

# The belief-bias effect in formal reasoning: The influence of knowledge on logic

R. REVLIN, V. LEIRER, H. YOPP, and R. YOPP  
*University of California, Santa Barbara, California 93106*

The present study examines the applicability of a rational model of categorical inference (e.g., Revlis, 1975b) to account for the apparently irrational decisions students reach on categorical syllogisms. In Experiment 1, students judged the logical validity of emotionally neutral conclusions to controversial premises. Of the reasoners' decisions, 80% can be accounted for by the application of rational rules to their idiosyncratic encoding of the syllogistic premises. In Experiment 2, students were asked to solve syllogisms whose conclusions varied in truth value. When asked to reason about controversial, if not emotional, material, students do not suspend rational choice, but rather, their decisions are judicious ones, flowing logically from their idiosyncratic understanding of the materials reasoned about. When errors do occur, they result from an interrupt to rational processes and reflect conflict between competing goals rather than a switch to irrational decision processes.

Categorical reasoning, as exemplified by the Aristotelian syllogism, has served as a standard of rationality for centuries. It is not surprising, therefore, that models developed recently to account for such reasoning also reflect rational decision rules (e.g., Erickson, 1978; Johnson-Laird & Steedman, 1978; Revlin & Leirer, 1978). These models posit basic inferential processes that determine the decisions reached by the problem solver on a variety of materials. Their approach is similar, insofar as they do not characterize students as erratic decision makers. These models are congruent with the general cast of information processing models in which apparently irrational outputs are a consequence of interrupts to a serial processor, rather than a result of a switch to a nonrational system during problem solving (e.g., Simon, 1967). This view of the rational reasoner has had a considerable measure of success in accounting for decisions resulting from reasoning with neutral materials (e.g., letters that stand for categories, or common category names; Revlin, Ammerman, Petersen, & Leirer, 1978). However, such models have not been evaluated in those special situations in which human rationality is often called into question, occasions when people reason about controversial relations and appear to make capricious and irrational decisions. The present study examines the viability of the "rational" reasoner approach and assesses the applicability of at least one of the rational models of reasoning, the conversion model (Revlis, 1975b), for its ability to account for apparently capricious decisions. In so doing, this work addresses the question of whether

rational models may offer plausible accounts of irrational as well as rational behavior.

The notion of a rational reasoner has been challenged in the social and clinical psychology literature, in which it is frequently concluded that untrained reasoners are not strictly logical in their inferences and that they base their decisions primarily on personal knowledge and biases (McGuire, 1960; Morgan & Morton, 1944; Thistlethwaite, 1950; Winthrop, 1946). For example, when syllogisms such as the following are solved, the content of the conclusions is said to have an effect on the reasoner's assessment of the validity of the overall argument: "(1) All Russians are Bolsheviks. *Some Bolsheviks are undemocratic people.* Therefore: (a) All undemocratic people are Russian. (b) No undemocratic people are Russian. (c) Some undemocratic people are Russian. (d) Some undemocratic people are not Russian. (e) None of the above is proven."

That is, when solving such syllogisms, students do not appear to base their judgments on the logical form of the arguments; instead, they appear to base their judgments on the believability of the conclusions (e.g., Feather, 1965; Gordon, 1953; Janis & Frick, 1943; Janis & Terwilliger, 1962; Kaufman & Goldstein, 1967; Lefford, 1946; Morgan & Morton, 1944; Wilkins, 1928; Winthrop, 1946). In the syllogism above, students are claimed to accept Conclusion c rather than Conclusion e, the logically required answer, because they are supposed to believe that Russians are undemocratic people.

The conversion model of syllogistic reasoning provides an alternative interpretation of these data that will be examined here, namely, that deductive errors on categorical syllogisms are only indirectly affected by the statement's truth value and that decisions do not reflect insufficiencies in the reasoner's logical skills. For

This research was supported in part by a grant from the National Science Foundation (BNS78-24763) to R. Revlin. The authors wish to express their appreciation to C. Clifton and the anonymous reviewers whose insightful comments have helped us to clarify our work.

example, one possible explanation of errors in Syllogism 1 is that the proposition "All Russians are Bolsheviks" is converted in the process of comprehension so that the reasoner understands it to mean that not only are all Russians Bolsheviks, but it is also the case that all Bolsheviks are Russians. If this converted interpretation is the first one employed by the reasoner when considering the composite meaning of the two premises, then the conclusion "Some undemocratic people are Russian" is a valid deduction from the two premises of the syllogism.

### Conversion

The conversion model directs our attention to the encoding process. It makes the strong claim that a conversion operation participates in the encoding of each quantified, categorical expression: When the reasoner is told "All A are B," he or she interprets this proposition to mean that both the intended inclusion relation and its converse, "All B are A," are true. The presence of conversion in encoding is found in the verification errors of adults (Revlis & Leirer, 1980), is consistent with the developmental sequence in the comprehension of quantified, categorical expressions described by Bucci (1978), and may reflect a heuristic assumption of symmetry in "is a" relations (i.e., "All A are B" and "All B are A"; cf. Tsai, 1977). Conversion as a source of errors in syllogistic reasoning was suggested by Chapman and Chapman (1959) and has been embodied in a formal, testable model by Revlis (1975a, 1975b). The importance of conversion for categorical inference is illustrated by comparing Syllogisms 2 and 3. "(2) All P are M. *Some M are S*. Therefore: (a) All S are P. (b) No S are P. (c) Some S are P. (d) Some S are not P. (e) None of the above is proven." "(3) All M are P. *Some S are M*. Therefore: (a) All S are P. (b) No S are P. (c) Some S are P. (d) Some S are not P. (e) None of the above is proven."

