Deduction as stochastic simulation

Sangeet Khemlani[1], J. Gregory Trafton[1], and P. N. Johnson-Laird[2,3]
sunny.khemlani.ctr@nrl.navy.mil, greg.trafton@nrl.navy.mil, phil@princeton.edu

[1]NCARAI, Code 5515, Naval Research Laboratory, Washington, DC 20375 USA
[2]Department of Psychology, Princeton University, Princeton, NJ 08540 USA,
[3]Department of Psychology, New York University, New York, NY 10012, USA

Abstract

Many theorists argue that deduction is based on the construction of mental models or simulations of descriptions. Individuals tend to reason intuitively from a single mental model, but on occasion they make a deliberate search for alternative models. Previous computer implementations of the theory were deterministic, but evidence from empirical studies suggested that a stochastic algorithm would have greater predictive power. We present such a system for inferences from assertions with single quantifiers, such as, "All the agents are lawyers". This system implements constraints on the size of model, the sorts of individual it represents, and on the likelihood of a search for alternative models. We show that the system yields quantitative predictions at a fine-grained level, and that they fit the data from two experiments better than previous accounts.

Keywords: deduction, mental models, reasoning, simulation, stochastic models

Introduction

Reasoners with no knowledge of logic can make deductive inferences. Consider the following examples:

All of the agents are lawyers. Does it follow that…
1. *none of the lawyers are agents?*
2. *all of the lawyers are agents?*

The correct answer to both (1) and (2) is 'no', but (1) is easier than (2), i.e., one study revealed 92% accuracy on the former inference, but only 67% accuracy on the latter inference (Newstead & Griggs, 1983, Experiment 1). Indeed, psychologists have known for a long time that similar inferences can differ in difficulty in a striking way (Begg & Harris, 1982; Wilkins, 1928). Their theories to explain these differences reflect three general approaches: the use of formal rules similar to the proof theory of logic (e.g., Rips, 1994), the use of heuristics for probabilistic validity (e.g., Oaksford & Chater, 2007), and the use of mental models (e.g., Johnson-Laird, 1983). The first two sorts of theory include parameters for fitting their predictions to the frequencies of different conclusions. Hitherto, however, the mental model theory has made parameter-free predictions about rank-order differences in difficulty. Yet, one of the major conclusions from a study of how individuals used external models in reasoning was that the individual human reasoning system is not deterministic. The same person constructs different models of the same premise from one inference to another (Bucciarelli & Johnson-Laird, 1999).

The mental model theory of higher cognition postulates that reasoning relies on the construction and manipulation of *mental models*, i.e., iconic simulations of possible situations (Johnson-Laird, 2006). As the theory posits, reasoners who are told that *all of the agents are lawyers* simulate the situation, that is, they build a model of a small arbitrary number of tokens that denote individual agents, and then add the property of *being a lawyer* to each of the agents. A model of such a situation is akin to the following diagram, in which each row denotes the properties of an individual:

```
agent   lawyer
agent   lawyer
agent   lawyer
agent   lawyer
```

Reasoners can immediately make deductions by inspecting the model. Since at least one of the lawyers is an agent, it does not follow that none of the lawyers are agents. An inference is *valid* if its conclusion is true in every case in which the premise is true (Jeffrey, 1981, p. 1). And so the initial model establishes the invalidity of the inference.

Other inferences require reasoners to revise initial models. For instance, does it follow that all of the lawyers are agents? It does in the initial model, but an alternative model of the premise is a counterexample:

```
 agent   lawyer
 agent   lawyer
 agent   lawyer
 agent   lawyer
¬agent   lawyer
```

where '¬' denotes the mental symbol for negation. Thus, the correct answer is 'no', but it calls for reasoners to construct multiple models. In sum, the model theory accounts both for valid responses and for relative difficulties in inference, and it does so using set-theoretic models rather than formal rules of inference or probabilities (Johnson-Laird, 2006).

