Memory & Cognition
2007, 35 (6), 1353-1364

Induction as conditional probability judgment

Sergey V. Blok
University of Texas, Austin, Texas

Douglas L. Medin
Northwestern University, Evanston, Illinois

and

Daniel N. Osherson
Princeton University, Princeton, New Jersey

Existing research on category-based induction has primarily focused on reasoning about *blank* properties, or predicates that are designed to elicit little prior knowledge. Here, we address reasoning about nonblank properties. We introduce a model of conditional probability that assumes that the conclusion prior probability is revised to the extent warranted by the evidence in the premise. The degree of revision is a function of the relevance of the premise category to the conclusion and the informativeness of the premise statement. An algebraic formulation with no free parameters accurately predicted conditional probabilities for single- and two-premise conditionals (Experiments 1 and 3), as well as problems involving negative evidence (Experiment 2).

Studies of inductive inference are usually framed in terms of projecting an unfamiliar (blank) property from one category to another, as in

1. Wolves *have sesamoid bones*, therefore bears *have sesamoid bones*.

Participants are typically asked to evaluate *arguments* like the one above with respect to the extent to which the premise supports the conclusion. Several models of induction have used similarity to explain willingness to project properties (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975). For example, the similarity–coverage model (Osherson et al., 1990) assumes that the strength of categorical arguments is related to the *similarity* between the premise and the conclusion categories and *coverage*, or the extent to which the premise category is representative of a superordinate that includes the premise and conclusion kinds (see Sloman, 1993, for a feature-based alternative).

Although the similarity–coverage model is able to capture reasoning about blank properties, it does not appear to extend to reasoning about nonblank or familiar predicates. Smith, Shafir, and Osherson (1993) presented examples in which similarity is unable to account for reasoning with some nonblank properties. For instance, their participants reliably chose Argument 2 over Argument 3:

2. Poodles *can bite through wire*, therefore German shepherds *can bite through wire*.

3. Dobermans *can bite through wire*, therefore German shepherds *can bite through wire*.

Smith et al. (1993) assumed that in evaluating statements like (2) and (3), people changed their representations so as to minimize the coherence *gap* between the premise facts in the arguments and prior knowledge of the categories and properties in question. Specifically, people might observe that the premise fact in (2), "Poodles *can bite through wire*," demands belief revision because it is surprising. The reasoner may then update his or her beliefs by assuming that poodles are stronger then previously believed. Alternatively, one may also close the coherence gap by concluding that biting through wire is easier than was previously thought (see Osherson, Smith, Myers, Shafir, & Stob, 1994, for a model that incorporates both processes).

Although the intuitions behind the gap model are important, the formulation presented in Smith et al. (1993) needs elaboration. Specifically, in order to generate a probability of an object's having a predicate, the model integrates object and predicate values for a given dimension, which are, in turn, converted into probabilities. Thus, if poodles have an a priori strength of 5 and it takes a strength of 7 to bite through wire, the model predicts the probability that poodles will be able to bite through wire as .33. Incoherence arises when we represent the attributes by rescaling the values. If, instead of attribute values of 5 and 7, we consider 10 and 14 (keeping the relative gap the same), the corresponding probability becomes .20 (see Blok, 2004, for more details).

SimProb MODEL

Here, we propose another model of induction with nonblank predicates, called *SimProb*. The inputs to our model

S. V. Blok, blok@psy.utexas.edu

are prior probabilities for premise and conclusion events and similarities between the categories involved. We believe that starting with context-independent probabilities is an improvement over the gap model because SimProb does not rely on strong hypotheses regarding the decomposition of categories and predicates into features. All we require are prior probabilities and similarity values. To predict the conditional probability of a conclusion, given the premise, the initial values are combined through an algebraic function whose behavior accords with a set of qualitative requirements. These constraints are grounded in limiting-case scenarios (e.g., "What should happen to the conditional probability when conclusion probability approaches 1.0?" or "What should happen when the similarity between the premise and the conclusion approaches 0.0?"). The requirements stemming from probability considerations are normatively sanctioned, although those that arise from similarity are psychological in nature. In this article, we present functions for predicting judgments about single- and two-premise arguments, as well as those involving negative and mixed evidence. To preview, SimProb provides a good account for all of these problem frames. This is accomplished without estimating any free parameters and requires only the prior probability and similarity estimates provided by participants. Thus, one of the benefits of our approach is the simplicity of its formulation and testing. Another is the relative ease of extending the theory to handle negative and mixed evidence.

Before proceeding to a more detailed discussion of the model, we will mention some of its limitations. The first has to do with the use of similarity to account for reasoning. Ever since Goodman's (1955) warnings about similarity's status as a "false friend," psychologists have been cautious about attributing explanatory power to this overly flexible construct. Any two objects can be similar, depending on the dimension of comparison. Work in reasoning parallels this skepticism by showing that inductive judgments are based on similarity with respect to a dimension picked out by the predicate (Heit & Rubinstein, 1994). In this article, we have selected our predicates so that the dimensions relevant to reasoning about them are likely to be the same as those that guide similarity judgments (we call these *stable* predicates). For example, reasoning about biological predicates (e.g., "has biotin") has been shown to be guided by taxonomic similarity between species (Heit, 2000; Osherson et al., 1990). In this context, "has biotin" is stable with respect to a set of biological categories, because biological similarity accounts for both the similarity ratings between category members and the projection of the property. By contrast, a nonstable predicate, such as "weigh more than 10 kilos," is likely to promote inductions not predicted by similarity ratings. The use of stable predicates, although common throughout induction research, is clearly a simplifying assumption made in the name of progress. We will leave the extension of SimProb to nonstable predicates to future work.

Another simplification is that we do not provide an account of the source of prior probabilities, only conditional ones (for a good model for the generation of priors, see Juslin & Persson, 2002). Finally, we note that the model we are proposing is not intended as a process account of reasoning; we do not suggest that people follow the steps we are specifying here, only that our formulas approximate the output of the reasoning procedure.

Specific Formulation

Our goal is to predict the conditional probability of a statement, given one or more others. In our experiments, statements always have subject–predicate form, as in "Foxes *have good night vision*." As in prior studies of induction, all the statements in an argument share the same predicate. In order to predict the conditional probability of the argument's conclusion, given its premise, we will need only the prior probabilities of the statements "Foxes *have good night vision*" and "Wolves *have good night vision*," as well the similarity between foxes and wolves.

To facilitate exposition, our notation for a single-premise conditional will be $\Pr(Qc \mid Qa)$, where Q stands for the predicate and c and a stand for the conclusion and premise categories, respectively. Similarly, a two-premise conditional will be expressed as $\Pr(Qc \mid Qa,Qb)$. We will also consider cases of negative evidence, such as the likelihood that "Foxes *have good night vision*, given that Wolves *do NOT have good night vision*." Such conditionals will be formalized as $\Pr(Qc \mid \neg Qa)$.