A student reasoning logically on Syllogism 2 should claim that no valid conclusion is possible (Conclusion e). However, if, while encoding the premises, the reasoner's understanding of the premises is such that he/she converts each one in turn, the problem would appear as Syllogism 3. This converted syllogism does have a solution (Conclusion c), "Some S are P," which is just the conclusion that reasoners accept when shown Syllogism 2. As a result of conversion, a new problem is produced with a conclusion that is inappropriate for the original syllogism.

It should be noted that the conclusions drawn as a consequence of a converted encoding will not always be inappropriate. Consider the following: "(4) No P are M. *Some M are S*. Therefore: (a) All S are P. (b) No S are P. (c) Some S are P. (d) Some S are not P. (e) None of the above is proven." "(5) No M are P. *Some S are M*. Therefore: (a) All S are P. (b) No S are P. (c) Some S are P. (d) Some S are not P. (e) None of the above is proven."

Although Syllogism 4 is transformed into Syllogism 5 by virtue of conversion, the appropriate conclusion is not altered: "Some S are not P" is the logically required conclusion to both syllogisms. When decisions are scored without independent reference to a reasoner's understanding of the premises, Syllogism 4 will be scored as correct and Syllogism 2 will be scored as incorrect. Here we see that conversion of some syllogisms will lead to a formally (logically) correct conclusion and conversion of other syllogisms will lead to a formally incorrect conclusion.

Revlis (1975a, 1975b) constructed a model of reasoning with conversion as well as other processing mechanisms at its core. This model has been shown to be effective in predicting the reasoner's decisions on both abstract problems (in which letters represent categories) and concrete ones (in which real-world categories are used). It specifies precisely those conditions under which knowledge of the categories reasoned about will affect (both positively and negatively) the deductions reached. The model claims that students will be correct in their judgments in either of two conditions: (1) when the premises are converted, but the conclusion is fortuitously the same in the converted and original forms of the problem (e.g., Syllogism 4 above), and (2) when the reasoner's knowledge of the world blocks conversion (e.g., if the student were told "All horses are animals," it is unlikely that he/she would conclude that the converse is true, i.e., "All animals are horses"). Students will accept incorrect conclusions primarily in those cases in which conversion during premise encoding transforms the syllogism into one with a conclusion different from that prescribed by logicians (e.g., Syllogism 2 above). In summary, every decision, whether logically correct or not, reflects premise encoding processes. The model addresses the issue of apparent irrationality by claiming that students make errors in reasoning when their encoding of the premises transforms the syllogism into another problem with a logical conclusion different from the original. In such cases, the reasoner's deduction is scored as an error only because it deviates from the prescribed conclusion and not because it is not rationally derived.

While the notion of premise conversion has been useful in accounting for decisions with neutral, abstract, and concrete materials, it remains an empirical question whether it can account for the ostensibly irrational decisions made by reasoners who are asked to make decisions concerning "emotional" or prejudicial materials. Since neither the conversion model nor any of the class of such process models can posit illogical inference rules, they can be supported only if it is shown that the "irrational" decisions follow from the way materials are encoded by the reasoner. In other words, the influence of personal bias on categorical reasoning must be a consequence of the reasoner's understanding of the premises and not due to a suspension of rationality.

The approach taken here characterizes the reasoner as

possessing logical processes (as yet unspecified) that are isomorphic to those used by a logician working with standard, first-order (quantificational) calculus. Discrepancies between decisions reached by the logician and decisions reached by the present subjects are not a consequence of logical rules per se, but of the quite different readings given to the propositions to be reasoned about.

To extend the assumptions of the conversion model to categorical reasoning with controversial materials, the present study examined whether students' decisions can be predicted from their encoding of the premises (independently assessed). This study also examined the model's claim that when encoding is controlled, reasoners will accept the logically prescribed conclusion, independent of its empirical truth value. These objectives were part of the rationale for Experiments 1 and 2, respectively.

### EXPERIMENT 1

The conversion model directs our attention not to the believability of the conclusions, but to the personal encodings of the premises of the syllogism. When the premises contain information about categories that are already available to the reasoner, long-term memory (LTM) may provide working memory with more information than may have been contained or intended in the presented material (cf. Scribner, 1977). If one accepts the notion that the meaning of terms is, at least in part, determined by the immediate relations into which those terms enter in LTM (cf. Collins & Quillian, 1969), then even a simple semantic reading of the premises might entail "awareness" of relations among categories not specifically called for in the reasoning problem. Hence, the reasoner makes his/her judgments based on too much information and not on the specific content of the problem.

