Memory & Cognition
2001, 29 (8), 1196-1208

Reasoning counterfactually: Combining and rending

RUSSELL REVLIN, CHRISTINA L. CATE, and TENA S. ROUSS
University of California, Santa Barbara, California

Counterfactual reasoning occurs when people are asked to assume for the sake of argument that a fact they previously thought was true is now false and to draw a conclusion on that basis. To accomplish this sort of reasoning requires a revising of one's beliefs, which was simulated in the present study. Students were shown a set of statements that they were to assure themselves was consistent. They were then asked to accept a counterfactual assumption as true and reconcile resulting inconsistencies among the set of statements. In these problems, one statement is a generality (e.g., *All trees on the plaza are elms*), another is a particular (e.g., *This tree is a pine*), and one is a counterfactual (e.g., *Assume this tree is on the plaza*). Students preferred to reconcile the inconsistency by identifying the generality as "true" and the particular as "false." They did this more often when the assumption *combined* categories than when it *dislodged* categories and when real beliefs were at stake rather than arbitrary generalities. This study tested current models of inference for their ability to account for counterfactual reasoning and found the results to be consistent with natural deduction system, mental models, and conceptual-integration network approaches to everyday reasoning.

Counterfactual reasoning is defined as "reasoning from false assumptions." It is commonplace in everyday thinking—from planning about the future (*If I were to die tomorrow, is my family protected*), to pretending (*Long ago, in a galaxy . . .*), to conjecturing about the importance of some hypothetical situation (*If you had studied, you would have gotten an A*) (see Farris & Revlin, 1989; Fauconnier, 1997; Simon & Rescher, 1966). However, its ubiquity does not make it any less of a challenge to understand the cognitive processes that participate in such inference. For example, let's suppose that you're asked to evaluate the consequences of the conjecture *If this animal (which is a fish) were a whale*, then either *this animal would be a mammal* or *not all whales are mammals*. One way to represent this sort of reasoning is to propositionalize the antecedent of the conditional as in statement 1d and examine its consequences for the set of statements you believe to be true and that are operative for the case at hand (statements 1a–1c). Although you believe that statement 1d is false (asterisk indicates falsehood), you must assume it is true and reconcile it with existing beliefs. This paradigm was developed by Rescher (1964), who called such situations *belief-contravening problems*.

(1a) All whales are mammals
(1b) This animal is not a whale
(1c) This animal is not a mammal
(1d) *this animal is a whale

The task at hand is to determine how to revise your beliefs (1a–1c) in light of the Counterfactual Assumption 1d. The question of what beliefs to revise seems deceptively simple. The assumption contradicts statement 1b, so the latter must be false. But notice that if the reasoner identifies 1a as true, then it and the assumption jointly contradict 1c; we will refer to this as the *generalist* solution (after Revlis, 1974) since its effect is to identify the generality as true. Alternatively, if the reasoner combines the assumption with statement 1c first, the two together contradict 1a; we will refer to this as the *particularist* solution since its effect is to identify the particular as true. In terms of standard logic, no solution path is more rational than another (e.g., Chisholm, 1946), and yet students express a reliable tendency to select the generalist path in situations described above (Redding-Stewart & Revlin, 1978; Revlis, 1974; Revlis & Hayes, 1972; Revlis, Lipkin, & Hayes, 1971). These choices made by human reasoners are potentially in conflict with artificial intelligence (AI) prescriptions of *belief revisioning*, which are automatic algorithms for treating new "facts" that are in conflict with established beliefs (see Elio & Pelletier, 1997; Thagard, 1989). The present study examines some of the underlying cognitive processes in counterfactual reasoning and how they may inform our understanding of theoretical accounts of belief revisioning.

Counterfactual Methods

We consider the predictive accuracy of three different approaches for the resolution of belief-contravening problems. These methods are, in historical order, predicate cal-

We wish to express our appreciation to M. Oaksford, S. Sloman, and an anonymous reviewer for their thoughtful comments on an earlier draft of this paper. Correspondence should be addressed to R. Revlin, Psychology Department, University of California, Santa Barbara, CA 93106 (e-mail: revlin@psych.ucsb.edu).

culus, modal logic, and natural deduction systems (NDS). Other approaches will be evaluated in the General Discussion.