Yet, the theory is limited. It explains the rank order of the difficulty of inferences (see Khemlani & Johnson-Laird, 2012, 2013, for reviews), but it has been parameterized to account for quantitative differences in only a limited way – a notable exception is Schroyens and Schaeken's (2003) formalization of its account of conditional reasoning (see also Oberauer, 2006). In other domains, such as the

inferences above, it cannot account for quantitative patterns of reasoning, e.g., why participants were correct for 92% of inferences of (1), but for 67% of inferences of (2). The problem is the deterministic computational implementations of the theory (e.g., Johnson-Laird & Byrne, 1991). Determinism is counterintuitive for three principal reasons. First, individuals vary one from another in how they reason even from just a pair of premises. Second, a given individual varies from one inference of the same sort to another. Third, as we illustrated earlier, a premise that refers to distinct possibilities is compatible with several distinct models, and individuals vary in the initial model that they construct – a difference that occurs both between and within individuals (see Polk & Newell, 1995, for a model-based theory that accommodates differences between individuals). Reasoners may vary in the number of tokens that they construct to represent a particular sort of individual, i.e., the size of the model varies (see Table 1). All of these factors, in turn, are susceptible to hidden unknown variables that yield noise, and that may therefore contribute to errors.

In the present paper, we outline a new stochastic system for building mental models. It relies on three components, which serve as the parameters of the system for predicting performance. We describe the system in detail, and then show how it provides a close fit to the data from an experiment on immediate inferences.

A stochastic model-building system

The model-building system embodies three stochastic parameters. The first parameter constrains the size of a model, that is, the number of different tokens representing entities that it contains. The second parameter constrains the contents of a model, and in particular the different sorts of entities that it represents. The third parameter constrains the likelihood that the system searches for an alternative to the initial model of premises. In summary, the system manipulates stochastically the size, the contents, and the revisions of models. We now describe in detail each parameter and its effects.

The size of a mental model (parameter λ)

The interpretation of a quantified assertion, such as, *All of the agents are lawyers* calls for a mental model of a set of individual agents that are lawyers. The cardinality of the set is not fixed, but varies depending on the inference, the capacity of working memory, and other cognitive constraints. These variables can be approximated in a discrete probability distribution. The distribution must be discrete, because the elements of a mental model are discrete, e.g., there can be no scenario in which 4.5 agents are lawyers. We therefore assume that the cardinality of a model varies according to a Poisson distribution, which is typically used to describe a given number of events occurring within a fixed interval of time. For example,

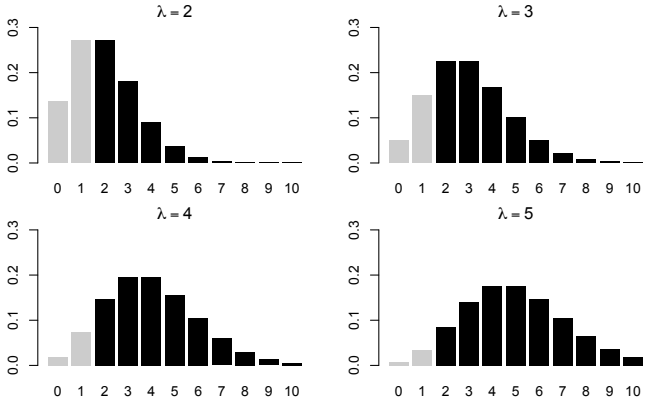

Figure 1. Left truncated Poisson distributions for various values of λ. The distributions establish the stochastic cardinality for building models. Gray bars indicate values that were truncated (0 and 1).

a Poisson distribution can describe the number of emails an individual receives in a day: some days she might get no emails, other days five, other days twenty. In the reasoning system, "events" correspond to the mental consideration of tokens in a model. When reasoners interpret the assertion *All of the agents are lawyers*, they may consider two such individuals, three, four, or maybe five. A single real-valued parameter, λ, governs the shape of a Poisson distribution, and λ is both the expected value and the variance of the distribution. In addition, the Poisson distribution is *left truncated*, i.e., zeroes and ones are excluded (see Deshpande, Gore, & Shanubhogue, 1995, p. 199). When building a model, the model's cardinality is established by a single sample drawn from a left-truncated Poisson distribution of parameter λ. Figure 1 presents the frequencies of such samples for various values of λ, and it shows, for example, that for λ = 3 the majority of them will be 2, 3, 4, or 5. Table 1 shows these models.