One way in which this may be manifested is in terms of the kinds of immediate inferences that the reasoners are willing to make when presented a proposition. For example, when shown the abstract relations, "All A are B," reasoners are quite willing to infer that all B are A. But, as stated earlier, they may block conversion in some concrete relations as a function of the pragmatic implications of the relations (Revlín et al., 1978). It is this kind of use of LTM or personal knowledge that the conversion model claims will affect the validity judgments on categorical syllogisms.

To test the hypothesis that the belief-bias effect is related to the personalized representation of the premises of a syllogism (and not to faulty inference rules), students were asked to solve syllogisms of the kind shown below: "(6) No black people in Newton are residents of Sea Side. *All black people in Newton are welfare recipients.* Therefore: (a) All welfare recipients in Newton

are residents of Sea Side. (b) No welfare recipients in Newton are residents of Sea Side. (c) Some welfare recipients in Newton are residents of Sea Side. (d) Some welfare recipients in Newton are not residents of Sea Side. (e) None of the above is proven." The premises of these syllogisms had real-world truth values (of a sort) and were of a controversial nature. However, the conclusions were quite neutral. In this way, we felt we could isolate the effects of premise encoding from response biases in the selection of conclusions.

We acknowledge that while the reasoners have no prior information about, for example, a mythical community of Newton and its residents, they may have knowledge about the categories of "blacks" and "welfare recipients." Indeed, it was expected that some of the reasoners would actually believe the converse of the major premises in Syllogism 6: "All welfare recipients are black." If so, they should readily accept Conclusion b, since it logically follows from this reading of the premise.

### Method

**Materials. Syllogisms.** Two groups of students ( $n = 13$ ) were asked to solve 16 categorical reasoning problems. Half of the problems were valid syllogisms; they are ones for which a single conclusion unambiguously follows from the premises (e.g., Syllogism 4 above) (in standard terminology they are EI-1, EI-2, EA-3, and EA-4).<sup>1</sup> Half of the problems were invalid syllogisms; they are ones for which no conclusion unambiguously follows from the premises (e.g., Syllogism 2 above) (they are IE-1, AO-1, AE-1, and AI-2). All but two of these problems were used by Revlin et al. (1978). The valid syllogisms were (1) "sames," on which the conversion model claims the reasoner will select the logical conclusions independently of whether the premises are converted (Syllogism 4 above), and (2) "differents," on which the model claims that the reasoner will always accept an erroneous conclusion because he/she has converted the premises of the syllogisms (e.g., Syllogism 6). The decisions reached on sames will always be found to be correct; those reached on differents will always be found to be incorrect.

Invalid syllogisms were composed of same-N and different-N problems. For invalid syllogisms, the logically required decision is "No conclusion is proven." The model distinguishes between two types of invalid syllogisms: (1) those for which conversion transforms the syllogism into one with a conclusion different from that which would be prescribed by a logician (these are called different-N syllogisms; e.g., Syllogism 2), and (2) those for which conversion produces a syllogism requiring the same conclusion is "None of the above"; these problems are called same-N syllogisms). On different-N syllogisms (as with valid differents), the reasoner will accept the conclusion that matches the encoded problem, and the conclusion will be scored as incorrect. For same-N syllogisms, the converted problem requires a "none" conclusion that evokes a "double-checking" procedure in which the reasoner reworks the problem with less derived interpretations of the premises (this is analogous to the effect of matching incongruity so familiar in sentence-picture verification models; e.g., Chase & Clark, 1972). Due to the time constraints in these tasks (60 sec/problem), students are predicted to be unable to complete the processing on such problems and are said to make a fair guess from among the alternatives. As a consequence, reasoners will accept the logically prescribed conclusion ("none") on only 20% of the problems (when there are five alternatives to select among, as in the present study). A fuller description of this mechanism is provided in Revlin (1975a).

**Table 1**  
**Potentially Convertible Relations**

1. All white people are bigots.
2. All black people are welfare recipients.
3. No intellectuals are athletic people.
4. All homosexuals are hairdressers.
5. All unconventional people are artistic.
6. All women are bad drivers.
7. Some members of large families are Catholics.
8. All women are nurses.
9. All old people are social security recipients.
10. Some quiet people are Japanese-Americans.
11. All college students are marijuana users.
12. All Republicans are upper class people.
13. All only-children are spoiled.
14. All political conservatives are policemen.
15. No secretaries are poor typists.
16. All shrewed money managers are Jews.

The syllogisms were similar to those already shown: two premises followed by five alternatives from which the students were asked to make their selections. The subject of each conclusion was always a category mentioned in the second premise, and the predicate of the conclusion was always a category mentioned in the first premise. The syllogisms were randomly ordered in a booklet, with the restriction that runs of three or more problem types were not permitted (e.g., valid-invalid, same-different).

*Questionnaire.* The syllogistic premises consisted of relations between concrete categories. Beliefs in these relations were assessed by a 60-item questionnaire. The questionnaire examined the student's beliefs about 30 category pairs. For example, students were asked what percentage of college-educated people are professionals and (30 items later) what percentage of professionals are college educated. Topics ranged from innocuous category relations (e.g., college educated/professionals) to controversial ones (e.g., black Americans/welfare recipients). Students were asked to answer each question on a 5-point scale that embodied the degree of category relation that they felt was appropriate for each of the 60 relations: (1) 0%-15%, (2) 15%-40%, (3) 40%-60%, (4) 60%-85%, and (5) 85%-100%. The students were assured that their answers would be kept confidential and were assigned tally numbers that they placed on their questionnaires and the reasoning test.

