilcoxon test, $z = 2.17$, $p = .03$, Cliff's $\delta = .33$). The result may reflect the indeterminacy of *all of the X are Y*, which leaves open whether or not all of the Y are X. No such indeterminacy occurs with *None of the X are Y*, which implies that *None of the Y are X*. Finally, we observed a marginal effect of figure, where figure 1 inferences were easier than figure 2 inferences (73% vs. 68%, Wilcoxon test, $z = 1.65$, $p = .10$, Cliff's $\delta = .20$). Analogous effects of figure occur in syllogistic reasoning (Khemlani & Johnson-Laird, 2012).

Of the three factors manipulated in the experiment (one vs. multiple models; *all* vs. *none*; and figure), which of them best predicts the difficulty of the inferences? To answer this question, we fit the data to a generalized mixed-effects model with a binomial error distribution and a logit link function using the *lme4* package (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2013). The model took into account the three fixed effects described above and one random effect, i.e., the variance in the participants' accuracy. The only significant predictor of performance was whether an inference required one or multiple models ($b = 3.31$, $SE = .42$, $p < .0001$). These results again corroborate the model theory's main prediction.

In general, the experiment shows that naive individuals can make valid set membership deductions, but that they often err in inferences that depend on multiple models.

## Simulating set membership inferences

We sought to simulate the results of the experiment by generating synthetic data from *mReasoner* v0.9, a unified computational implementation of the mental model theory of reasoning (Khemlani & Johnson-Laird, 2013), and matching it against the dataset from the experiment. The program implements three general systems:

a) A linguistic system for parsing premises and building up intensional representations used in model building. This system's purpose is to map an assertion's syntax to an underlying semantics (the intension).
b) An intuitive heuristic system (System 1) for building an initial mental model and for drawing inferences from it.
c) A deliberative system (System 2) that interrogates the initial model to search for alternative models. This system can manipulate and update the representations created in System 1, and it can modify conclusions, but it too can fall prey to systematic errors (Johnson-Laird & Savary, 1999; Khemlani & Johnson-Laird, 2009).

System 1 does not have access to working memory, whereas system 2 does. As a result, system 1 is faster and more prone to err than system 2.

To simulate the non-determinism inherent in human reasoning, *mReasoner* is equipped with three parameters that govern how models are built and how the two inferential systems are engaged (Khemlani, Trafton, & Johnson-Laird, 2013). The first parameter stochastically varies the size of a mental model, i.e., the number of individuals it contains. This parameter can have no effect on the inferences in the present experiment. The second parameter varies the properties of the individuals in models, e.g., in the case of *All of the A are B*, it affects whether or not B's that are not A's occur in a model. But, this variable can also have no effect on the present inferences (see Table 1). Hence, only the third parameter should affect them. This parameter, σ, sets the probability of a System 2 search for alternative models. Such a search can corroborate a putative conclusion or else provide a counterexample to it, i.e., a model in which the premises are true but the conclusion is false, and evidence shows that individuals are able to construct such counterexamples (Bucciarelli & Johnson-Laird, 1999; De Neys, Schaeken, & d'Ydewalle, 2003; Johnson-Laird & Hasson, 2003). In general, a search for counterexamples does not guarantee a correct response, but for the simple inferences, such those in our experiment, the model assumes that the search for alternative models always yields a correct answer.

An exhaustive exploration of the parameter space yielded an optimal σ value of .58, i.e., the system optimally modeled the data when it engaged a search for counterexamples 58% of the time. We generated synthetic data by running 1000 simulations of the 8 inferences. Figure 1 shows the proportion of correct responses in the observations (histograms with error bars) and predictions (circles) in the study as a function of the inference. The computer model matched the participants' performance in the experiment well ($R^2 = .75$, $RMSE = .11$). The predictions of the computer model were in the 99$^{th}$ percentile relative to