Predicate calculus. The paradigmatic belief-contravening problem can be stated symbolically as in statement 2, where $p \rightarrow q$ represents the generality *If x is a whale, then x is a mammal* (i.e., *All whales are mammals*); $\sim p$ indicates that *x is not a whale* and $\sim q$ indicates that *x is not a mammal*.

(2) GIVEN: $[p \rightarrow q, \sim p, \sim q]$
ASSUME: $[p]$

The assumption $[p]$ replaces one of the givens $[\sim p]$ and leaves the reasoner to decide whether under the generality $[\sim q]$ should be replaced with $[q]$ or whether in the face of the existence of $[\sim q]$, the generality itself should be denied. Clearly, standard predicate calculus can demonstrate the propositional conflict, but as we have stated, the resolution of the counterfactual eludes standard predicate calculus, which prescribes no valid inference when the premises are false. This is because when you reason from false assumptions, any conclusion is possibly valid (Chisholm, 1946; Rescher, 1964). Predicate calculus therefore does not provide a guide as to which statements should be retained and which should be discarded. This is unfortunate because counterfactual reasoning potentially allows us to gain new understandings and new knowledge (Peirce, 1877). It offers an opportunity to adjust our system of beliefs, but there are no guidelines—at least not from standard logic—that tell us how to accomplish this. In a sense, the central question for counterfactual reasoning is, how should we revise our beliefs?

Modal logic. In describing the rational conflict inherent in counterfactual reasoning, Rescher (1964) created the paradigm of belief-contravening problems illustrated in statement 1. He proposed that the solution to belief revision should be based on a *modal logic* analysis in which the facts directly related to the problem at hand should be organized along the lines of *degrees of necessity* and that the reasoner should seek to retain those statements with the lowest modal value (the most necessary). Rescher claimed that this strategy would be the most natural solution to counterfactual conditionals. Students' actual selection of the generalist path corresponds to the predictions of this modal logic analysis of the problems (Revlis et al., 1971). Consequently, the inherent modal logic components have been embodied in a process model called the generality coding model (GCM; Revlis, 1974), which asserts that reasoning with counterfactual conditionals proceeds in three stages. First, the reasoner constructs a possible world in which the counterfactual assumption can be true. The ability to do this is not restricted to belief-contravening problems, but may be a broadly applicable reasoning process. For example, Sternberg and Gastel (1989), employing categorization, series-completion, and analogical reasoning tasks, showed that the ability to integrate counterfactual propositions is related to fluid intelligence.

In the second stage, the reasoner orders the available descriptions about the world in terms of his/her ability to structure the domain—that is, in terms of modality (Rescher, 1961, 1964). The more general the proposition, the more instances it predicts in a well-defined universe of discourse, and therefore the more central the statement will be for reasoning about new, possible worlds. In the third stage, reasoners seek to retain those propositions that provide an ordering rule within the domain of the problem and sacrifice less central, more particularist propositions in order to retain the more general ones.

The GCM leads us to anticipate that reasoners will exhibit a form of *entrenchment* where one or more propositions are protected from elimination. In statement 1 this would be the lawlike generality. Other forms of counterfactual reasoning support the notion that some statements are less mutable than others. For example, when asked how a situation might be changed to have made the outcome better, people are less willing to sacrifice the "normal" than the odd event (Byrne, Segura, Culhane, Tasso, & Berrocal, 2000). In a sense, the generality might be construed as the normative and therefore the more reliable statement of a state of affairs—*ground* for the hypothetical, possible world—so that reasoners are less willing to make changes to it and are more likely to alter the isolated particular. That reasoners might be motivated to retain the most general statement is seen in an independent task by Johnson-Laird and Steedman (1978), who had students generate conclusions to syllogistic premises and noticed that they tended to avoid restatement or summaries and preferred to capture a more general relationship (even at the risk of being incorrect).

