The role of similarity in deductive categorical inference

DUSTIN P. CALVILLO and RUSSELL REVLIN
University of California, Santa Barbara, California

The category inclusion rule specifies that categories inherit the properties of their superordinates. For example, given that *all metals are pentavalent*, it can be concluded that *all iron is pentavalent*. Sloman (1998) showed that people do not fully endorse conclusions that follow from the category inclusion rule. He claims that people rely on the similarity between the premise and the conclusion categories (*metals* and *iron*), rather than applying the category inclusion rule. By allowing reasoners to rate their certainty for category relations (e.g., *iron is metal*), as well as for conclusions, the present study shows that similarity has only an indirect effect on the certainty of conclusions: Reasoners are more certain that similar categories have a category inclusion relation, and this in turn affects the certainty of conclusions based on this relation.

It is generally believed that similarity plays an important role in cognition (e.g., Sloman & Rips, 1998), although there is no consensus on why certain concepts are considered similar (e.g., Tversky, 1977; Wallach, 1958). Some examples of the importance of similarity in cognition have come from research on categorization (e.g., Goldstone, 1994), induction (e.g., Osherson, Smith, Wilkie, López, & Shafir, 1990), hypothesis testing (e.g., Kincannon & Spellman, 2003), and analogical reasoning (e.g., Markman, 1997). Sloman (1998) suggested that similarity also plays a role in deductive reasoning, despite the fact that prominent theories of deduction, such as mental logic (Braine & O'Brien, 1998) and mental models (Johnson-Laird & Byrne, 1991), preclude any use of similarity. In the present study, we examine the role of similarity in deductive categorical inference and address the findings of Sloman (1998).

Sloman (1998) had participants rate the probability of a conclusion's truth from a given fact. He found that probability ratings were less than certain and that they varied with the similarity between the category in the conclusion and the category in the fact. He claims that it is the similarities between these categories that make some arguments appear stronger than others. We test an alternate interpretation here—namely, that the reasoners' decisions follow a rational procedure, in which similarity plays only a minor role.

Before we explain our approach, we will describe the structure of the reasoning task and will review the findings of Sloman (1998). The reasoner sees two statements: a premise presented as a fact and a conclusion to be evaluated, as shown below:

- (1) Fact: All metals are pentavalent.
Conclusion: All iron is pentavalent.

If the fact entails the conclusion, the probability of the fact should not be greater than the probability of the conclusion (e.g., Edgington, 1995). Therefore, if reasoners treat the fact as a certain truth and the conclusion follows from the fact, they should endorse the conclusion with certainty. However, in the problem above, the conclusion does not follow from the fact; it can be derived from the fact only if the tacit premise (*All iron is metal*) is made explicit. Arguments with a missing premise are known as *enthymemes*¹ (e.g., Madden, 1952) and are common in mathematical proofs (Fallis, 2003), advertising (Areni, 2002), language understanding (Revin & Hegarty, 1999), and everyday reasoning (Levi, 1995).

Sloman (1998) presented participants with enthymemes containing familiar categories, but with properties that were intended to be unfamiliar to the participants (henceforth, *blank predicates*), so that the participants could not rely on knowledge of the conclusion (as illustrated in Enthymeme 1). Sloman (1998) found that the participants did not fully endorse these conclusions, despite having previously affirmed the tacit premise (e.g., *iron is a metal*). He asked the participants in a separate questionnaire whether the category in the conclusion (*iron*) was an instance of the category in the fact (*metal*) and analyzed the conclusion ratings only for trials in which the participants affirmed this category inclusion relation.

Sloman (1998) reported two phenomena related to the importance of similarity in deduction: the *inclusion simi-*

Portions of this study were presented at the Fifth International Conference on Thinking, Leuven, Belgium, July 2004. We thank Steven A. Sloman, Ulrike Hahn, and an anonymous reviewer for comments on a previous version of the manuscript and Dana Goodgame and Susan Landgraff for assisting with data collection. Correspondence concerning this article should be addressed to D. P. Calvillo, Department of Psychology, University of California, Santa Barbara, CA 93106 (e-mail: calvillo@psych.ucsb.edu).

larity phenomenon and the *premise specificity phenomenon*. The inclusion similarity phenomenon is analogous to the *typicality effect* in semantic retrieval (e.g., Smith, Shoben, & Rips, 1974): Reasoners assign a greater probability of truth to a conclusion when it contains a more typical member of the category mentioned in the fact. Consider the following argument:

- (2) Fact: All metals are pentavalent.
Conclusion: All platinum is pentavalent.