The adoption of a stochastic value for the cardinality of a model can potentially capture more of the variation in reasoners' mental models. But, as the examples in Table 1 show, merely varying the size of a model does not always yield inferential flexibility. That is, the same inference (e.g., that it follows that some lawyers are agents) can be drawn from a model of three individuals as can be drawn from a model of five individuals. The second component in the system, however, directly affects inferences.

Cardinality									
2		3		4		5			
A	L	A	L	A	L	A	L		
A	L	A	L	A	L	A	L		
		A	L	A	L	A	L		
				A	L	A	L		
						A	L		

Table 1. Possible models of "All of the agents are lawyers", where A denotes an individual agent, and L denotes an individual lawyer.

The contents of a mental model (parameter ε)

The second component governs the contents of a model, i.e., the particular sorts of individual or entity that it represents. The models in Table 1 are canonical according to the model theory in that they represent only agents that are lawyers. But, the assertion is compatible with the situation in which there are some lawyers that are not agents. Thus, the following models are all feasible representations of *All the agents are lawyers*:

```
A   L        A   L        A   L        A   L
A   L        A   L        A   L        A   L
    L            L        A   L        A   L
    L            L        A   L        A   L
                 L            L        A   L
```

The system embodies the principle that some proportion of models deviate from the canonical sort. Models are built by drawing samples from one of two sets, either the *canonical* set of individuals for a particular type of quantified assertion, or else the *full set* of possible individuals compatible with the assertion (see Appendix). Every time the system draws a sample, the probability that it will draw it from the full set of possible individuals is given by ε, e.g., if ε = .10, then there is a 10% chance that the system will draw a model from the full set of individuals. To illustrate, the premise *All the agents are lawyers*, has only one canonical individual:

```
A   L
```

but two sorts of individual in the full set of possibilities:

```
¬A   L
¬A  ¬L
```

And so when ε = 0, any model built from drawing from the canonical set of individuals will comprised only agents that are lawyers, as in Table 1. But when ε = .5, models are built from drawing individuals from the full set of possibilities on half the occasions:

```
 A   L        A   L        A   L        A   L
¬A   L       ¬A  ¬L        A   L        A   L
             ¬A  ¬L       ¬A   L       ¬A   L
                          ¬A  ¬L       ¬A   L
                                       ¬A  ¬L
```

All of the models satisfy the premise, but variations in their contents afford different inferences. For example, in some of the models above, all of the lawyers are agents. In others, some of the lawyers are not agents. And in yet others, most, but not all, of the lawyers are agents.

The λ and ε parameters constrain the system's construction of initial models, and an underlying assumption of the model theory is that individuals tend to reason based on their intuitions, i.e., to draw conclusions based on initial mental models alone. Nevertheless, reasoners can make a deliberate search for alternative models, and the system contains a third parameter concerning the likelihood of such a search.

Modifying initial models (parameter σ)

The model theory assumes that reasoners tend to reason from a single model or simulated scenario (see also Vul et al., 2009). Nevertheless, individuals can search for alternative models by modifying their initial model. In this way, they may corroborate a putative conclusion or they may discover a counterexample to it, i.e., a model in which the premises are true but the conclusion is false (Bucciarelli & Johnson-Laird, 1999; De Neys, Schaeken, & d'Ydewalle, 2003; Neth & Johnson-Laird, 1999; Johnson-Laird & Hasson, 2003). Indeed, the simpler the inference, the more likely reasoners should be to modify their initial models. The likelihood of modifying initial models, i.e., engaging in a search for alternative models, is specified by the final parameter σ in the system, It fixes the proportion of inferences in which a search occurs, e.g., when σ = .3, the system searches for alternative models on 30% of inferences. In general, a search for alternative models is not guaranteed to return a correct response, but the prototype of the stochastic system operationalizes σ as the probability of obtaining the correct answer.