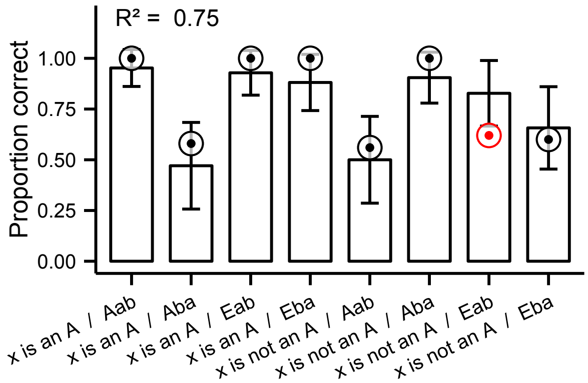


**Figure 1.** Observed (histograms with error bars) and predicted (circles) proportions of correct response for the eight different inferences in the experiment. Error bars show 95% confidence intervals. Black circles indicate when the predictions fell within the confidence interval of the observed proportion of correct responses, whereas the red circle indicates a deviation from the prediction to the observation. The premises are abbreviated using the conventions of Scholastic logicians: Aab = "All As are Bs", Eab = "No As are Bs" and likewise for Aba and Eba.

hypothetical datasets (Khemlani & Trafton, under review). However, the model underestimated performance on one inference (see Figure 1). When the inference was eliminated from the analysis, the model performed optimally ($R^2$ = .94). The system's inability to capture the inference may be a result of the determinacy of *None of the X are Y*, which implies that *None of the Y are X*. This symmetry does not hold for *All of the X are Y* (see above), and it may affect how models are initially constructed.

The computational model yielded a close fit to the data from the study, and successfully simulated the predicted difference between one- and multiple-model inferences.

## General Discussion

Set membership is a fundamental concept in set theory and elementary mathematics. It is also a rudimentary cognitive behaviour in object naming (Riddoch & Humphreys, 1987) and categorization (Murphy, 2002). Many higher order inferences presuppose the ability to reason about membership and non-membership. It is no accident that in the very first Western treatise on formal logic, the *Prior Analytics*, Aristotle discusses the example (70a25):

16. Pittacus is ambitious.
    Ambitious men are generous.
    Therefore, Pittacus is generous.

Given that the second premise refers to the set of ambitious men, the inference is valid. Our experimental results show that logically untrained individuals can readily make this inference, but that they tend to make invalid inferences akin to:

17. Pittacus is generous.
    All ambitious men are generous.
    Therefore, Pittacus is ambitious.

The model theory explains their performance: when the initial model yields the correct inference, as in (16), they are usually accurate. But, when the initial model does not yield the correct inference, they often err. They need to engage system 2 in a search for an alternative model. When they do so, they tend to be correct, or they may modify their initial conclusion to allow that it concerns only a possibility, e.g., *Pittacus is possibly ambitious*. A computer program implementing the theory, *mReasoner*, replicated participants' data by building and searching for models in the manner in which the theory posits.

Unlike set inclusion, membership is not a transitive relation, and so it is critical to distinguish between them. But, as we illustrated earlier (examples 6-8), it is not easy to do so. It is difficult to see how theories of reasoning based on formal rules of inference (e.g., Rips, 1994) or on probabilistic heuristics (e.g., Oaksford & Chater, 2007) can account for the following sort of valid inference:

18. Pat is one of the pianists.
    All the pianists are virtuosi who play the Minute Waltz in varying times down to 60 seconds.
    Therefore, Pat is a virtuoso who possibly plays the Minute Waltz in 60 seconds.

The conclusion that Pat is a virtuoso is an inference from set inclusion, whereas the inference about the speed of her performance is a complicated matter from the standpoint of formal logic. Nevertheless, it emerges from an iconic representation of the premises.