Despite the fact that the conclusions in both Enthymemes 1 and 2 follow² from the category inclusion rule (i.e., categories inherit the properties of their superordinates), reasoners rated the conclusion in Enthymeme 1 as more probable than the conclusion in Enthymeme 2, since *iron* is a more typical *metal* than is *platinum*.

The premise specificity phenomenon is analogous to the *semantic distance effect* in semantic retrieval (e.g., Smith et al., 1974): Reasoners assign a greater probability of truth to a conclusion when it contains a category that is a direct subordinate of the category mentioned in the fact than when it is more distant. Consider the following arguments:

- (3) Fact: All flowers are susceptible to thrips.
Conclusion: All roses are susceptible to thrips.
- (4) Fact: All plants are susceptible to thrips.
Conclusion: All roses are susceptible to thrips.

Reasoners rated the conclusion in Enthymeme 3 as more probable than the conclusion in Enthymeme 4 because they perceived *roses* as more similar to *flowers* than to *plants*, since *flowers* is a more specific category than *plants*. These two phenomena were reliable across three domains: natural kinds, artifacts, and social kinds.

Sloman (1998) concluded that "people neglect category inclusion relations in reasoning about everyday categories in favor of similarity relations" (p. 22). He came to this decision because he had used only the data from the students that had endorsed the category inclusion relations (the tacit premises). Therefore, the reasoners could supply the tacit premises to make the conclusions valid; thus, reasoners should endorse the conclusions with certainty and similarity should not affect these ratings. However, in establishing whether the participants had knowledge of the tacit premises, they were allowed only to respond *yes* or *no* and could not express uncertainty in this relation (*maybe* was an option in Experiment 3). Sloman (1998) acknowledged the possibility that reasoners respond *yes* to an inclusion relation although harboring some uncertainty but dismissed this explanation as unlikely.

We agree that similarity plays a role in categorical inference. However, we propose that this is because similarity affects the reasoner's certainty for the tacit premises,

which, in turn, affects certainty for the conclusions based on these premises. For example, reasoners may be more certain that *iron* is a *metal* than that *platinum* is a *metal* and will rate conclusions based on the former relation higher than conclusions based on the latter. In the present study, we examined whether uncertainty for conclusions may result from uncertainty for tacit premises; this will provide a more refined estimate of the role that similarity plays in deductive, categorical inference. If the influence of similarity is in the process of evaluating the tacit premise in enthymemes and not in the evaluation of the conclusions, reasoners' decisions are rational and not fallacious, as Sloman (1998) suggested.

EXPERIMENT 1

We had reasoners rate the probability that conclusions were true, as well as the probability that the category inclusion relations were true. We made several predictions. First, we predicted that we would replicate the phenomena reported by Sloman (1998): that conclusions would not be fully endorsed and that endorsements would be lower when the fact and the conclusion contained less similar categories than when they contained more similar categories (i.e., we would find an inclusion similarity and a premise specificity effect). Second, we predicted that inclusion relations (i.e., the tacit premises of the enthymemes) would not be fully endorsed and that they would be endorsed less when the inclusion statements contained categories that were less similar than when they contained more similar categories. Finally, we predicted that probability ratings of the inclusion relations and ratings of conclusions would not differ, suggesting that reasoners use their knowledge of category inclusions when evaluating conclusions.

Method

Participants and Design

Forty-four undergraduates from the University of California, Santa Barbara, participated to fulfill a research requirement for an introductory psychology course. Data from 1 participant who failed to follow instructions were omitted. We employed a 2 (task: category inclusion probability or conditional probability) $\times$ 2 (phenomenon: inclusion similarity or premise specificity) $\times$ 2 (similarity: high or low) $\times$ 3 (domain: natural kinds, artifacts, or social kinds) within-subjects design.