The algebraic formulation of our theory is constrained by a set of conditions concerning limiting cases. For example, a trivial observation is that the conditional probability should fall between 0.0 and 1.0. More substantively, constraints arise when we ask what the output of the function should be when premise–conclusion similarity or priors reach certain limit values. For example, it seems psychologically reasonable that as premise–conclusion category similarity approaches identity, the conditional judgment should approach 1.0, as in

$$\Pr(\text{Hogs } \textit{have } X \mid \text{Pigs } \textit{have } X) \approx 1.0.$$

Probability considerations can also impose constraints on the function. For example, when conclusion probability $\Pr(Qc)$ approaches 1.0, so should the conditional. This constraint reflects the intuition that the conditional is positively correlated with the conclusion prior. Yet another probability-based constraint is that surprising premise facts should have a greater impact on the conditional than do expected or noninformative ones (Lo, Sides, Rozelle, & Osherson, 2002; Smith et al., 1993). Appendix A presents the full set of limiting case constraints applicable to single- and two-premise judgments.

Equation 1: Single-Premise Formulation of SimProb

$\Pr(Qc \mid Qa) = \Pr(Qc)^{\alpha}$, where

$$\alpha = \left[\frac{1-\mathrm{sim}(a,c)}{1+\mathrm{sim}(a,c)}\right]^{1-\Pr(Qa)}.$$

Equation 1 is derived to satisfy the constraints described above. [A negative evidence version of Equation 1 is presented in Appendix B. This formula is designed to predict $\Pr(Qc \mid \neg Qa)$ and is symmetrical to the positive evidence theory in a straightforward way.] The conditional probability $\Pr(Qc \mid Qa)$ is expressed in terms of the prior probability of the conclusion statement $\Pr(Qc)$, the prior probability of the premise statement $\Pr(Qa)$, and the similarity between the conclusion and the premise categories $\mathrm{sim}(a,c)$. SimProb can be interpreted in terms of belief revision. The reasoner begins with his or her prior probability for the conclusion and revises it to the extent warranted by the evidence contained in the premise. Two factors determine the extent to which the conclusion prior probability is revised. First, the premise category has to be sufficiently *relevant* to the conclusion category. In SimProb, relevance is represented by similarity. Generally, facts about highly dissimilar categories are discounted. In terms of the formulation above, when similarity tends to 0.0, the revision exponent α tends to 1.0. Consequently, the conditional $\Pr(Qc \mid Qa)$ approaches $\Pr(Qc)$, indicating that no revision should take place. Conversely, facts about categories that are psychologically close should exhibit maximum revision and push the conditional $\Pr(Qc \mid Qa)$ to 1.0. In terms of the formulation, this means that α should approach 0.0. One can verify that when $\mathrm{sim}(a,c)$ approaches 1.0, α indeed approaches 0.0, and conditional probability approaches 1.0. The second factor governing revision is *informativeness*, or the extent to which a premise provides new information, rather than telling the reasoner what he or she already knows. Informativeness is expressed as the inverse of the premise prior probability, or $1-\Pr(Qa)$. When the premise fact is perfectly unsurprising or uninformative, no belief revision should take place, and the conditional should remain at the level of the conclusion prior. In terms of the formulation, a perfectly uninformative premise fact has a prior $\Pr(Qa)$ approaching 1.0. If this is the case, the revision exponent α approaches 1.0, and the conditional $\Pr(Qc \mid Qa)$ approaches $\Pr(Qc)$.

Informativeness captures the difference between poodles and dobermans as premises in the example discussed earlier (Arguments 2 and 3). The premise fact about poodles being able to bite through wire is more informative (surprising) than the premise fact about dobermans; hence, there will be greater preference for the former than for the latter. We assume that although the lower similarity of poodles and German shepherds should make (2) less preferred than (3), the gain in strength for (2) due to higher informativeness should outweigh the loss due to lower similarity.

We do not claim that Equation 1 is the only possible formula that satisfies the limiting-case constraints we have outlined. More reasonably, it is likely to be an instance of a class of models that do so. We suspect that any model that satisfies the constraints will have the same fit to the data as the present formulation. A reader may also wonder why the more complex exponential form was chosen over a potentially simpler linear combination of variables. The answer is that we have not discovered any simpler linear formula that captures the constraints.

We now will extend Equation 1 to two-premise conditionals. Since none of our experiments involve negative conclusions, we only derive formulas that vary the valence of the premise categories. Hence, we will consider just the cases $\Pr(Qc \mid Qa,Qb)$, $\Pr(Qc \mid \neg Qa,\neg Qb)$, and $\Pr(Qc \mid Qa,\neg Qb)$. Before presenting Equation 2, we introduce the notion of a dominant premise, described in terms of a confirmation function.

Definition 1: The Confirmation Function

The *confirmation* exhibited by the conditional $\Pr(Qc \mid Qa)$ is

$$\frac{\Pr(Qc \mid Qa) - \Pr(Qc)}{1 - \Pr(Qc)}.$$

The *confirmation* exhibited by the conditional $\Pr(Qc \mid \neg Qa)$ is

$$\frac{\Pr(Qc) - \Pr(Qc \mid \neg Qa)}{\Pr(Qc)}.$$

To illustrate the function, we will focus on its positive version. The numerator reflects the impact of the premise fact, expressed as a change in probability between the prior and the conditional. The denominator captures the maximum possible impact of a premise. Thus, the confirmation function is the actual impact of the premise normalized by its potential impact. The negative premise function is symmetric to the positive formula. A variety of confirmation measures are analyzed in Eells and Fitelson (2002). In Tentori, Crupi, Bonini, and Osherson (2007), they are compared for their ability to predict shifts of opinion in an experimental setting involving sampling from urns.

The *dominant* premise in a two-premise argument is the one that yields the one-premise argument of greatest confirmation. The one-premise probabilities are derived from the theory of one-premise arguments offered above. We now will present our theory of two-premise arguments. Equation 2 presents the formulation for positive two-premise conditionals.

Equation 2: Two-Premise Formulation of SimProb

Conditionals of the form $\Pr(Qc \mid Qa,Qb)$ with Qa dominant:

$$\begin{aligned}\Pr(Qc \mid Qa,Qb) = \Pr(Qc \mid Qa) + [&(1 - \Pr(Qc \mid Qa))\\ &\times (1 - \mathrm{sim}(a,b))\\ &\times (\Pr(Qc \mid Qb) - \Pr(Qc))].\end{aligned}$$

Conditionals of the form $\Pr(Qc \mid Qa,Qb)$ with Qb dominant:

$$\begin{aligned}\Pr(Qc \mid Qa,Qb) = \Pr(Qc \mid Qb) + [&(1 - \Pr(Qc \mid Qb))\\ &\times (1 - \mathrm{sim}(a,b))\\ &\times (\Pr(Qc \mid Qa) - \Pr(Qc))].\end{aligned}$$

In words, Equation 2 reflects the idea that the reasoner starts out with the conditional probability resulting from only the dominant premise [$\Pr(Qc \mid Qa)$ if Qa is dominant]. He or she then adds a fraction of the remaining *lack* of confidence $1 - \Pr(Qc \mid Qa)$ that dominant conditional "leaves behind." The size of the fraction is determined by the similarity between premise categories $\mathrm{sim}(a,b)$ and the separate impact of the nondominant premise on the conclusion prior $\Pr(Qc \mid Qb) - \Pr(Qc)$. The similarity component is designed to diminish the impact of the nondominant premise if the premises are redundant [i.e., sim(a,b) is high].