The simulation of immediate inferences

The system contains three stochastic components, λ (the number of individuals in a model), ε (the sorts of individual in a model), and σ (the search for alternative models, which yield the correct response). We implemented the system in R (R Core Team, 2012), and its code can be accessed at http://goo.gl/nATWH. We assessed its ability to account for immediate deductive inferences from a single quantified premise, as illustrated in the introductory examples in the present paper. Many studies have examined these inferences (e.g., Begg & Harris, 1982; Newstead & Griggs, 1983; Wilkins, 1928) but there is no theory that makes quantitative predictions about them. The inferences are based on four different moods of singly-quantified premise:

All the Xs are Ys
Some of the Xs are Ys
None of the Xs are Ys
Some of the Xs are not Ys

and 8 different sorts of conclusion (4 moods by 2 arrangements of terms X and Y). There are therefore 32 possible immediate inferences based on these premises. A typical inference is:

Some of the alchemists are barbers.
Does it follow that all of the barbers are alchemists?

We chose immediate inferences as a test case, because the model theory can make qualitative predictions about the relative difficulties of three sorts of immediate inference: a) zero-model inferences, b) one-model inferences, and c)

multiple-model inferences. Zero-model inferences yield a correct response without requiring a model, e.g., the conclusion is identical to the premise. One-model inferences yield the correct response from the initial model. Multiple-model inferences yield a correct response only from an alternative to the initial model. Recent analyses corroborate the predicted trend that zero-model inferences should be easier than one-model inferences, which in turn should be easier than multiple-model inferences (Khemlani, Trafton, Lotstein, & Johnson-Laird, 2012; Khemlani & Trafton, 2012).

Our present goal, in contrast, was to assess both quantitative and qualitative predictions of difficulty as derived from the stochastic system. We therefore simulated two experiments from Newstead and Griggs (1983) that examined all 32 immediate inferences. Experiment 1 used assertions containing letters, and Experiment 2 used assertions containing letters, and assertions containing terms such as, "artists" and "beekeepers".

Results and discussion

We carried out a search for the values of the parameters using the data from Experiment 1. Each triplet of parameters was used to simulate each of the 32 inferences 100 times. The search found optimal fits when $\lambda = 4$, $\varepsilon = .3$, and $\sigma = .4$. In other words, the search yielded the best fit to the data when models consisted of around 4 separate entities; when 30% of models contained non-canonical individuals; and the search for alternative models, which yield the correct response, occurred for 40% of the inferences.

We then carried out simulations for Experiments 1 and 2 in Newstead and Griggs (1983). The parameter values obtained from the parameter search were used to simulate Experiment 1, and to cross-validate the system on the data from Experiment 2. Each simulation consisted of 1000 simulations of the 32 inferences. The deterministic theory classified the inferences as zero-, one-, or else multiple-model inferences (see above). Hence, Figure 2 shows the proportion of correct responses in the observations (histograms with error bars) and predictions (circles) in the two studies as a function of the type of inference. The stochastic system closely matched the performance of Newstead and Griggs' participants for Experiment 1 (R^2 = .97, RMSE = .12) and Experiment 2 (R^2 = .93, RMSE = .12). The data from both studies are comparable to one another, and so the fit for Experiment 2 revealed successful cross-validation of the parameters, and suggested that the parameters did not overfit the data from Experiment 1.

The stochastic system yields more detailed predictions, because it predicts performance for each of the 32 separate inferences. Figure 3 shows the system's fit for each of the inferences. As expected, the goodness of fit metrics were not as close as those for the fits with group means (R^2 = .57, RMSE = .21 and R^2 = .61, RMSE = .23 for Experiments 1 and 2 respectively). However, the figure reveals that the

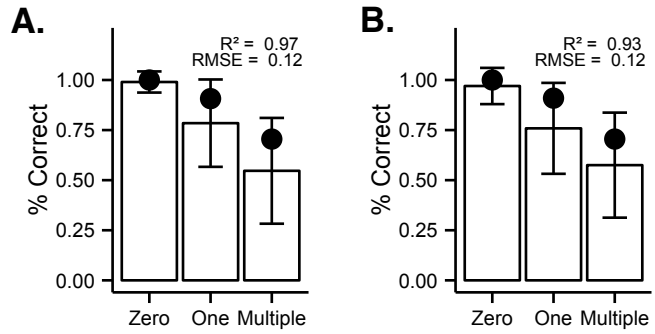

Figure 2. Observed (histograms with error bars) and predicted (circles) proportions of correct response for the three types of inference (i.e., zero-, one-, and multiple-model). Panel A shows the data from Experiment 1, and Panel B shows the data from Experiment 2. Error bars show 99% confidence intervals.

majority of the system's predictions fall within the confidence intervals of the observations. What explains the decline in the coefficients of determination?