Materials and Procedure

Conditional probability. The participants rated the probability that a conclusion was true, given that a fact was true. The problems were similar to Enthymemes 1–4 in the introduction, all containing blank predicates. The participants rated 96 conclusions on a scale from 0 (*definitely false*) to 1 (*definitely true*). Half of the conclusions were from critical arguments, and half were from filler arguments, which contained a conclusion that did not follow from the fact, as in Enthymeme 5:

- (5) Fact: All melons are in the cucurbitacea order.
Conclusion: All avocados are in the cucurbitacea order.

Category inclusion probability. We presented the participants with 96 statements that asserted that one category was a subordi-

nate of another category. These statements were the tacit premises from the enthymemes used in the conditional probability task. The participants rated the probability that these assertions were true. For example, the participants rated the probability that it is true that "roses are flowers" and the probability that it is true that "roses are plants." They used the same rating scale as in the conclusion probability questionnaire. The participants rated 96 statements, 48 containing category pairs from the critical arguments of the conclusion probability questionnaire and 48 containing category pairs from the filler arguments. All of the statements from the critical arguments were true, whereas all of the statements from the filler arguments were false, such as "avocados are melons."

Similarity ratings. We presented the participants with two pairs of categories, and they circled the pair that they considered more similar. There were 24 sets of two-pair categories. Twelve of the two-pair sets contained one pair that included a typical member of a category and its superordinate category and one pair that included an atypical member of a category and its superordinate category. For example, the participants chose the more similar of (1) iron-metal and (2) platinum-metal. The other 12 two-pair sets contained one pair that included a category and its immediate superordinate category and one pair that included a category and its more distant superordinate. For example, the participants chose the more similar of (1) roses-flowers and (2) roses-plants. This measure was included as a manipulation check for the similarity variable; we intended for some pairs of categories to be more similar than others, and this measure allowed us to determine whether the participants viewed the similarity between categories as we intended.

Half of the participants completed the three questionnaires in the order above (conclusion probability followed by category inclusion probability, then similarity ratings), and the other half completed the category inclusion probability before the conclusion probability ratings. We employed these two orders to control for possible order effects (see Sloman, 1998, Experiment 5). The participants received instructions for all three tasks prior to beginning each of them and completed questionnaires at their own pace in groups of up to 6 in a session, which lasted approximately 45 min.

Results and Discussion

Similarity Ratings

The participants chose, between two pairs of categories, the pair that they believed was most similar. There

were 24 decisions made by each participant. The participants rated the pair we designed to be more similar as the more similar pair in .817 ($SE = .013$) of the trials, which is significantly greater than if they were responding at chance levels [.50; $t(42) = 24.22, p < .001$]. The participants rated four of the pairs differently than we intended. Since the manipulation for these four pairs of items was not effective, subsequent analyses did not include items with these categories.

Probability Ratings

Each participant provided probability ratings for the 48 critical conditional probabilities and 48 critical category inclusion probabilities. The means for each condition (minus the eight categories omitted due to similarity manipulation being ineffective) are presented in Table 1. All the means in Table 1 are significantly less than 1.0 (all $ps < .001$), demonstrating that, as was predicted, the reasoners did not fully endorse conclusions or category inclusion relations. However, it is unknown whether conditional probability ratings were less than 1.0 because of uncertainty in the category inclusion relations, as we suggest, or whether the participants were failing to apply the category inclusion rule, as Sloman (1998) suggested. These two possibilities will be examined below.

To test our remaining hypotheses, we conducted a four-way ANOVA.³ Across both tasks, probability ratings of more similar items ($M = .874, SE = .010$) were higher than those of less similar items [$M = .704, SE = .012$; $F(1,42) = 241.7, p < .001$]. Although we predicted no difference in the ratings of the category inclusion and the conditional probabilities, we found that overall ratings were higher in the category inclusion probability task ($M = .805, SE = .011$) than in the conditional probability task [$M = .773, SE = .011$; $F(1,42) = 9.6, p = .003$].