Only 16 of the relations were used in the test; they were selected on an a priori basis as controversial and in keeping with the belief-bias literature. A list of the category relations is shown in Table 1. By putting together the students' answers to these questions (see Results), we were able to characterize each reasoner as either a converter or a nonconverter for each syllogism and to make predictions concerning the answers that would be logically dictated by the hypothesized encoding of the reasoner.

*Procedure.* One group of students was asked to complete the questionnaire before seeing the reasoning task; the second group completed the questionnaire following the task. All students were instructed on the rules of solving categorical syllogisms and were allowed as much time as they needed to complete the questionnaire. They were limited to 60 sec to solve each reasoning problem.

*Subjects.* The subjects were 26 men and women fulfilling a course requirement. They were run in two groups, in sessions lasting 45 min.

**Results**

The reasoners' accuracy was calculated for each problem type (valid same and different, and invalid

same-N and different-N). The results will be evaluated in two parts. First, the overall reasoning performance will be examined, to assess whether the ordinal predictions of the model hold. Second, we will determine whether reasoners' decisions can be predicted from their encoding of the syllogistic premises (as measured by a questionnaire) and whether such findings can inform us concerning the belief-bias effect.

The reasoning accuracy of students is presented in Table 2, which shows that the ordinal predictions of the conversion model were affirmed. Reasoners' decisions were more accurate for sames than for differents [ $F(1,24) = 46.2, p < .001$ ], and the decisions were as inaccurate on invalid different-N problems as on valid different problems, which is congruent with the model's claim that the same processes are entailed in both sorts of judgments. Invalid same-N problems were no more accurately solved than different-N problems; neither significantly differed from chance, which may have been due to a floor effect, since the model claims that students will guess on same-N problems and be minimally accurate on different-N problems. No difference on any of these measures was found between the two treatment groups (questionnaire first, questionnaire second), and there was no interaction between groups and problem type.

The overall reasoning accuracy in this study fell short of previous ones. For example, Revlis (1975b) found that reasoners were accurate on 73% of the decisions for valid sames, which is similar to the accuracy shown in Experiment 2 (74.8%). In spite of the slightly lower than expected performance in the present experiment, Table 3 shows that the model is consistently accurate in predicting the actual decisions reached across problem types and groups. Notice that while reasoners'

**Table 2**  
**Percent Correct Reasoning Decisions**

	Valid Syllogisms		Invalid Syllogisms	
	Sames	Differents	Same-N	Different-N
Questionnaire First	66.2	19.2	23.1	20.5
Questionnaire Second	63.1	25.0	30.8	25.6
Average	64.7	22.1	26.9	23.1

**Table 3**  
**Accuracy of the Conversion Model in Specifying Reasoners' Decisions**

	Same	Different	Different-N
Questionnaire First	66.2	73.1	82.1
Questionnaire Second	63.1	59.6	71.8
Average	64.7	66.4	76.9

*Note—Values are expressed in percentages. Accuracy on invalid same-N cannot be assessed, since the model predicts only that the subject will guess on these problems and be correct 20% of the time (actually observed accuracy = 26.9%; see Table 2).*

accuracy on valid and invalid differentials was at chance levels, the conclusions appear to have been selected systematically. The conversion model correctly predicts the decisions reached on 70% of the problems (ranging from 59.6% to 82.1% accuracy).<sup>2</sup> This clearly differs from what would be predicted from any simple, random selection process (for which accuracy in predictions should approximate 20%).

The importance of premise encoding for categorical decisions was assessed in two phases. In the first phase, each student, on each relation, was characterized as having a convertible or nonconvertible "belief" about the categories. This was accomplished by using the students' estimates of the degree of "overlap" between the categories expressed in the questionnaire. A convertible interpretation was one for which the reasoner's questionnaire response indicated a belief in the converse of the relation that was expressed in the syllogism. For example, corresponding to the syllogistic premise "All blacks are welfare recipients" (see Syllogism 6) were two questionnaire items: "(a) \_\_\_% of blacks are welfare recipients (Question 3)" and "(b) \_\_\_% of welfare recipients are black (Question 33)." If the student's response to Question b was "60%" or more, the student was characterized as holding a convertible belief on that relation (since this is the converse of the syllogistic premise). Such students are termed "converse believers"; all others are "nonconverse believers." Both converse and nonconverse believers can also be characterized in terms of their decisions on Question a. If the response was "60%" or more on Question a, the student is said to hold a belief in the premise as stated ("stated belief"); all others are nonbelievers in the premise as stated ("stated disbelief").

In the second phase, we considered the impact of sentence encoding on the categorical decisions. Our criterion was twofold: (1) whether the two types of encoders (converse and nonconverse) could be discriminated on the basis of their relative likelihood of accepting the "model-predicted" conclusion and (2) the probability of correctly predicting the decision, given the type of encoder. While the latter measure may appear more interesting, the former provides an assessment of how much of the collected judgments are directly related to the reasoner's encoding.