The constraints outlined earlier are satisfied by Equation 2. For example, the formula implies that $\Pr(Qc \mid Qa,Qb) \approx 1$ if $\mathrm{sim}(a,c) = 1$. Note that our proposal ensures that strength increases with additional premises—that is, $\Pr(Qc \mid Qa,Qb) \geq \Pr(Qc \mid Qa)$, $\Pr(Qc \mid Qb)$. This property is desirable given the character of the predicates in our experiments; briefly, they are *noncompetitive*, in the sense that instantiation by one category does not reduce the probability of instantiation by another.

The remaining cases, including those involving negative and mixed evidence, are listed in Appendix B. They are predictable from Equation 2 by switching the direction of similarity and "the probability left behind" as appropriate.

In the experiments that follow, participants were asked to determine whether graduates of a particular college (e.g., Oklahoma State University) tended to earn a high starting salary, given that this was true of graduates of another college (e.g., Harvard University). The participants were also asked for prior probabilities and similarities for the set of colleges in the experiment. Each participant's prior and similarity judgments were then used to predict their conditionals. No parameter estimates were required. In Experiment 1, we tested SimProb for single- and two-premise conditionals involving positive evidence [cases $\Pr(Qc \mid Qa)$ and $\Pr(Qc \mid Qa,Qb)$]. In Experiment 2, we replicated the design from Experiment 1 while adding cases of negative and mixed evidence [$\Pr(Qc \mid \neg Qa)$, $\Pr(Qc \mid \neg Qa, \neg Qb)$ and $\Pr(Qc \mid Qa, \neg Qb)$]. Finally, Experiment 3 replicated Experiment 1 with biological categories and predicates.

EXPERIMENT 1
Positive Evidence

Method

Eighteen Northwestern University undergraduates participated for course credit. They read a printed instruction sheet, which listed the colleges and the salary predicate. The colleges used were Harvard University, Yale University, Connecticut State University, Arkansas State University, and Oklahoma State University. The predicate was stated as "more than 60% of the graduates earned a starting salary of more than $50,000 in their first job after college." This predicate had intrinsic interest for our undergraduate participants and proved to elicit a wide range of prior probabilities across the colleges. The instructions also provided a brief explanation of prior and conditional probabilities with an example to illustrate the distinction.

The participants provided judgments by responding to a computer program. For every question, the participants moved a slider to enter a value corresponding to an answer. The scale was anchored between 0 and 1, with intermediate values being displayed as the slider moved. Once a value was established, the participants pressed a button to register an answer and advance the trial.

Similarity block. The first task involved questions about the similarity of the colleges in the experiment. On each trial, the participants were asked to estimate "the similarity of [*A*] and [*B*]," where *A* and *B* were two colleges from the stimulus set. Across trials, *A* and *B* were assigned categories, so that a complete set of 10 unique pairwise similarity judgments were collected. The order of presentation of categories in a similarity comparison was random. The following brief instruction appeared at the top of each screen: "We now request a similarity between two items. The answer should be a number between 0 (for extremely low similarity) and 1 (for virtual identity)."

Prior probability block. For each trial of the prior probability block, the participants were asked to estimate "the probability that [*X*] has over 60% earning more than $50K." *X* varied over the five colleges in the experiment via random permutation. The participants were instructed to enter a "number between 0 (for impossibility) and 1 (for certainty)."

Conditional probability block. The third task requested conditional probability judgments of one- and two-premise statements. The instruction read, "We now request the conditional probability of one statement given another. . . ." For the one-premise trials, the participants judged "the probability that [*X*] *has over 60% earning more than $50K*, given that this is true of [*Y*]." For the two-premise trials, participants estimated "the probability that [*X*] *has over 60% earning more than $50K*, given that this is true of [*Y*] and [*Z*]." Assignment of categories to the roles of *X*, *Y*, and *Z* was constrained so that no participant provided $X \mid Y,Z$ and $X \mid Z,Y$. This yielded 20 single- and 30 two-premise conditionals. All the questions in the conditional block were randomly ordered. Across the three blocks, there was a total of 65 trials.

Results and Discussion

Similarity. Table 1 shows the mean pairwise similarity ratings collapsed across direction of comparison. We first tested for any asymmetries in the similarity data (Tversky, 1977) and did not find any. The universities typically perceived as more prestigious, Harvard and Yale, clustered together, as indicated by high similarity ratings, and appeared to differ sharply from less prestigious Arkansas State and Oklahoma State. The high similarity of the latter schools also reflected an effect of geographical proximity (both schools are located in the South of the United States). Both of these universities exhibited only moderate similarities to Connecticut State (the remaining state university; .53 and .60, respectively).

Prior probability judgments. To the extent that a predicate is nonblank, we should observe a range of prior probability judgments. As was expected, confidence in whether the salary predicate held true of a given school depended on the prestige of the university. The prestigious Harvard and Yale received high priors (.80 and .79, respectively). Arkansas and Oklahoma State received low priors (.29 and .34, respectively). Connecticut State was assigned an intermediate prior at .38.

Conditionals. The conditional probability data set consisted of single- and two-premise judgments. For clarity, we will present the data for one- and two-premise judgments separately. Table 2 shows the mean one-premise conditional probability estimates. The premise categories are listed in the rows of the table, whereas the conclusion

Table 1
Mean Similarity Ratings for the Universities in Experiment 1

University	Harvard	Yale	ConnSt	ArkSt	OkSt
Harvard					
Yale	.87				
ConnSt	.22	.30			
ArkSt	.16	.23	.53		
OkSt	.17	.18	.60	.71	

Note—ConnSt, Connecticut State University; ArkSt, Arkansas State University; OkSt, Oklahoma State University.

Table 2
Mean One-Premise Conditional Probabilities in Experiment 1

Premise	Conclusion: Harvard	Yale	ConnSt	ArkSt	OkSt	*M*	Premise Prior
Harvard		.89 (.91)	.38 (.40)	.36 (.30)	.38 (.36)	.50	.80
Yale	.93 (.92)		.41 (.42)	.33 (.31)	.35 (.36)	.50	.79
ConnSt	.94 (.84)	.94 (.84)		.60 (.53)	.64 (.63)	.78	.38
ArkSt	.95 (.83)	.95 (.82)	.73 (.66)		.71 (.73)	.84	.29
OkSt	.90 (.83)	.94 (.82)	.73 (.68)	.67 (.65)		.81	.34
M	.93	.93	.56	.49	.52		
Conclusion prior	.80	.79	.38	.29	.34		

Note—Conclusion categories are in columns, and premise categories are in rows. Numbers in parentheses are SimProb predicted values. ConnSt, Connecticut State University; ArkSt, Arkansas State University; OkSt, Oklahoma State University.

categories are listed in the columns. The "mean" estimate is the average across premise categories for a conclusion and across conclusion categories for a row. The conclusion and premise priors are included for reference. The mean estimates are helpful for seeing the main effects of premise and conclusion priors. SimProb predictions are included in parentheses and will be addressed shortly.