One reason is that the coefficients do not take into account the number of separate data points to be predicted: a fit is bound to be smaller for 32 inferences than for 3 group means. We therefore used an alternative metric, the Bayesian Information Criterion (BIC; see Schwarz, 1978), as a selection measure that takes into account both the number of data points that were fit and the number of parameters that were used to fit them. Table 2 presents the BIC values for the two experiments. Lower values of BIC indicate a better fit, and so as the Table shows the fit was better for the 32 inferences than for the 3 types of inference.

To assess the relative importance of each parameter, we systematically disabled them and compared the results to the version of the system in which all parameters are enabled (cf. Altmann & Trafton, 2002). The results are shown in Table 3. They suggest that disabling any of the parameters has a deleterious effect on the fit between the predictions and the data. However, disabling the λ parameter (the number of individuals in a model) appeared to diminish the system's performance the most. In contrast, while disabling the σ parameter (the likelihood of searching for an alternative model, which yielded the correct response) affected on the R^2 value negatively, it seemed to yield a slight improvement for the RMSE.

	Bayesian Information Criterion (BIC)	
	by inference type (n = 3, k = 3)	by immediate inference (n = 32, k = 3)
Experiment 1	-8.43	-87.51
Experiment 2	-8.40	-82.31

Table 2. The values of the Bayesian Information Criterion (BIC) for the fit of the predictions to the observed data for the 3 sorts of inference and for the 32 individual inferences (in Experiments 1 and 2). The system uses three parameters to fit the data, i.e., k = 3.