Using the likelihood of a model-predicted decision as a criterion, the model directs our attention to two decision patterns. First, for valid same syllogisms, the probability of the model-predicted decision should be equivalent for both the converse and nonconverse believers, since they are both predicted to reach the logically prescribed decision. Second, the probability of the model-predicted error should be greater for converse believers on valid and invalid different syllogisms, since only the converse believers are supposed to make errors on these problems. Table 4 shows that both claims are supported by the present findings: (1) The model-

**Table 4**  
Probability of a Conversion Response (Various Encodings)

	Converse Believer			Nonconverse Believer		
	B	D	Mean	B	D	Mean
Questionnaire First						
Same	.60	.40	.50	.67	.57	.62
Different	.83	.79	.86		.54	.49
Different-N	.90	.91			.44	
Questionnaire Second						
Same	.64	.63	.64	.53	.64	.59
Different	.71	.65	.77		.33	.49
Different-N	.81	.89			.62	

*Note*—Cells with no entries contained less than 2% of the responses. B = belief and D = disbelief as stated.

predicted decisions occurred with roughly equal probability on same syllogisms for both the converse and nonconverse believers ( $p = .57$  and  $.61$ , respectively; chance = .20), and (2) the probability of the model-predicted error for differentials was substantially greater for converse than for nonconverse believers ( $p = .82$  and  $.49$ , respectively).

The probability that the predicted decision would occur was relatively independent of the student's belief in the relation as stated in the syllogism. There was no difference in the probability of the predicted decision between those cases in which a reasoner believed the stated relation and those in which he/she did not, so long as he/she believed in the converse of those relations (belief in the stated relation:  $p = .81$ ; disbelief in the stated relation:  $p = .82$ ).

When the reasoners' knowledge about the premise categories was assessed (convertible and nonconvertible beliefs), the model correctly predicted 80% of all decisions, at least 53.3% of which would otherwise be ascribed to irrational inference rules rather than to sentence encoding.

## Discussion

These findings illustrate that a model that incorporates conversion as an encoding component can account not only for reasoners' decisions on neutral, abstract, and concrete problems (Revin et al., 1978), but also for decisions with controversial materials. Based solely on the conversion model's strong claims concerning the reasoning process, it is able to account for approximately 70% of all decisions; distinguishing among students' encodings of the premise relations increases the predictive accuracy to 80%.

## EXPERIMENT 2

The first experiment demonstrates that when students reason with controversial materials, their decisions can be predicted from their encoding of the premises rather than premise believability. However, it still remains possible that decisions about controversial

conclusions are affected by the believability of those conclusions rather than by the logical force of the arguments in which they are embedded. The conversion model claims that when premise encoding is controlled, there will be no difference in the acceptance of conclusions that vary in real-world truth value (i.e., believability). To examine this hypothesis, students were asked to solve same and same-N syllogisms whose conclusions varied in truth value, but whose structure was such that premise encoding (i.e., convertibility) would not differentially affect the logical decisions. Consequently, incorrect decisions could be ascribed to irrational rules rather than to premise encoding.

### Method

**Materials.** Twenty-five students were asked to solve 28 concrete, categorical syllogisms. Each problem was presented on a separate page, and the entire set was arranged in a booklet in a random order. Eighteen of these problems had an unambiguous answer (valid same: EI-1 and EI-2), and 10 did not (invalid same-N: II-1 and II-2). The first four problems in the set were treated as practice; they consisted of two valid and two invalid syllogisms, leaving the ratio of valids to invalids in the basic set 2:1.

For half of the valid syllogisms, the logical conclusion was one that the reasoner would select if his decisions were based on real-world truth values. For example, in the following syllogism, Answer d is both the logically valid and the empirically true conclusion: "(7) No members of the ad-hoc committee are women. *Some U.S. senators are members of the ad-hoc committee.* Therefore: (a) All U.S. senators are women. (b) No U.S. senators are women. (c) Some U.S. senators are women. (d) Some U.S. senators are not women. (e) None of the above is proven." Answer c is also empirically correct, but it does not pose a problem for analysis (only 3% of the decisions on these problems are attributable to selecting Answer c). These problems are termed LA syllogisms because logic agrees with belief.

For the remaining valid syllogisms, the logical conclusion would not be selected if the reasoner based his judgments on real-world truth values. For example, while the logical conclusion to the following syllogism is Answer d, the empirically true conclusion corresponds to Answer b; "(8) No U.S. governors are members of the Harem Club. *Some Arabian sheiks are members of the Harem Club.* Therefore: (a) All Arabian sheiks are U.S. governors. (b) No Arabian sheiks are U.S. governors. (c) Some Arabian sheiks are U.S. governors. (d) Some Arabian sheiks are not U.S. governors. (e) None of the above is proven." These problems are termed LC syllogisms, since logic conflicts with belief.