Task interacted with many other variables.⁴ In an attempt to interpret these interactions, we conducted t tests

Table 1
Mean Category Inclusion and Conditional Probability Ratings (With Standard Deviations) as a Function of Similarity, Phenomenon, and Domain

	Natural Kinds		Artifacts		Social Kinds		Mean	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Inclusion similarity								
More similar								
Category inclusion	.938	.087	.895	.116	.870	.130	.901	.087
Conditional	.922	.100	.827	.144*	.700	.183*	.817	.101
Less similar								
Category inclusion	.695	.141**	.654	.144**	.694	.147**	.681	.105
Conditional	.706	.155**	.620	.147**	.672	.149	.666	.111
Premise specificity								
More similar								
Category inclusion	.824	.177	.942	.104	.927	.114	.898	.095
Conditional	.814	.164	.924	.105	.904	.115	.878	.087
Less similar								
Category inclusion	.805	.154	.753	.187**	.659	.206**	.739	.101
Conditional	.809	.192	.746	.189**	.632	.235**	.741	.116

*Conditional rating is significantly less than the corresponding category inclusion rating ($p < .05$). **Mean rating of less similar items is significantly less than corresponding mean rating of more similar items ($p < .05$).

to examine the consistency of the task effect (differences between category inclusion and conditional probability ratings) in each of the 12 combinations of the other variables (phenomenon, similarity, and domain). The single asterisks in Table 1 show the 2 conditions in which conditional probability ratings differed from their category inclusion probability counterparts. We also examined the consistency of the effect of similarity by comparing ratings of more similar and less similar items across the 12 conditions of the other variables (task, phenomenon, and domain). The double asterisks in Table 1 show that similarity had an effect in 9 of the 12 conditions. These analyses show that similarity had a relatively consistent effect, whereas differences between category inclusion and conditional probability ratings rarely occurred.

EXPERIMENT 2

Experiment 1 showed that ratings of category inclusion and conditional probabilities are similar. Sloman (1998) made the claim that judgments of conditional probability are based on similarity, whereas we claim that these judgments result from uncertainty in underlying category inclusion relations. To better test these two possibilities, Experiment 2 allowed similarity to be rated on the same scale as category inclusion certainty and conditional probability ratings so that we could determine which is the better predictor of conditional probabilities (similarity or category inclusion similarity).

Method

Participants and Design

Thirty-two undergraduates from the University of California, Santa Barbara, participated to fulfill a research requirement for an introductory psychology course. Data from 2 participants who failed to follow instructions were omitted. The design matched that in Experiment 1, with the addition of a similarity condition to the task variable.

Materials and Procedure

Conditional probability. This task was similar to that in Experiment 1, but only the items used by Sloman (1998) were included as critical items in this experiment (this was a subset of the items used in Experiment 1). These 36 arguments were embedded in a set of 36 filler items. The scale for making these ratings was changed from a range of 0–1 to a range from 0 (*definitely false*) to 100 (*definitely true*), because this may be an easier scale to use in making judgments (e.g., Gigerenzer & Hoffrage, 1995).

Category inclusion certainty. This task was also similar to that in Experiment 1, with three exceptions. First, only 36 critical and 36 filler items were used. Second, the participants were instructed to rate their certainty for these statements, rather than the probability that these statements were true. Finally, the scale was changed from 0–1 to 0–100.

Similarity ratings. We presented the participants with 72 pairs of categories and asked them to rate how similar these were, on a scale from 0 to 100. These pairs were the categories used in the other questionnaires (e.g., *iron–metal*). We provided no definition of similarity.

All the participants completed the three questionnaires in the order above. The participants received instructions for all three tasks prior to beginning any of them and completed questionnaires

at their own pace in groups of up to 7 per session, which lasted approximately 45 min.

Results and Discussion

Task Ratings

Mean ratings for the three tasks for each condition are presented in Table 2. Excluded from these means are the categories in which the participants believed that an inclusion relation did not exist. For example, if a participant believed that *iron* was not a *metal*, all ratings of *iron–metal* for this participant would be excluded. To be consistent with Experiment 1, we analyzed the consistency of the effects of task [$F(2,58) = 10.85, p < .001$] and similarity [$F(1,29) = 50.02, p < .001$]. For task, post hoc analyses (Sidak) revealed that ratings of conditional probability ($M = 82.9, SE = 1.8$) were less than those of similarity ($M = 87.8, SE = 1.1$) and category inclusion certainty ($M = 88.7, SE = 1.0$; both $ps < .01$). The asterisks in Table 2 show the 6 of the 12 conditions in which there existed a difference between the category inclusion, similarity, and conditional probability ratings. Five of these reflect differences in category inclusion and conditional probability ratings. The pluses in Table 2 show the conditions in which more similar items received higher ratings than did less similar items.