Reasoning about membership is a precursor to higher order inference. When reasoners can make set membership inferences about individuals, they can cope in principle with inferences about properties (Johnson-Laird, 2006, Ch. 10; Khemlani & Johnson-Laird, 2012). When entities are in a set, they inherit any properties that hold for the complete set. Indeed, given a premise asserting that a property holds for an entire set, another premise based on almost any sort of affirmative quantifier yields a non-trivially valid inference, e.g.:

19. All artists are benefactors.
    Most minimalists are artists.
    Therefore, most minimalists are benefactors.

A corresponding model of the premises is as follows:

```
artist   benefactor   minimalist
artist   benefactor   minimalist
artist   benefactor
```

These inferences are known as "monotone increasing" (Barwise & Cooper, 1981), and theorists have sometimes proposed rules designed to capture them (Geurts, 2003). Yet, they are merely emergent consequences of iconic models. The exceptions to such inferences are quantifiers that fix the number of members of a set (so-called "nonmonotone" quantifiers, see Barwise & Cooper, 1981). The following inference is therefore invalid:

20. All artists are benefactors.
    Exactly two minimalists are artists.
    Therefore, exactly two minimalists are benefactors.

The model theory predicts that naïve reasoners will sometimes make the inference because the initial model of the premises yields it. But, those who search for an alternative model may find the following counterexample:

```
  artist   benefactor   minimalist
  artist   benefactor   minimalist
  artist   benefactor
¬ artist   benefactor   minimalist
```

Hence, three minimalists could be benefactors.

In a similar way, when individuals are not members of a set, they are not in any of its subsets, e.g.:

21. All artists are benefactors.
    Most corporate raiders are not benefactors.
    Therefore, most corporate raiders are not artists.

These inferences are known as "monotone decreasing" (Barwise & Cooper, 1981). Once again, however, they are emergent properties of iconic models.

We conclude that the model theory provides a sensible account of reasoning that hinges on set membership, which extends naturally to reasoning about set inclusion. It has the advantage that particular inferences of considerable logical complexity emerge from models of the premises. Inferences become difficult for reasoners only when the correct response depends on moving from the intuitions of system 1 to the deliberations of system 2 and its construction of alternative models. People represent sets with sets of mental tokens, and this sort of representation allows them to treat quantified assertions in the way that Boole (1854) advocated, i.e., as relations between sets.

## Acknowledgements

This research was supported by a National Research Council Research Associateship awarded to SK and by NSF Grant No. SES 0844851 to PJL to study deductive and probabilistic reasoning. We are grateful to Ruth Byrne, Monica Bucciarelli, Sam Glucksberg, Adele Goldberg, Geoffrey Goodwin, and Marco Ragni, for their helpful comments and criticisms.

## References

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

Bates, D., Maechler, M., Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. http://CRAN.R-project.org/package=lme4.

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

Bucciarelli, M., & Johnson-Laird, P.N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*, 247-303.

Byrne, R.M.J., & Johnson-Laird, P.N. (1989). Spatial reasoning. *Journal of Memory and Language*, *28*, 564-575.

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

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

Goodwin, G., & Johnson-Laird, P.N. (2005). Reasoning about relations. *Psychological Review*, 112, 468-493.

Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, *86*, 223–251.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*, 31, 581-595.

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review*, *75*, 550–560.

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

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

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

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

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

Johnson-Laird, P.N., & Savary, F. (1999). Illusory inferences: a novel class of erroneous deductions. *Cognition, 71,* 191-229.

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

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

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

Khemlani, S., Trafton, J. G., & Johnson-Laird, P. N. (2013). Deduction as stochastic simulation. In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*.

Khemlani, S. & Trafton, J. G. (2014). Percentile analysis for goodness-of-fit comparisons of models to data. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Murphy, G.L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

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

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

Peirce CS (1931-1958). Hartshorne, C., Weiss, P., & Burks, A. (eds), *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

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

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

Steedman, M. (2012). *Taking scope: The natural semantics of quantifiers*. Cambridge, MA: MIT Press.

R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Riddoch, M. J., & Humphreys, G. W. (1987). Object naming. In G. W. Humphreys & M. J. Riddoch (Eds.), *Visual object processing: A cognitive neuropsychological approach* (pp. 107–143). Hove: Erlbaum.

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