For invalid syllogisms, the logically prescribed conclusion, "None of the above," was always contrasted with the true conclusion of the form "All A are B" or "No A are B." The invalid syllogisms provide a test for whether reasoners tend to prefer conclusions with certain terms (cf. Revlis, 1974) or whether there might be a response bias toward certain conclusions due to their top-down sequencing on the page. These possibilities are important to examine, since it is not possible (with these syllogisms and the constraints of the paradigm) to permit the empirically true conclusion to vary equally across all possible answers, a through d. Consequently, the invalid syllogisms offer an opportunity to examine potential conflicts between Conclusions a and e and between Conclusions b and e. The valid syllogisms contrast Conclusions b and d, as well as Conclusions c and d.

There were two types of believed-true conclusions in this study: (1) empirical relations that can be said to be based on

general knowledge (e.g., "No ex-U.S. vice presidents are women") and (2) definitional relations that are true in part because of their a priori meaning (e.g., "No robots are people"). If students do select the conclusions on the basis of personal knowledge and belief, then such a tendency might reflect the epistemological basis for those beliefs (cf. Rescher, 1964). Clearly, this distinction between empirically true and definitionally true relations is a difficult one to maintain (e.g., Goodman, 1952). However, it has heuristic value for the hypotheses under consideration. Consequently, half of the true statements in the reasoning problems were only empirically true, and half were both definitionally and empirically true.

**Procedure.** Students were told that they would be required to solve reasoning problems and their goal was to decide which of five possible conclusions had to follow unambiguously from the given premises. The subjects read the rules for solving such problems and were shown a sample problem that was not repeated in the experimental set. The subjects were instructed to work each problem in the 60 sec allotted and to proceed to the next problem in their booklet only when told to do so.

**Subjects.** The subjects were 25 introductory psychology students fulfilling a course requirement. They were run in a single session lasting 45 min. None of the subjects had been exposed to a course in logic.

### Results

The accuracy score of each reasoner for the basic 24 problems was determined by summing the percentage correct for each problem type (LA and LC) for valid syllogisms and LC problems for invalid syllogisms (there were no invalid LA problems). The results are presented in Table 5, which shows that when solving valid, categorical syllogisms, students were less accurate in their decisions when logic and knowledge conflicted than when they agreed [ $F(1,24) = 13.5, p < .01$ ]. While there was no difference overall in accuracy between problems with definitionally true or empirically true conclusions, basis for belief did interact with problem type [ $F(1,24) = 9.1, p < .01$ ]. That is, when logic and knowledge conflicted, reasoners' accuracy deteriorated primarily in those cases in which the conflict was between a logical conclusion and a definitionally true conclusion. While this trend toward belief bias reached conventional levels of significance, it should be kept in mind that the effect of believability of the conclusions was quite limited: Inspection of Table 5 shows that, on valid problems, reasoners overwhelmingly selected the logically correct conclusion (observed = 74.8% accuracy; chance = 20%). This was also true for invalid syllogisms, for

**Table 5**  
Reasoning Accuracy When Conclusions Vary in Believability

Problem Type	Type of Belief		Mean
	Definitionally True	Empirically True	
LA	88.0	78.0	83.0
LC Valid	60.0	73.0	66.6
LC Invalid	63.0	68.0	65.5

*Note*—Values are expressed as percentages. LA = logic agrees with belief; LC = logic conflicts with belief.

which the logically required answer ("None of the above") was accepted 65.5% of the time.<sup>3</sup>

While reasoning errors were highest when belief conflicted with logic, the specific errors that were made were not completely accounted for by students' preference for the believed-true conclusion. On valid syllogisms, when logic and belief conflicted, less than half of the reasoning errors (valids: 47.8%; invalids: 4.3%) occurred when reasoners selected the conclusion that agreed with their personal knowledge (Answer b). Yet, when logic and belief agreed, students also accepted Conclusion b (35.3% of the time). In this case, the conclusion was actually disbelieved.

In addition to overall reasoning accuracy, there were some discernible response patterns. First, on valid syllogisms, there was a tendency for some of the students to select Alternative b independently of its truth value. Of the belief errors in the LC condition, 72% were contributed by only five subjects who made the same error in the LA condition: They accepted Conclusion b. Again, this error in the LA condition is tantamount to accepting a conclusion that these subjects must clearly disbelieve (e.g., "No college students are seniors"). It is plausible, therefore, that much of the effect of personal bias was the result of a few students following an unpredicted reasoning process, but one that is clearly orthogonal to accepting conclusions based on their real-world truth values. This preference for Conclusion b appears to reflect confusion with the meaning of the quantifier "some." Students were informed that "some" should be given a distributive definition, so that if "Some A are not B" is true, it is possible that "No A are B" is also true. This special definition for the quantifier is occasionally confusing to students (Frase, 1966), and it may have been especially troublesome in the present study because the logically required response for the valid syllogisms was of the form "Some A are not B." This hypothesis seems plausible, since in no case did any reasoner accept Conclusion b when solving invalid syllogisms, problems for which logically they would not have considered a conclusion with "some" as the quantifier.