Several of SimProb's qualitative predictions were supported. First, there was a large effect of conclusion prior probability—that is, conclusions with higher priors received higher conditionals, and vice versa. For example, reasoning about Harvard generated a higher mean conditional than did reasoning about Oklahoma State (.93 vs. .52, respectively). Across items, the correlation between the conditional and the conclusion prior was .85.

The second finding is the effect of premise surprisingness. Low-prior (surprising) premises provided for stronger generalization of the salary predicate than did high-prior premises, as is shown by the column means. The mean conditional for the premise Arkansas State was .84, whereas the mean conditional for the premise Harvard was .50. Across items, the correlation between premise probability and the conditional was $-.64$.

The participants also appeared to be sensitive to the similarity of a pair of colleges, although analyses of means did not reveal a clear pattern consistent with this hypothesis. This is not surprising, given that priors clearly play a role in conditionals. For many pairs, premise priors were negatively correlated with similarity, thus making determining the source of the variance difficult. In order to test whether similarity made an independent contribution to conditionals, we carried out a multiple regression analysis. The dependent variable was the conditional probability provided by each participant for an item.

Each observation was predicted by linear regression from three values provided by a participant: (1) conclusion prior probability, (2) premise prior probability, and (3) premise conclusion similarity. SimProb predicts positive regression weights for conclusion prior and similarity factors and a negative regression weight for the premise prior probability factor. The weight for the premise prior factor is negative because premise surprisingness is an inverse of premise probability.

The regression model fit well ($R^2_{\mathrm{adj}} = .73$ for the single-premise analysis and .72 for the two-premise analysis, with all predictor coefficients reliably above 0). First, there was a strong positive effect of conclusion prior ($\beta = .87$). Second, there was a reliable negative effect of premise prior ($\beta = -.40$). Finally, we found a strong positive effect of category similarity ($\beta = .40$). This analysis indicated that, controlling for probability factors, people used similarity in generalizing the salary predicate. The two-premise regression estimates exhibited the same pattern as those for single-premise conditionals. Overall, SimProb appears to have the right qualitative properties for fitting the data from Experiment 1.

SimProb Fits. SimProb was fit to group means (across items), as well as to data from each participant. We also contrast SimProb with a set of baseline variants, to be described shortly. We will discuss the group fits in the text and present the individual fits in Appendix C. The individual results are the same as the group fits with respect to the differences between SimProb and the contrasting models. As a reminder, the two-premise predictions were based on the confirmation functions computed using the single-premise formulation (Equation 1). These were never computed using the one-premise estimates provided by our participants. There were no reliable effects of presentation order in our two-premise data, and we collapsed across this variable.

Since the regression analyses have shown that the conclusion prior factor was a strong predictor of conditional probabilities, we first need to demonstrate that SimProb outperforms this simple baseline. When the conclusion prior served as the only predictor, this model exhibited a correlation of .87 and a mean squared error (MS_{e}) of .054. SimProb's predictions correlated with the data at $R = .97$, and the MS_{e} was .004 (see Table 3 for model fits and Table 2 for a side-by-side comparison of single-premise judgments and model predictions). SimProb's improvement over a conclusion prior baseline was reliable [$t(59) = -6.19, p < .001$].

Table 3
Fits of SimProb and Contrasting Models to Averaged Data From Experiments 1 and 2

	Experiment 1 (1 and 2 Premises, $n = 50$)		Experiment 2 (1 Premise, $n = 24$)		Experiment 2 (2 Premises, $n = 48$)	
Model	R	MS_e	R	MS_e	R	MS_e
SimProb	.97	.004	.95	.005	.95	.009
PC−	.37	.063	.07	.077	−.40	.173
PP−	.91	.008	.78	.032	.56	.077
SIM−	.96	.007 (n.s.)	.62	.058	.40	.137

Note—MS_e is the mean squared error. All competing models have MS_es that are reliably higher than SimProb, except where indicated as n.s. PC−, PP−, and SIM− are versions of SimProb without the conclusion prior, premise prior, and similarity components, respectively.

Baseline models. In addition to SimProb, we tested a series of baseline models. The set of alternatives was designed to assess the relative contributions to the overall fit of SimProb from each of its three components (i.e., conclusion prior, premise prior, and similarity). For this purpose, we tested a set of simplified SimProb formulations in which each variant was missing one of the three main components. The general motivation behind testing a set of *reduced* SimProb models is to show that each of the qualitative constraints that motivated the model's formulation is necessary in that it improves the model's performance, relative to a formulation in which a particular variable is not considered.

For example, in order to show that the conditional is positively related to the conclusion prior, we test a version of SimProb identical in all ways to the original, but one that was missing the conclusion prior variable. Specifically, this PC− model (as in "PC minus") was formulated as

$$\Pr(Qc \mid Qa) = \mathrm{sim}(a,c)^{\Pr(Qa)}.$$

It is important to note that baseline formulations, like SimProb itself, were constructed to reflect the remaining constraints (less those concerning the variable being trimmed) and to produce reasonable behavior. For example, the PC− model is constrained so that the premise probability is inversely related to the final conditional probability. In addition to PC−, we also generated two other baseline variants. SIM− is a model that handles all but the similarity factor, whereas the PP− combines all variables except premise probability.

SimProb showed better fits (lower MS_e) than did the contrasting formulations (see Table 3). Among the alternatives, SIM− fit the best, but somewhat less well than SimProb [$t(49) = -1.48$, $p = .15$, n.s.]. Overall, we expected that SIM− would fit well because premise–conclusion probability differences tended to be confounded with similarity differences. That is, categories that differed on priors also differed on similarity. We attempted to separate these factors to a greater extent in Experiment 2. We also attributed the lack of a significant difference between SIM− and SimProb to a low number of participants. This was also remedied in Experiment 2.

The remaining contrasting models fit substantially worse than SimProb [$t(49) = -6.68$, $p < .001$, for a contrast between SimProb and PC−; $t(49) = -4.37$, $p < .001$, for the contrast between SimProb and PP−]. Taken together, the performance ordering of the baseline models suggests that the conclusion prior played the most important role in fitting SimProb, followed by premise prior and similarity. Note that this finding is consistent with the relative ordering of weights in the regression analyses.