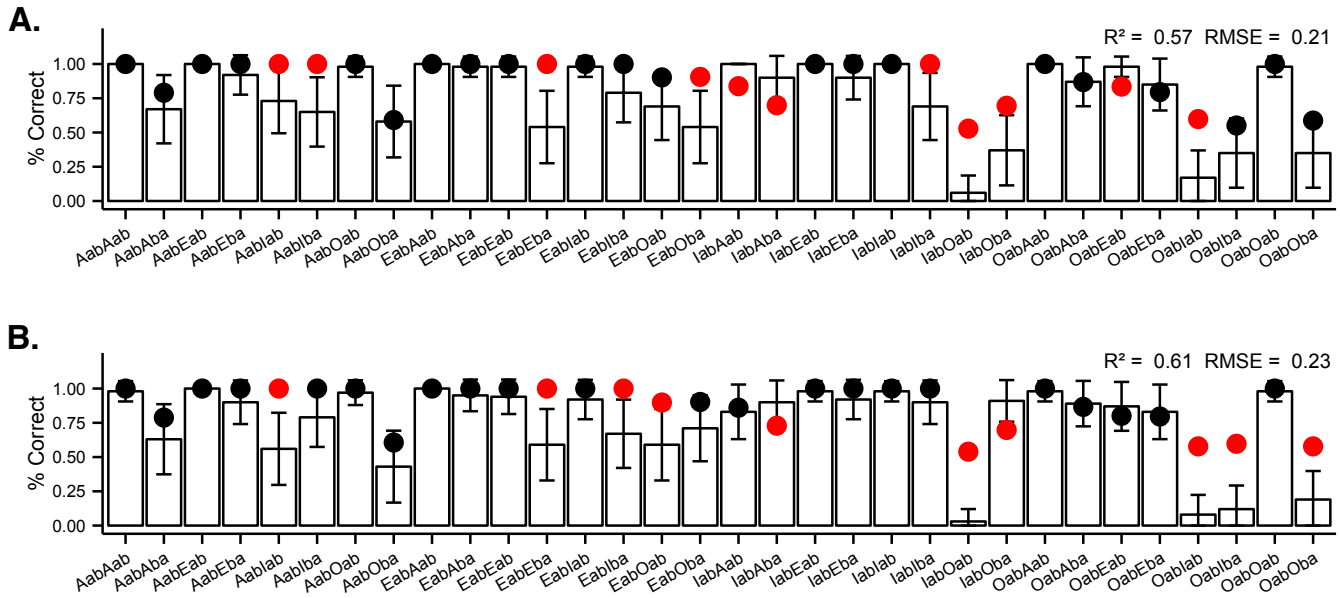

Figure 3. Observed (histograms and error bars) and predicted (circles) proportions of correct response for 32 immediate inferences. Panel A shows the data from Experiment 1 in Newstead and Griggs (1983), and Panel B shows the data from their Experiment 2. Error bars reflect 99% confidence intervals. Black circles indicate when the predictions fell within the confidence interval of the observed proportion of correct responses, while red circles indicate deviations from the predictions and the observation. Premises and conclusions are abbreviated using the conventions of Scholastic logicians, as follows: *Aab* = "All As are Bs", *Iab* = "Some As are Bs", *Eab* = "No As are Bs", and *Oab* = "Some As are not Bs," and analogous conventions for *Aba*, *Iba*, *Eba*, and *Oba*. Each inference is abbreviated as a premise concatenated with its putative conclusion, e.g., *AabAab* denotes the inference, "All As are Bs. Does it follow that all As are Bs?"

In sum, the stochastic system provided quantitative predictions of accuracy across the different types of inference (zero-, one-, and multiple-model inferences) and across the 32 individual inferences themselves, and those predictions closely matched the results of two experiments.

	All parameters enabled	Disabled parameters			
		λ	ε	σ	all
R^2	.57	.38	.52	.53	.34
RMSE	.21	.24	.22	.20	.25

Table 3. The goodness of fit between the predictions and the observations (Newstead & Griggs, 1983, Experiment 1) when all parameters are enabled; when λ is disabled, i.e., models are built with a fixed cardinality of 3; when ε = 0; when σ = 0; and when all parameters are disabled.

General Discussion

The stochastic system for building mental models rests on the idea that the fundamental engine of deductive inference is the process of constructing, manipulating, and inspecting set-theoretic models – a constant in the theory from Johnson-Laird (1983) to Khemlani and Johnson-Laird (2012). The stochastic system embodies three parameters that allow variation from inference to inference in the number of individuals represented in a model, in the sorts of individuals, which may be canonical or from the set of possible individuals as a whole, and in the likelihood that a search is made for alternatives to the initial model, which is treated in the system as a way to guarantee the correct response.

Can the system be extended to provide a closer fit to the 32 individual immediate inferences? Figure 3 suggests so in two respects. First, many of the participants' errors are cases of "scalar implicatures" (Gazdar, 1979; Horn, 2004). In particular, the stochastic system responded that "Some of the As are Bs" follows from "All of the As are Bs," whereas the participants drew this conclusion only 73% in Experiment 1 and 56% in Experiment 2. We suspect that they rejected the inference on the grounds that the conclusion is less informative than the premise, and so it violates the conversational convention to be informative (Grice, 1975). The present system might therefore be enhanced if it incorporated a computational model of scalar implicature (e.g., Goodman & Stuhlmüller, 2012). Second, most predictions that deviate from the data (the red circles in the figure) are overestimates of accuracy, and therefore suggest that the parameter search may have yielded too large a value of σ. Extensions of the model should therefore correct for these two factors.

The three components of the stochastic system reflect psychologically motivated constraints on inferential processes. In the present analysis, we compared the system's ability to predict accuracy across items, i.e., we assumed that the noise contributed by each participant was a random effect. The opposite approach is viable as well: the

stochastic system's parameters could be tweaked for individual reasoners. For example, the λ parameter may theoretically correspond to working memory span, and the ε and σ parameters may correspond to analytic processing (Stanovich & West, 2000). Thus, the system may be a foundation for developing computational cognitive models that analyze individual differences in deductive reasoning.