A second aspect of these data merits interest. While the belief-bias effect appeared to be strongest when definitionally true statements and logic conflicted (LC), the dominant error for these syllogisms (62.5% of all errors) was due to the assertion "No conclusion is proven," rather than to the acceptance of the believed conclusion. This is in contrast to the error profile on empirically true statements, for which the indeterminate response did not differ from chance expectation (18.5% of all errors). Overall, the frequency of such responses was greater for definitionally true statements than for empirically true statements [ $F(1,24) = 5.2, p = .01$ ] and was shown primarily in the LC condition [Belief Type by Problem Type:  $F(1,24) = 18.3, p < .001$ ]. This suggests that when there is an inherent conflict between belief and logic, it tends to be resolved by the reasoner's

opting for the logically correct conclusion or, if an error is made, by selecting an indeterminate response ("No conclusion") rather than selecting the believed but illogical conclusion.

For invalid syllogisms, this conflict between belief and logic is also resolved by accepting the "none" conclusion. Fortuitously, for these problems, the conclusion is the one that is logically prescribed. As a result, an increased accuracy on these problems was observed over Experiment 1.

A more detailed, subject-by-subject analysis of the data supported the foregoing and argued persuasively that personal bias per se had only an indirect influence on reasoning decisions. The most appropriate subjects to look at in evaluating the belief-bias effect are those who made no errors in the LA condition. If students accept conclusions based on their believability, then one would predict optimal accuracy in the LA condition and minimal accuracy in the LC condition. Therefore, we compared the performance of the 11 students who made no errors when belief and logic agreed (LA) with their performance when belief and logic conflicted (LC). We found that while their accuracy in the LC condition was lower than in the LA condition, it was not appreciably lower (88.7% compared with 100%, respectively). In addition, there was no clear difference between definitionally true conclusions and empirically true ones (86.4% compared with 90.9%). It is important to note that the 11 subjects did not tend to select believed conclusions in the LC condition. Such errors represented only 2% of all their responses, an insignificant component of the total data.

### Discussion

There are at least two aspects of these findings that are critical for a rational model explanation of the belief-bias effect. First, reasoners tend to make rational judgments even when such decisions conflict with their personal knowledge: There is only a modest tendency to reject the logically correct conclusion (34%). Second, when students fail to reach the logical conclusion, it is not due to a total suspension of rationality in favor of personal biases: Only a small proportion of errors can be attributed to belief in the conclusions. When logic and belief conflict, the dominant error (52.2% of all errors) is to opt for no conclusion rather than to make a choice between logic and belief. This indicates that students are sensitive both to the extraexperimental belief values and to the logical properties of statements. Their decisions reflect an interrupt to the flow of rational inference rather than a bias toward believed conclusions.

### GENERAL DISCUSSION

In the experiments presented here, we have undertaken to show that at least one rational model, the conversion model, can account for the apparent reason-

ing bias that historically has been said to act "irrationally," leading to the acceptance of conclusions based on their truth values. The model accomplishes this without having to posit post hoc mechanisms and without having to violate its underlying assumption that the decisions people reach are based on logical operations applied to their understanding of the materials reasoned about (although these operations have not been specified). The conversion model claims that LTM and normal language comprehension mechanisms participate in the encoding of the materials, creating a personalized representation from whence rational decisions are made. Some of these decisions are the same as those that a logician would reach (sames) and some are not (differents). In the present study, failures to accept the predicted conclusion are not the result of a belief bias, but rather, they reflect interrupts in the decision process resulting from an implicit awareness of the conflict between logic and belief. As such, students respond that no conclusion is possible rather than accept the believed statement over the logically prescribed one. Reasoners' decisions appear to reflect rational processes: The conclusions reached on categorical syllogisms do not in any simple way reflect the external truth values of the premises or of the conclusions.

The approach taken here is consistent with that of Henle (Henle, 1962; Henle & Michael, 1956), who succinctly illustrated how reasoners might interpret the propositions of a syllogism in ways that the experimenter or logician did not intend. An advantage of the present formulation is that it extends Henle's criticisms of the existent literature by pointing out a potential confound in the research on beliefs in formal reasoning; namely, such research may have manipulated not only the belief status of the conclusions, but also the kind of relations expressed in the premises. Syllogism 1 is typical, in that one cannot tell whether the reasoners' decisions are a result of faulty inference or the consequence of a personalized representation, as claimed by the conversion model.

Scribner's (1977) cross-cultural observations of an "empirical bias" also argues that reasoners do not treat premises as abstract relations, but rather, as having real-world content. In such cases, the stated relations are "portmanteau" expressions, carrying in them packets of derived relations that function as new conditions that may dictate quite different inferences than would be prescribed by logic. Equally as important, such relations may contribute to apparently logical answers, but for reasons that would be considered inappropriate by a logician. The present analysis suggests, therefore, the viability of a rational framework for theories of reasoning and the importance of further research on sentence comprehension and encoding for a fuller understanding of logical processes and human inference.