Other independent evidence for entrenchment has been found in *subtyping*, which occurs when people are presented with counterarguments to their stereotypes (i.e., generalities) and they dismiss these disconfirming examples as "the exception that proves the rule." Subtyping has even been shown to actually strengthen stereotypic belief (Hewstone, Hopkins, & Routh, 1992; Hewstone, Macrae, Griffiths, & Milne, 1994). The opposite process, called *conversion* (Rothbart, 1981), occurs when a particular instance disconfirms a general rule (equivalent to the *particularist* strategy). Findings with the sort of belief-contravening problems illustrated above show evidence for the entrenchment of universally quantified relation—subtyping rather than conversion—and this entrenchment is predicted by a modal logic analysis of the task domain.

Notice that this finding contrasts with traditional views of hypothesis testing in science, in which a single disconfirming datum should overthrow the generalization, and formal models of scientific reasoning prescribe just that (e.g., Thagard, 1989). This may be especially sensible in cases where the generalizations are based on potentially fallible observations. However, one reason why theories in science are not so easily displaced is because they are more than accidental generalizations (Kuhn, 1996). As Goodman (1952) has illustrated, a statement of the form *All coins in my pocket are silver* is an accidental generality in the case where there is no principled reason for the event in question. In this case, no matter how many times the coins are observed, a single disconfirming ob-

servation in which there is a copper coin, for example, would overturn the generalization. This is in contrast to lawlike statements (e.g., *All whales are mammals*) that are based on principles as well as other observations and hold across space and time conditions. Such generalities are more than a summary of repeated measurements because they provide predictions of new observations. In a sense, they are entrenched because they function as an inference ticket in science (Ryle, 1949). It should come as no surprise, then, that research that seeks to simulate scientific reasoning (e.g., Elio & Pelletier, 1997) may be misled by presenting students with statements that are syntactically equivalent to lawlike conditionals, but are contextually framed as accidental generalizations.

Natural deduction. Another resolution procedure is associated with NDS (e.g., Braine & O'Brien, 1998; Rips, 1994). In these proof systems, there is a set of inference rules (a subset of the standard ones) that act as operators that move the system from one state to another. The system contains the heuristic-like application of rules that allows it to bridge the gap between the supposition (the counterfactual assumption) and one of the remaining statements in the initial ensemble. Of all the available rules, modus ponens would be at the top of the stack of available options (IF elimination in the PSYCOP model of Rips, 1994). In statement 1, the generality might be glossed as follows:

(3a) *If x is a whale then x is a mammal*
(3b) *x is not a whale*
(3c) *x is not a mammal*
(3d) **this x is a whale*

Statement 3d requires immediate denial of 3b. Then, the application of modus ponens would require the combination of 3a and 3d and the conclusion that *x must be a mammal,* and therefore the rejection of 3c. The application of a modus ponens rule creates a simple resolution to the problem corresponding to the generalist path. The particularist path is unmotivated in this account of NDS, evoking a circuitous series of operations, which makes it more cognitively laborious. On counterfactuals, then, the NDS and modal logic approaches (*qua* GCM) predict entrenchment of the generality because it is consistent with the application of the modus ponens rule.[1]

Types of Counterfactuals

The type of counterfactual portrayed in the belief-contravening problem may be called a *combining problem* because a new relation is *added* to the set of beliefs, which the reasoner must reconcile (e.g., that the *animal* is now a *whale*). However, counterfactuals could take a different form: They could deny a salient relation that already exists. Such situations would be exemplified by *rending problems*, illustrated as follows:

(4a) *All whales are mammals* [$p \rightarrow q$]
(4b) *This animal is a whale* [p]
(4c) *This animal is a mammal* [q]
(4d) **this animal is not a mammal* [$\sim q$]

The GCM supports the generalist path as long as it expresses a necessary condition (i.e., is lawlike: its truth is not delimited in space and time and it is believed on the basis of generally available information—Revlis & Hayes, 1972). Consequently, the modal logic approach would predict that reasoners would resolve this belief-contravening problem by taking the generalist path: retain 4a and reject 4b.