Multiple Regression Analysis

A multiple regression analysis was conducted to determine whether the relationship between similarity and conditional probability certainty is mediated by category inclusion certainty. Sloman (1998) claimed that reasoners neglect category inclusion information and, instead, rely on similarity when judging the probability of conclusions. This is expressed as the similarity-alone model in Figure 1. We predicted that the effect of similarity in evaluating the probability of conclusions would be mediated by certainty for the underlying category inclusion relation. This is expressed as the similarity-mediated model in Figure 1.

For each of the 36 pair of categories, the participants provided ratings for similarity, category inclusion certainty, and conditional probability. Correlations for these three ratings are presented in Table 3. We examined the similarity-alone model by conducting a regression analysis with similarity as a predictor and conditional probability ratings as the criterion. Similarity accounted for 64.2% of the variance in conclusion endorsements ($p < .001$). Next, we examined the similarity-mediated model by conducting a multiple regression analysis with similarity and category inclusion certainty as predictors and with conditional probability again as the criterion variable. These two predictors together accounted for 67.8% of the variance in conditional probability judgments ($p < .001$). The effect of similarity on conditional probability was no longer significant when category inclusion certainty was simultaneously considered in the similarity-mediated model (see Figure 1), demonstrating that the effect of similarity on conditional probability ratings is mediated by category inclusion certainty. That is, similar-

Table 2
Mean Category Inclusion Certainty, Similarity, and Conditional Probability Ratings
(With Standard Deviations) as a Function of Similarity, Phenomenon, and Domain

	Natural Kinds		Artifacts		Social Kinds		Mean	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Inclusion similarity								
More similar								
Category inclusion	96.0	7.7	93.2	7.7	87.7	12.0	92.3	6.0
Similarity	94.7	7.0	94.4	8.0	83.2	17.1	90.7	7.9
Conditional	90.4	14.8 ^a	91.5	11.1	76.0	19.0 ^{ab}	86.0	11.2
Less similar								
Category inclusion	87.8	12.4 [†]	81.3	13.7 [†]	75.6	12.5 [†]	81.6	7.9
Similarity	86.7	12.8 [†]	84.6	13.7 [†]	73.2	15.8 [†]	81.6	9.7
Conditional	84.3	16.3	76.3	20.2 ^{ac†}	63.1	24.1 ^{ab†}	74.6	14.6
Premise specificity								
More similar								
Category inclusion	94.6	7.7	92.9	10.6	90.6	9.5	92.7	6.0
Similarity	91.0	10.8	89.9	12.1	92.2	9.3	91.0	6.7
Conditional	89.3	13.0 ^{ab}	93.7	10.4	85.1	24.5	89.4	10.9
Less similar								
Category inclusion	88.4	12.2 [†]	88.8	14.3	87.6	11.3	88.3	8.9
Similarity	87.2	14.1	88.2	13.2	88.7	13.2	88.0	10.0
Conditional	89.7	11.8	81.5	21.0 [†]	74.1	23.5 ^{at}	81.8	11.6

^aConditional rating is significantly less than corresponding similarity and category inclusion rating (Sidak post hoc analyses, $p < .05$). ^{ab}Conditional rating is significantly less than corresponding category inclusion rating, but not the similarity rating (Sidak post hoc analyses, $p < .05$). ^{ac}Conditional rating is significantly less than the corresponding similarity rating, but not the category inclusion rating (Sidak post hoc analyses, $p < .05$). [†]Mean rating of less similar items is significantly less than the corresponding mean rating of more similar items ($p < .05$).

ity predicts category inclusion certainty, which predicts conditional probability ratings.

GENERAL DISCUSSION

Reasoners' endorsements of conclusions appear to be weaker than required by the category inclusion rule. As in Sloman (1998), the participants in the present study more strongly endorsed conclusions when the category of the conclusion was similar to the category of the premise. This has led to the claim that reasoners employ similarity as the sole criterion in making their endorsements of a conclusion and ignoring inclusion relations. However, by allowing reasoners in the present study to estimate the probability of tacit premises, as well as the probability of conclusions, we found that, in most cases, these two probabilities were equivalent. We propose that it is the uncertainty for the tacit premises that leads to uncertainty for the conclusions.