As an aside on the difference between correlation and absolute fit measures, we find that correlations can be high despite a model's exhibiting a relatively poor fit to the data. The absolute error measure is more informative because it is better able to capture the accuracy of prediction beyond accounting for the effects of the conclusion prior. The correlation values are presented mainly to give the reader a sense of how well the models capture first-order patterns in the data (we are also not able to perform statistical comparisons of correlations computed for group fits).

EXPERIMENT 2
Negative Evidence

So far, we have considered cases in which the premise information *confirms* (increases belief in) the conclusion. Real-world reasoning often demands making inferences in situations in which the evidence is likely to disconfirm or decrease belief in the conclusion. We will refer to such cases as *reasoning from negative evidence*. In addition to considering simple cases of single-premise negative evidence [$\Pr(Qc \mid \neg Qa)$], we also examined more complex two-premise scenarios. Specifically, we considered *all-negative* cases [$\Pr(Qc \mid \neg Qa, \neg Qb)$], as well as *mixed* cases [$\Pr(Qc \mid Qa, \neg Qb)$, $\Pr(Qc \mid \neg Qa, Qb)$]. We also replicated the positive evidence findings discussed in Experiment 1.

Method

The method was the same as that in the previous study, except for three differences. First, the set of colleges used as categories was reduced from five to four to accommodate the increased number of questions resulting from the addition of negative evidence conditionals. Second, the categories used were adjusted in order to better separate similarity and priors. In the set of colleges used in Experiment 1, large dissimilarities were confounded with large differences in probabilities (e.g., Harvard and Arkansas State). In the design for Experiment 2, for some pairs, relatively high similarity was maintained despite large differences in probability of earning a high salary (e.g., Harvard University and Harvard Divinity School). Third, because of the greater number of conditionals arising from the addition of a negative predicate, one- and two-premise conditional judgments were made by different groups of participants. Both groups completed the initial priors and similarity judgments.

Ninety-two Northwestern University undergraduates participated for course credit (42 in the one-premise condition and 50 in the two-premise condition). The categories used in this study were Harvard University, Texas Tech, Harvard Divinity School, and Texas Bible College. The positive predicate read, "Graduates earned an average salary of MORE than $50,000 in their first job after college." The corresponding negative version of the predicate read, "Graduates earned an average salary of LESS than $50,000 in their first job after college." The conclusion predicate was always positive.

The procedure was identical to that in Experiment 1. The participants were first told which colleges and predicates they would be reasoning about. They then provided a set of similarity, prior probability, and conditional probability judgments (either one or two

Table 4
Similarities and Prior Probabilities for the Universities in Experiment 2

	Similarity			Prior Pr()	
	Harvard	TexTech	HarvDiv	>50k	<50k
Harvard				.81	.26
TexTech	.49			.57	.46
HarvDiv	.38	.22		.35	.64
TexBible	.17	.32	.65	.26	.70

Note—TexTech, Texas Tech University; HarvDiv, Harvard Divinity School; TexBible, Texas Bible College.

premises). The participants in the one-premise group made a total of 24 conditional judgments (12 positive and 12 negative evidence questions). Those in the two-premise group completed a total of 48 conditionals (12 positive, 12 negative, and 24 mixed).

Results and Discussion

Similarity and prior probability judgments. Table 4 shows the mean similarity ratings (scale from 0 to 1) along with mean priors assigned for the positive ("more than 50k") and negative ("less than 50k") versions of the predicate. We did not find any reliable asymmetries in the similarity judgments.

As was expected, similarity judgments were based on the relative prestige of schools and the compatibility of educational goals (e.g., religious vs. technical schools). For priors, the judgments were also based on the relative prestige of each school and mirrored each other for the two predicates.

Single-premise arguments. Table 5 shows the observed one-premise conditional probabilities for positive and negative predicates. The data for the positive predicate replicated the findings from Experiment 1. There was an effect of conclusion probability, so that higher probability conclusions yielded higher conditionals. There was also, as was expected, an effect of premise prior, so that lower priors led to higher conditionals. For example, when reasoning about Texas Tech, our participants found the fact about Harvard Divinity students earning more than the target salary to be more informative than discovering this fact about Harvard University [.67 and .57; $t(41) = 2.40$, $p < .05$]. This effect cannot be attributed to similarity, because Texas Tech was judged more similar to Harvard than to Harvard Divinity [.46 vs. .20; $t(41) = 7.59$, $p < .001$].

For the negative predicate, the participants were also sensitive to the premise prior probability, so that surprising premise facts tended to lower the conditional to a greater extent than did expected facts. For example, when reasoning about Texas Tech, the participants assigned a much lower conditional after learning that Harvard did not have the property than after finding out that Texas Bible did not have the property [.29 vs. .61, respectively; $t(41) = 6.27$, $p < .001$].

SimProb fits. Table 3 (presented earlier) shows the results of the model fits for the single- and two-premise conditionals. Across items (group data), we observed a SimProb correlation of .95 for one-premise judgments and .95 for two-premise estimates (the MS_es were low and comparable to those observed in Experiment 1). In contrast to Experiment 1, the PP− (SimProb without the premise prior component) model was the next best. The PP− model exhibited reliably higher error levels than did SimProb [$t(23) = -2.85$, $p < .05$, for one-premise arguments; $t(47) = -5.30$, $p < .01$, for two-premise arguments]. SimProb also showed reliably lower error levels than did the SIM− model [$t(23) = -2.57$, $p < .05$, for single-premise items; $t(47) = -5.81$, $p < .01$, for two-premise items]. We interpret the worse fit of the SIM− model, relative to its performance in Experiment 1, to mean that our efforts at separating probability and similarity factors were at least somewhat successful.

In this experiment, we replicated the basic findings from Experiment 1 and extended our proposed model to reasoning about arguments that involved negative evidence. As in the previous study, SimProb outperformed competing versions of the model that lack a probability component (PC− and PP−). This experiment also showed

Table 5
Conditional Probabilities for One-Premise Positive and Negative Items in Experiment 2

	Conclusion					Premise
Premise	Harvard	TexTech	HarvDiv	TexBible	*M*	Prior
		Predicate: Positive				
Harvard		.57	.37	.34	.43	.81
TexTech	.81		.45	.39	.55	.57
HarvDiv	.80	.67		.64	.70	.35
TexBible	.83	.73	.76		.77	.26
M	.81	.66	.53	.46		
		Predicate: Negative				
Harvard		.29	.18	.15	.21	.26
TexTech	.61		.28	.19	.36	.46
HarvDiv	.73	.62		.30	.55	.64
TexBible	.72	.61	.47		.60	.70
M	.69	.51	.31	.21		
Conclusion prior	.81	.57	.35	.26		

Note—TexTech, Texas Tech University; HarvDiv, Harvard Divinity School; TexBible, Texas Bible College.

that SimProb reliably outperforms its SIM− alternative (SimProb without a similarity component).