## REFERENCES

- BUCCI, W. The interpretation of universal affirmative propositions. *Cognition*, 1978, 6, 55-77.
- CHAPMAN, L., & CHAPMAN, J. Atmosphere effect re-examined. *Journal of Experimental Psychology*, 1959, 58, 220-226.
- CHASE, W., & CLARK, H. H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.
- COHEN, M. R., & NAGEL, E. *An introduction to logic*. New York: Harcourt, 1934.
- COLLINS, A., & QUILLIAN, M. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 240-248.
- ERICKSON, J. Research on syllogistic reasoning. In R. Revlin & R. Mayer (Eds.), *Human reasoning*. Washington, D.C: Winston, 1978.
- FEATHER, N. Acceptance and rejection of arguments in relation to attitude strength, critical ability, and intolerance of inconsistency. *Journal of Abnormal and Social Psychology*, 1965, 69, 127-136.
- FRASE, L. T. Validity judgments of syllogisms in relation to two sets of terms. *Journal of Educational Psychology*, 1966, 57, 239-245.
- GOODMAN, N. The problem of counterfactual conditionals. In L. Linsky (Ed.), *Semantics and the philosophy of language*. Urbana: University of Illinois Press, 1952.
- GORDON, R. Attitudes toward Russia on logical reasoning. *Journal of Social Psychology*, 1953, 37, 103-111.
- HENLE, M. On the relation between logic and thinking. *Psychological Review*, 1962, 69, 366-378.
- HENLE, M., & MICHAEL, M. The influence of attitudes on syllogistic reasoning. *Journal of Social Psychology*, 1956, 44, 115-127.
- JANIS, I., & FRICK, P. The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. *Journal of Experimental Psychology*, 1943, 33, 73-77.
- JANIS, I., & TERWILLIGER, R. An experimental study of psychological resistances to fear arousing communications. *Journal of Abnormal and Social Psychology*, 1962, 65, 403-410.
- JOHNSON-LAIRD, P., & STEEDMAN, M. The psychology of syllogisms. *Cognitive Psychology*, 1978, 10, 64-99.
- KAUFMAN, H., & GOLDSTEIN, S. The effects of emotional value of conclusions upon distortions in syllogistic reasoning. *Psychonomic Science*, 1967, 7, 367-368.
- LEFFORD, A. The influence of emotional subject matter on logical reasoning. *Journal of General Psychology*, 1946, 34, 127-151.
- MCGUIRE, W. A syllogistic analysis of cognitive relationships. In M. Rosenberg & C. Hovland (Eds.), *Attitude organization and change*. New Haven: Yale University Press, 1960.
- MORGAN, J., & MORTON, J. The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, 1944, 20, 39-59.
- RESCHER, N. *Hypothetical reasoning*. Amsterdam: North Holland, 1964.
- REVLIN, R., AMMERMAN, K., PETERSEN, K., & LEIRER, V. Category relations and syllogistic reasoning. *Journal of Educational Psychology*, 1978, 70, 613-625.
- REVLIN, R., & LEIRER, V. O. The effects of personal biases on syllogistic reasoning: Rational decisions from personalized representations. In R. Revlin & R. Mayer (Eds.), *Human reasoning*. Washington, D.C: Winston, 1978.
- REVLIN, R., & LEIRER, V. O. Understanding quantified categorical expressions. *Memory & Cognition*, 1980, 8, 447-458.
- REVLIN, R. Prevarication: Reasoning from false assumptions. *Memory & Cognition*, 1974, 2, 87-95.
- REVLIN, R. Syllogistic reasoning: Logical decisions from a complex data base. In R. Falmagne (Ed.), *Reasoning: Representation*



- and process. Hillsdale, N.J.: Erlbaum, 1975. (a)
- REVLIS, R. Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 180-195. (b)
- SCRIBNER, S. Modes of thinking and ways of speaking: Culture and logic reconsidered. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press, 1977.
- SIMON, H. A. Motivational and emotional controls of cognition. *Psychological Review*, 1967, **74**, 29-39.
- TSAL, Y. Symmetry and transitivity assumptions about a non-specified logical relation. *Quarterly Journal of Experimental Psychology*, 1977, **29**, 677-684.
- THISTLETHWAITE, D. Attitude and structure as factors in the distortion of reasoning. *Journal of Abnormal and Social Psychology*, 1950, **45**, 422-448.
- WILKINS, M. The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, 1928, **16**, 83.
- WILSON, W. The effect of competition on the speed and accuracy of syllogistic reasoning. *Journal of Social Psychology*, 1965, **65**, 27-32.
- WINTHROP, H. Semantic factors in the measurement of personality integration. *Journal of Social Psychology*, 1946, **24**, 149-175.

## NOTES

1. The letters designate the quantified relation in each premise: A = All A are B; E = No A are B; I = Some A are B; O = Some A are not B. The numbers indicate the configuration of subject and predicate terms in the premises; four arrangements are possible (after Cohen & Nagel, 1934):

$$1 = \begin{matrix} MP \\ SM \end{matrix}, 2 = \begin{matrix} PM \\ SM \end{matrix}, 3 = \begin{matrix} MP \\ MS \end{matrix}, 4 = \begin{matrix} PM \\ MS \end{matrix}$$

2. This does not include decisions on invalid same-N syllogisms, since only chance responding is predicted for these problems.

3. The absence of a strong belief-bias effect for valid syllogisms cannot be attributed solely to a response bias against "No A are B" type answers (the believed-true propositions), since such conclusions are accepted 75% of the time when predicted in Experiment 1. In addition, when the logically required conclusion is contrasted with believed conclusions other than "No A are B," a similar error pattern is shown on invalid syllogisms: "All A are B," 1.5% error; "Some A are B," 3.0% error.

(Received for publication September 18, 1979;  
revision accepted May 5, 1980.)