The regression analysis showed that similarity has only an indirect effect on conclusion endorsements, an effect that is mediated by category inclusion certainty. Baron and Kenny (1986) have presented four criteria for proposing that a variable mediates an outcome. All four criteria were met in the present study; similarity was correlated with conclusion ratings; similarity was correlated with category inclusion ratings; category inclusion ratings affected conclusion ratings when similarity was also included as a predictor; and similarity no longer affected conclusion ratings when category inclusion was also included as a predictor.

We adopt the approach suggested by Grennan (1994) to calculate the probability of an enthymematic conclusion. He claimed that the probability of an enthymematic conclusion should be greater than or equal to the product of the probability of the presented premise and the probability of the tacit premise. When presented as a fact, the explicit premise is assumed to be true and to have a probability of 1.0. The probability of the conclusion should, therefore, be greater than or equal to the probability of the tacit premise. For illustration, reconsider Enthymemes 1 and 2. Endorsements of the conclusions in Enthymemes 1 and 2 should depend on one's certainty for the tacit premises (i.e., *iron is metal* and *platinum is metal*, respectively). If one is more certain that iron is metal than that platinum is metal, one could rate the conclusion in Enthymeme 1 as more probable than the conclusion in Enthymeme 2.

Furthermore, if these tacit premises are made explicit in the argument, a reasoner should fully endorse both of these conclusions. Support for this prediction comes from Sloman's (1998) fourth experiment. He presented the category inclusion relation as a premise (this was tacit in all the other experiments). For example, *all iron is metal* was included as a fact, along with the other fact and the conclusion in Enthymeme 1. When this information was presented as a premise, the reasoners fully endorsed the conclusion. He claims that this occurs because "inclusion relations are not always used because they are not always accessed during reasoning" (p. 19). We believe that providing the category inclusion relation as a premise removes uncertainty for this relation, which, in turn, removes uncertainty for the conclusion. More generally,

A) Similarity-Alone Model

B) Similarity-Mediated Model

Figure 1. Two models of the relationship between similarity and conditional probability ratings: (A) similarity alone and (B) similarity mediated (both with standardized path coefficients). * $p < .05$. *** $p < .001$.

making a tacit premise explicit removes any uncertainty for this proposition.

This approach of determining the probability of conclusions applies only to conclusions with blank predicates and only to arguments in which the conclusion category is a member of the premise category. If nonblank predicates are used, we would expect reasoners' knowledge of the conclusion to affect their certainty for its truth. For example, if one knows that iron is pentavalent, rating the conclusion would not rely on the certainty for the tacit premise. Even without the knowledge that iron is metal, the conclusion in Enthymeme 1 could be strongly endorsed. Furthermore, if one believes that the conclusion category is not a member of the premise category, the category inclusion rule would not apply. These arguments lose their deductive quality, and reasoners may rely on inductive strategies, such as similarity relations (e.g., Rips, 1975; Sloman, 1993). We believe that the arguments in the present study are deductive in nature and that reasoners apply the category inclusion rule to determine conditional probabilities, whereas arguments in which no rule applies require inductive reasoning, which is a separate process (e.g., Rips, 2001). This is a less parsimonious explanation than Sloman's (1993, 1998), which posits the same process for both types of argument, but it provides a better fit of the data in the present study.

Sloman and Rips (1998) have discussed four possible roles for similarity (as opposed to rule-based processing) in cognition: strong, weak, feeble, and none. They claim that similarity probably plays a weak or feeble role in most cognitive processes, in which reliance on similarity is context dependent (weak) or similarity is just a shorthand way of discussing property-based cognitive processing (feeble). However, Sloman (1998) appears to suggest that similarity plays a strong (primitive and fundamental) role in categorical inference. Our findings indicate that similarity between categories appears to affect certainty for category inclusion relations, which determines conclusion certainty. Similarity's effect is only superficially part of the inference process, since similarity affects certainty for knowledge that supports a conclusion when that information is tacit in the argument. Therefore, the role of similarity in deductive categorical inference is not strong but feeble, since reasoners make these inferences primarily by rule: They determine the probability of a conclusion by considering the product of the probabilities of the stated and the unstated premises.