Here, we make another note about the formulations under consideration. Our rules for negative evidence are straightforward reversals of their positive counterparts, constrained by the kind of limiting cases outlined earlier. A tempting simplification is to derive the negative evidence formulas from those for positive evidence by applying basic rules of probability. This is possible only if the agent offers coherent probabilities—for example, respecting the rule that $\Pr(Qc \mid Qa) \times \Pr(Qa) \leq \Pr(Qc)$. Given that the estimates offered by our participants do not always conform to this rule, we prefer to enumerate separately the formulas for the negative cases. (For more on probabilistic incoherence, see Tentori, Bonini, & Osherson, 2004, and the references cited there.)

So far, our experiments have employed a fairly narrow range of categories. Would SimProb be able to account for reasoning in other domains? In a follow-up experiment, we addressed inductions about a set of familiar mammals.

EXPERIMENT 3
Inductions About Biological Stimuli

The goal of this experiment was to extend our theory of induction to reasoning about biological stimuli. It was possible that reasoning in the biological domain would not be captured by SimProb because prior probabilities would not be perceived as being as relevant here as they are in other conceptual areas. One potential reason is that typical U.S. undergraduates have relatively sparse folk-biological knowledge (Wolff, Medin, & Pankratz, 1999). Consequently, participants might have little confidence in their priors and might choose not to use them in reasoning. They might, instead, convert the conditional probability problems into argument strength ones and judge only the reason to believe in the conclusion, irrespective of the conclusion prior probability. If this turned out to be the case, SimProb would not be able to fit their judgments, because our model relies critically on participants' keeping track of their conclusion priors. Alternatively, if presented with a predicate that elicited prior knowledge, participants would use prior probabilities in the way specified by SimProb.

In order to test this hypothesis, we asked participants to decide whether a particular mammal species "has more than 20 chromosomes" or "has exactly 24 chromosomes." We chose these predicates because they were likely to elicit different prior probabilities. Since having 24 chromosomes entails having more than 20, the participants would be likely to treat statements associated with the "more than 20" predicate as being more likely than statements about the "exactly 24" property.

With respect to inductions, SimProb predicts that participants will generate higher conditional probabilities for arguments about a high-prior predicate than for those about a low-prior predicate. It also predicts that people will rely on similarity and premise probability information in their judgments.

Method

The method was the same as that used in the previous experiments. Thirty Northwestern University students provided conditional probability judgments about mammal stimuli. The categories were *squirrels*, *rabbits*, *lions*, *tigers*, and *elephants*. The high- and low-probability predicates were "have more than 20 chromosomes" or "have exactly 24 chromosomes." Every participant provided judgments about both predicates, although the order of the questions was random. As in the previous experiments, the set of conditionals was generated by assigning each category in the set of five stimuli to the roles of conclusion and premise (for one-premise conditionals) or conclusion, first premise, and second premise (for two-premise conditionals). There were a total of 40 single-premise conditionals and 60 two-premise conditionals. The conditionals were preceded by similarity and prior probability judgments (totals of 10 and 10, respectively). The participants were tested in groups. Printed instructions listed the five mammals and told the participants that they would be "reasoning about how many chromosomes each kind of mammal has." The three types of tasks in the experiment were briefly described.

Results and Discussion

Similarity and probability judgments. We first tested for order effects in similarity judgments; none of the tests were significant. Mean similarity judgments presumably reflected perceived taxonomic relations between the categories. Large felines (lions and tigers) and small rodents (rabbits and squirrels) received high ratings (.80 and .68, respectively). Ratings for mammals in different taxonomic groupings were low (e.g., .27 for the similarity of lions and squirrels). Elephants exhibited moderate similarity to the felines and low similarity to the rodents.

The priors varied both by predicate and by category. The high-probability predicate ("$>$20 chromosomes") had mean prior of .60, and the low-probability predicate ("exactly 24 chromosomes") had a mean prior of .41. There was also an effect of the mammal's size, so that the larger lions, tigers, and elephants were more likely to have the property than were the smaller squirrels and rabbits ($M = .54$ and .46, respectively). We submitted the prior probability judgments to a 2 (predicate) $\times$ 5 (category) repeated measures ANOVA. There was a strong main effect of the predicate, with the "more than 20" predicate receiving higher priors than the "exactly 24" predicate [$F(1,29) = 23.10$, $p < .0001$]. There was also a reliable effect of category [$F(4,116) = 8.70$, $p < .001$]. There was no interaction between predicate and category factors. Arranging categories by presumed perceived size (i.e., squirrels, rabbits, lions, tigers, and elephants) yielded a reliable linear trend [$F(1,29) = 20.24$, $p < .001$]. Together, the analyses suggest that the predicates differed in their priors and that perceived size played a reliably positive role in judgments.

Conditional probability judgments. It appears that the participants were sensitive to their priors in generating conditionals. First, the overall single-premise mean for the high-prior property ("$>$20 chromosomes") was higher than that for the low-probability property ("exactly 24 chromosomes") [$M = .70$ and .55, respectively; $t(19) = 6.9$, $p < .001$].

Similarity also played an important role in predicting judgments. For example, in reasoning about lions having exactly 24 chromosomes, tigers' having the property was the strongest piece of evidence, followed by elephants, rabbits, and squirrels, respectively. This same rank ordering applies to similarity judgments between categories. Overall, the itemwise correlation between conditionals (collapsed across predicates) and similarity was .62.

Model fits. Table 6 presents the SimProb and contrasting model fits broken down by predicate. As in the previous experiments, we will focus mainly on group data (fits for the participants are presented in Appendix C). First, SimProb provides roughly equally good fits to data for both predicates ($MS_{e>20}$ = .003 and $MS_{e=24}$ = .005; $t(49) = 1.74$, n.s.]. The correlations are in the .9 range. The MS_es for SimProb are reliably lower than those for the contrasting models. For the ">20" predicate, the closest alternative formulation to SimProb was SIM− (MS_e = .010). The difference between SimProb and SIM− was reliable [$t(49) = 4.24, p < .001$]. For the "exactly 20" predicate, the closest competitor was PP− (MS_e = .010). This error level was also reliably higher than that of SimProb [$t(49) = 4.08, p < .001$].

One goal of Experiment 3 was to see whether SimProb can fit induction with biological stimuli. As with our college stimuli in Experiments 1 and 2, SimProb exhibited an excellent fit to the conditional probabilities. Two specific predictions of the model were supported. First, the participants provided conditional estimates that revealed sensitivity to their own levels of prior belief. The participants provided higher conditional estimates for arguments about a high-prior predicate than for those about a low-prior predicate. Second, the participants showed that in addition to using the conclusion prior information, they also relied on premise prior and similarity information.