The stochastic system is relevant to the broader context of developing unified theories of inference. We have taken pains to describe the system as "stochastic" and not "probabilistic", even though the system is based on the interaction of three variables with values chosen according to probabilities. The reason is that "probabilistic" is often used to characterize theories in which inferences are based on subjective probabilities, transformations between them, such as Bayes's theorem, and heuristic approximations of probabilistic validity (see, e.g., Oaksford & Chater, 2007). In our account, however, mental models, i.e., set-theoretic models of possibilities, are the primary representation that makes inferences possible. The stochastic components constrain their construction, but it is the models themselves that allow reasoners to draw deductive conclusions. Thus, our use of the term "stochastic" refers, not to the meaning of quantified assertions, but to the system's non-determinism.

In conclusion, stochastic model building is a way to approximate the variation in mental simulations that underlie deductive inference. The size of a model, its contents, and its propensity for revision, are each critical in accounting for inferences with quantifiers.

Acknowledgments

This research was funded by a National Research Council Research Associateship awarded to SK; ONR Grant #s N0001412WX30002 and N0001411WX20516 to JGT; and NSF Grant No. SES 0844851 awarded to PJL. We are grateful to Chand Chandrasekaran and Tony Harrison for their helpful comments and criticisms. The views and conclusions contained in this document do not represent the official policies of the US Navy.

References

Altmann, E., & Trafton, J.G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26,* 39-83.

Begg, I., & Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behavior*, *21*, 595–620.

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

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition, 31,* 581-595.Deshpande, J.V., Gore, A. P., Shanubhogue, A. (1995). *Statistical analysis of nonnormal data.* Wiley Interscience.

Gazdar, G. (1979). *Pragmatics, Implicature, Presupposition, and Logical Form.* New York: Academic Press.

Goodman, N., & Stuhlmüller, A. (2012). Knowledge and implicature: Modeling language understanding as social cognition. In N. Miyake, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Grice, H. P. (1975). Logic and conversation. In: P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts*, (pp. 41-58). New York: Academic Press.

Horn, L. (2004) Implicature. In: L. R. Horn and G. Ward (Eds.), *The Handbook of Pragmatics*. Oxford: Basil Blackwell.

Johnson-Laird. P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

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

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

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. (2012). mReactr: A computational theory of deductive reasoning. In N. Miyake, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Khemlani, S., Trafton, J. G., Lotstein, M., & Johnson-Laird, P. N. (2012). A process model of immediate inferences. In N. Rußwinkel, U. Drewitz, & H. van Rijn (Eds.), *Proceedings of the 11th International Conference on Cognitive Modeling*. Berlin, Germany: University of Groningen.

Neth, H., & Johnson-Laird, P. N. (1999). The search for counterexamples in human reasoning. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (p. 806). Mahwah, NJ: Erlbaum.

Newstead, S.E., & Griggs, R.A. (1983). Drawing inferences from quantified statements: A study of the square of opposition. *Journal of Verbal Learning and Verbal Behavior*, *22*, 535–546.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford: Oxford University Press.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories . *Cognitive Psychology*, *53*, 238-283.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 140–149.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Stanovich, K., & West, R. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-726.

Vul E., Goodman N.D., Griffiths T.L., & Tenenbaum J.B. (2009). One and Done? Optimal decisions from very few samples. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Appendix

Canonical and full model sets for the four types of assertions that the stochastic model building system simulates. Canonical sets reflect the sorts of individuals people are most likely to build, whereas full sets reflect all individuals compatible with the assertion.

	All As are Bs	Some As are Bs	No As are Bs	Some As are not Bs
Canonical	{ A B }	{ A B, ¬A B }	{ A ¬B, ¬A B }	{ A B, A ¬B, ¬A B }
Full	{ A B, ¬A B, ¬A ¬B }	{ A B, A ¬B, ¬A B, ¬A ¬B }	{ A ¬B, ¬A B, ¬A ¬B }	{ A B, A ¬B, ¬A B, ¬A ¬B }