REFERENCES

- ARENI, C. S. (2002). The proposition-probability model of argument structure and message acceptance. *Journal of Consumer Research*, **29**, 168-187.
- BARON, R. M., & KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality & Social Psychology*, **51**, 1173-1182.
- BRAINE, M. D. S., & O'BRIEN, D. P. (1998). *Mental logic*. Mahwah, NJ: Erlbaum.
- EDGINGTON, D. (1995). On conditionals. *Mind*, **104**, 235-329.
- FALLIS, D. (2003). Intentional gaps in mathematical proofs. *Synthese*, **134**, 45-69.
- GEORGE, R. (1972). Enthymematic consequence. *American Philosophical Quarterly*, **9**, 113-116.
- GIGERENZER, G., & HOFFRAGE, U. (1995). How to improve Bayesian

Table 3
Correlations Between Category Inclusion Certainty,
Similarity Rating, and Conditional Probability for
Categories in Experiment 2

	1	2	3
1. Category inclusion certainty	—	.887**	.819**
2. Similarity rating	—	—	.801**
3. Conditional probability	—	—	—

** $p < .01$.

- inference without instruction: Frequency formats. *Psychological Review*, **102**, 684-704.
- GOLDSTONE, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, **52**, 125-157.
- GRENNAN, W. (1994). Are "gap-fillers" missing premises? *Informal Logic*, **16**, 185-196.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.
- KINCANNON, A., & SPELLMAN, B. A. (2003). The use of category and similarity information in limiting hypotheses. *Memory & Cognition*, **31**, 114-132.
- LEVI, D. S. (1995). The case of the missing premise. *Informal Logic*, **17**, 67-88.
- MADDEN, E. H. (1952). The enthymeme: Crossroads of logic, rhetoric, and metaphysics. *Philosophical Review*, **61**, 368-376.
- MARKMAN, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, **21**, 373-418.
- OSHERSON, D. N., SMITH, E. E., WILKIE, O., LÓPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.
- REVLIN, R., & HEGARTY, M. (1999). Resolving signals to cohesion: Two models of bridging inference. *Discourse Processes*, **27**, 77-102.
- RIPS, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, **14**, 665-681.
- RIPS, L. J. (2001). Two kinds of reasoning. *Psychological Science*, **12**, 129-134.
- SLOMAN, S. A. (1993). Feature-based induction. *Cognitive Psychology*, **25**, 231-280.
- SLOMAN, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, **35**, 1-33.
- SLOMAN, S. A., & RIPS, L. J. (1998). Similarity as an explanatory construct. *Cognition*, **65**, 87-101.
- SMITH, E. E., SHOBEN, E. J., & RIPS, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, **81**, 214-241.
- SORENSEN, R. A. (1988). Are enthymemes arguments? *Notre Dame Journal of Formal Logic*, **29**, 155-159.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.
- WALLACH, M. A. (1958). On psychological similarity. *Psychological Review*, **65**, 103-116.

NOTES

1. It has been argued that enthymemes are not arguments at all (Sorensen, 1988), but for the present discussion, we will consider enthymemes to be a special type of argument: a syllogism with a missing premise.
2. Neither conclusion follows logically, but both are enthymematic consequences (e.g., George, 1972).
3. Order was a control variable, and a five-way ANOVA revealed that it neither exercised a main effect nor interacted with any other variables in this study, so the two orders were combined for all the analyses.
4. The interactions were task $\times$ similarity [$F(1,42) = 11.3, p = .002$], task $\times$ phenomenon [$F(1,42) = 9.03, p = .004$], task $\times$ domain [$F(2,84) = 8.05, p = .001$], task $\times$ phenomenon $\times$ similarity [$F(1,42) = 5.50, p = .024$], task $\times$ phenomenon $\times$ domain [$F(2,84) = 3.42, p = .037$], and task $\times$ phenomenon $\times$ similarity $\times$ domain [$F(2,84) = 4.15, p = .019$].

(Manuscript received May 20, 2004;
revision accepted for publication March 2, 2005.)