Although SimProb exhibited an equally good fit to both properties, some interesting differences between the predicates are worth noting. First, it appears that similarity played a bigger role in predicting reasoning about the "exactly 24" predicate than about the ">20" predicate. Conversely, the probability factors seemed somewhat more important for the ">20" predicate than for the "exactly 24" predicate. To illustrate the differences, consider again the model fits in Table 6. The PC− model fits better for arguments about the "exactly 24" predicate than for those about the ">20" predicate. Conversely, the SIM− model fits better for the ">20" than for the "exactly 24" predicate. One explanation for the difference is that the two predicates put different emphases on probability and similarity information. The ">20" predicate is likely to emphasize the relative ordering of animals on the chromosome dimension. This could make probability information salient, because the relative rank ordering on the chromosome dimension is likely what drives the prior probability judgments. By contrast, the exact predicate should make similarity more salient than does the relative predicate. This is likely attributable to the background knowledge people bring to the task. Specifically, participants may believe that similar species are more likely to share the exact number of chromosomes than are different species.

Table 6
Fits of SimProb and Contrasting Models to Averaged Data From Experiment 3 by Predicate

	Predicate			
	>20 Chromosomes (n = 50)		24 Chromosomes (n = 50)	
Model	R	MS_e	R	MS_e
SimProb	.89	.003	.91	.005
PC−	.53	.049	.87	.042
PP−	.76	.011	.91	.010
SIM−	.74	.010	.49	.056

Note—MS_e is the mean squared error. All the competing models have MS_es that are reliably higher than those for SimProb.

If the two predicates render similarity and probability factors differentially salient, how does SimProb fit reasoning about both? The answer lies in the formulation of the model. Specifically, SimProb's conditional probability ranges between the conclusion prior Pr(Qc) and 1.0. Keeping priors constant, similarity always operates over a range between 1.0 and 1− Pr(Qc). The larger the value of 1− Pr(Qc) [the lower the value of Pr(Qc)], the greater the range over which similarity can exert an influence. This prediction can also be conceptualized in terms of a ceiling effect on the role of similarity as the conclusion prior probability grows.

A related question is whether SimProb can handle inductions about "completely" blank predicates (e.g., the famous blank property "have sesamoid bones"). We believe that SimProb can handle such predicates, on the assumption that the overall priors for blank predicates are uniform and relatively low. When priors are low, SimProb relies on similarity to generate conditionals. The use of similarity guarantees that many of the "classic" blank property induction phenomena can be captured by SimProb. This is because similarity (or some stand-in) is the best predictor of judgments about blank predicates (e.g., Osherson et al., 1990).

GENERAL DISCUSSION

Existing theories of induction typically address reasoning with blank predicates. These types of properties, by definition, do not evoke any prior knowledge about the likelihood that categories will have them. A truly general model must also extend to nonblank properties. In order to capture the effects of property knowledge, we require a way of adequately representing levels of prior belief. We chose to represent confidence in terms of probability judgments.

We arrived at the specific formulations of the model by first considering a set of limiting-case scenarios. For example, when the similarity between the categories in a single-premise conditional approaches 1.0, so should the conditional probability. Conversely, when the similarity approaches 0.0, the conditional should remain unchanged, relative to the conclusion prior, because the premise appears to be irrelevant to the conclusion. Other limiting-case scenarios are based on probability considerations. For example, the conditional should be positively related to the conclusion prior, so that higher conclusion priors should signal higher conditionals and vice versa.

We believe that starting out with the limiting-case scenarios is a unique contribution of the present approach because our conditional probability functions are able to handle any input in a sensible way. Not all the constraints are normative. Some constraints are, indeed, normatively defensible (e.g., as the conclusion prior probability approaches 1.0, so should the conditional probability). Others may only be psychological in nature, as in the intuition that as premise–conclusion similarity approaches 0.0, the conditional should remain unchanged, relative to the prior. Some constraints may capture more abstract information-processing heuristics. For example, the constraint just mentioned may be seen as reflecting the reasoners' intuitions about the relevance of a datum (Sperber & Wilson, 1995; for a model based on normative principles, see Kemp & Tenenbaum, 2003). Overall, our results strongly supported the proposed model. SimProb performed well across different types of data sets. The data from Experiment 1 were fit with an across-item correlation of .92. The fits were equally close for judgments involving negation in Experiment 2 and when applied to biological categories in Experiment 3.

One interpretation of our findings is that the participants were under a task demand to use information about priors and similarity because they were asked for them before being invited to provide conditionals. In an experiment reported in Blok, Osherson, and Medin (2007), we replicated our findings with a between-subjects design. A group of participants evaluated the similarity/priors questions that was separate from those answering the conditional probability questions. We observed a correlation of .83 between the predictions of SimProb and the data, suggesting that a task demand alone is insufficient to explain our results.

Another limitation may be a theoretical one: How does similarity stand in for relevance? Earlier in this article, we put forth a caveat about similarity; we do not propose or specify the mechanism that connects it with relevance. We have sidestepped this issue by using predicates that can be termed as stable. A predicate is stable if its presence yields category similarities that are the same as those in its absence. In our stimulus set, the same dimensions (e.g., prestige) were relevant to both similarity and induction. Although any complete model of induction will have to move beyond stable predicates, we believe that starting out with stable predicates is challenging enough and serves as a necessary departure point. Moving beyond stable predicates will likely necessitate expanding the concept of relevance to causal relations. For example, Sloman (1997) has shown that conditional probabilities are boosted (with respect to the conclusion prior) if the premise and conclusion statements can be explained by the same causal mechanism. One way of handling such effects is to collect similarity judgments of statements, rather than just categories, on the assumption that similarity judgments reflect the compatibility of causal relations between premise and conclusion representations.

SimProb can be extended to a variety of inductive frames, as in arguments with general conclusions, for example,

Wolves have rods and cones *therefore*
All mammals have rods and cones,

and arguments involving diverse predicates, for example,

Howler monkeys will eat cheddar cheese, *therefore*
Spider monkeys will eat Swiss cheese.

Models able to capture the latter kind of inference will need to take account of predicate similarity in addition to category similarity.

As was noted earlier, our model is so far limited to monotonic predicates—those that increase belief in the conclusion relative to the prior. An interesting extension would be to properties that sometime decrease belief in the conclusion. For example, whereas determining that Dell had a profitable fourth quarter may increase the belief that HP also had a profitable fourth quarter, determining that Dell increased market share in the fourth quarter can decrease the corresponding belief about HP.

The theory presented here predicts conditional probabilities only from prior probabilities of categories having the property in question and the pairwise similarities of the categories involved in the argument. The specific formulation is constrained by limiting-case scenarios that relate extreme values of predictor variables (similarity and probability) to what the resulting conditional ought to be. Across three experiments, SimProb exhibited very close fits to observed data. Despite initial progress, serious theoretical challenges remain. Perhaps the most pressing is the consideration of predicates in which the relevant dimensions of similarity and reasoning diverge. This important endeavor is likely to be informed by advances in normative models of induction, theories of causal reasoning, and the broadening of the scope of problems considered by induction researchers.

AUTHOR NOTE

This article is based on a Ph.D. dissertation submitted to Northwestern University by the first author. The work was supported by NIH Grant MH68889 to S.V.B., NSF DLS 0132469 to D.L.M., and NSF IIS-9978135 to D.N.O. The authors thank Lance Rips, Peter Vishton, Mike Oaksford, Russ Burnett, and two anonymous reviewers for their comments on this work. Address correspondence to S. V. Blok, Department of Psychology, University of Texas, Mail Code A8000, Austin, TX 78712 (e-mail: blok@psy.utexas.edu).

REFERENCES

Blok, S. V. (2004). Modeling induction as conditional probability judgment. *Dissertation Abstracts International*, **65**, 457.

Blok, S. V., Osherson, D., & Medin, D. L. (2007). From similarity to chance. In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 137-166). Cambridge: Cambridge University Press.

Eells, E., & Fitelson, B. (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies*, **107**, 129-142.

Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge: Harvard University Press.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, **7**, 569-592.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 411-422.

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, **26**, 563-607.

KEMP, C., & TENENBAUM, J. B. (2003, July). *Theory-based induction*. Paper presented at the Twenty-Fifth Annual Conference of the Cognitive Science Society, Boston.

LO, Y., SIDES, A., ROZELLE, J., & OSHERSON, D. N. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science*, **26**, 181-206.

OSHERSON, D. [N.], SMITH, E. E., MYERS, T. S., SHAFIR, E., & STOB, M. (1994). Extrapolating human probability judgment. *Theory & Decision*, **36**, 103-129.

OSHERSON, D. N., SMITH, E. E., WILKIE, O., LOPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.

RIPS, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, **14**, 665-681.

SLOMAN, S. A. (1993). Feature-based induction. *Cognitive Psychology*, **25**, 231-280.

SLOMAN, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking & Reasoning*, **3**, 81-110.

SMITH, E. E., SHAFIR, E., & OSHERSON, D. N. (1993). Similarity, plausibility, and judgments of probability. *Cognition*, **49**, 67-96.

SPERBER, D., & WILSON, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.

TENTORI, K., BONINI, N., & OSHERSON, D. N. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science*, **28**, 467-477.

TENTORI, K., CRUPI, V., BONINI, N., & OSHERSON D. N. (2007). Comparison of confirmation measures. *Cognition*, **103**, 107-119.

TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.

WOLFF, P., MEDIN, D. L., & PANKRATZ, C. (1999). Evolution and devolution of folk-biological knowledge. *Cognition*, **73**, 177-204.

APPENDIX A
Qualitative Constraints on Positive Single- and Two-Premise Conditional Judgments

$$\Pr(Qc \mid Qa) = \begin{cases} 1.0 \text{ as } \mathrm{sim}(a,c) \to 1.0 \\ \Pr(Qc) \text{ as } \mathrm{sim}(a,c) \to 0.0 \\ 1.0 \text{ as } \Pr(Qc) \to 1.0 \\ \Pr(Qc) \text{ as } \Pr(Qa) \to 1.0 \end{cases}$$

$$\Pr(Qc \mid Qa, Qb) = \begin{cases} 1.0 \text{ as } \mathrm{sim}(a,c) \text{ or } \mathrm{sim}(b,c) \to 1.0 \\ \Pr(Qc) \text{ as } \mathrm{sim}(a,c) \text{ and } \mathrm{sim}(b,c) \to 0.0 \\ 1.0 \text{ as } \mathrm{sim}(a,c) \text{ and } \mathrm{sim}(b,c) \to 1.0 \\ \Pr(Qc \mid Qa) \text{ as } \mathrm{sim}(b,c) \to 0.0 \\ \Pr(Qc \mid Qb) \text{ as } \mathrm{sim}(a,c) \to 0.0 \\ \Pr(Qc \mid Qa) \text{ as } \mathrm{sim}(a,b) \to 1.0 \\ \Pr(Qc \mid Qb) \text{ as } \mathrm{sim}(a,b) \to 1.0 \\ \Pr(Qc \mid Qb) \text{ as } \Pr(Qa) \to 1.0 \\ \Pr(Qc \mid Qa) \text{ as } \Pr(Qb) \to 1.0 \end{cases}$$

(*Continued on next page*)

APPENDIX B
Formulas for Negative Evidence Conditionals

Single-Premise Negative Evidence Conditionals

$$\Pr(Qc \mid \neg Qa) = 1 - (1 - \Pr(Qc))^{\alpha},$$

where

$$\alpha = \left[\frac{1 - \mathrm{sim}(a,c)}{1 + \mathrm{sim}(a,c)}\right]^{1-\Pr(\neg Qa)}.$$

Two-Premise Negative Evidence Conditionals

For predicting conditionals of form $\Pr(Qc \mid \neg Qa, Qb)$ with $\neg Qa$ dominant:

$$\Pr(Qc \mid \neg Qa, Qb) = \Pr(Qc \mid \neg Qa) + [(\Pr(Qc \mid \neg Qa)) \times \mathrm{sim}(a,b) \times (\Pr(Qc \mid Qb) - \Pr(Qc))].$$

[Conditionals of form $\Pr(Qc \mid Qa, \neg Qb)$ with $\neg Qb$ dominant are treated similarly.]

For predicting conditionals of form $\Pr(Qc \mid \neg Qa, Qb)$ with $\neg Qb$ dominant:

$$\Pr(Qc \mid \neg Qa, Qb) = \Pr(Qc \mid Qb) + [(1 - \Pr(Qc \mid Qb)) \times \mathrm{sim}(a,b) \times (\Pr(Qc \mid \neg Qa) - \Pr(Qc))].$$

[Conditionals of form $\Pr(Qc \mid Qa, \neg Qb)$ with Qa dominant are treated similarly.]

For predicting conditionals of form $\Pr(Qc \mid \neg Qa, \neg Qb)$ with $\neg Qa$ dominant:

$$\Pr(Qc \mid \neg Qa, \neg Qb) = \Pr(Qc \mid \neg Qa) + [(\Pr(Qc \mid \neg Qa)) \times (1 - \mathrm{sim}(a,b)) \times (\Pr(Qc \mid \neg Qb) - \Pr(Qc))].$$

[Conditionals of form $\Pr(Qc \mid \neg Qa, \neg Qb)$ with $\neg Qb$ dominant are treated similarly.]

APPENDIX C
SimProb and Contrasting Model Fits to Data From Individual Participants

	Experiment 1 (1 and 2 Premises, $n = 50$)		Experiment 2 (1 Premise, $n = 32$)		Experiment 2 (2 Premises, $n = 50$)		Experiment 3 (1 and 2 Premises, $n = 30$)	
Model	Mdn R	MS_e	Mdn R	MS_e	Mdn R	MS_e	Mdn R	MS_e
SimProb	.67	.048	.74	.051	.62	.066	.46	.071
PC−	.37	.134	.03	.152	−.28	.309	−.04	.277
PP−	.63	.063	.57	.081	.34	.130	.45	.076
SIM−	.73	.048	.49	.104	.20	.212	.16	.167

Note—MS_e is the mean squared error. Mdn R is the median correlation across participants between predicted and observed values.

(Manuscript received April 1, 2005;
revision accepted for publication June 29, 2006.)