Recall that the natural deduction approach, as embodied in the mental logic of Braine and O'Brien (1998) or PSYCOP of Rips (1994), makes no predictions concerning entrenchment of generalities and applies only syntactically relevant rules. Therefore, in seeking the application of a mediating rule for rending problems, it would be sensitive to the introduction of an assumption that denies a relation (e.g., 4d), which in turn creates an argument structure equivalent to modus tollens [$p \rightarrow q$, $\sim q$, therefore $\sim p$; but p]. However, neither of these NDSs contains a modus tollens rule, and therefore they do not possess a direct method for resolving the logical conflict and should not predict a difference between preferences for alternate solution paths.[2]

Model-Based Predictions

We have described three psychologically relevant methods that can be applied to direct the resolution of counterfactuals; these are summarized in Table 1. The predicate calculus approach treats combining and rending problems equivalently. It does not specify the entrenchment of some statements over others, and so it is sensible to conclude on the basis of standard predicate calculus that no preference will be shown by reasoners on either of these problems. On both problems, students' preferences will be at chance levels. In contrast, the modal logic approach as embodied in the GCM predicts entrenchment of the generality in the rending problems as well as the combining problems. Finally, the natural deduction approach supports the application of the modus ponens rule for the combining problems—leading to the entrenchment of generalities. However, absent an operative forward rule, the natural deduction approach would anticipate no specific preference in rending problems. We anticipate that if reasoners were following the precepts of natural deduction, then any preference for the generalist path on rending problems would approximate normative accuracy on modus tollens problems—50% (e.g., Girotto, Mazzocco, & Tasso, 1997). One purpose of the present study is to assess the applicability of the three approaches to logical inference for their

Table 1
Preference for Generalities in Experiment 1: Predictions of Three Approaches

	Counterfactual Problems	
Approach	Combining (%)	Rending (%)
Predicate Calculus	50	50
Modal Logic (GCM)	100	100
Natural Deduction	100	50

Note—GCM, generality coding model.

ability to account for the pattern of decisions shown by students when reasoning about two types of counterfactuals.

The GCM has been effective in accounting for the entrenchment of generalities expressed as categorical relations (e.g., *All x are included in y*), but less effective with property-assignment relations (e.g., *All x have property y*) (Revlis, 1974). Neither predicate calculus nor NDS have been examined for their ability to explain counterfactual reasoning with property-assignment relations. Consequently, another purpose of this study will be to replicate and extend previous findings with property-assignment counterfactuals.

EXPERIMENT 1

Method

Design and Procedure. Two groups of subjects were asked to solve 24 counterfactual conditionals as illustrated in statement 1 and to indicate which sentences they wish to reject in order to establish a consistent set of statements. Instructions were read aloud by the experimenter while the students followed along with their own copy of the instructions. A sample problem that contained symbolic letters to stand for the categories was used to illustrate the task. The two possible solutions were demonstrated. Students were told to draw a line through the statement on each problem that they wished to discard in order to create a set of consistent statements. The experimenter asked them to proceed through the problems at their own rate and to work on each problem in the order presented in the booklet (one problem per page). When students questioned the logical ambiguity of the solutions, they were instructed to choose the "best" alternative for them. Finally, students were told that the rationale for this experiment was to see how students would prefer to resolve the equally logical contradictions.

Subjects. Sixty-one students participated in this study to fulfill a course requirement. They were run in groups of up to 5 in sessions lasting approximately 30 min. The data from 5 students were excluded from analysis for failure to follow instructions.

Materials. One group of 28 students solved problems in which the critical propositions expressed class-inclusion relations (e.g., *All snakes in the forest are reptiles*), and the other group of 33 students solved problems where the critical propositions expressed property-assignment relations (e.g., *All snakes in the forest have cold blood*). The problems in each booklet represented the orthogonal contrast of two variables: order of appearance on the page and polarity. *Order* refers to whether the generality was the first mentioned proposition or the second; it was included as a variable to control for the possibility that students would be biased to assign truth values as a function of which statements were first mentioned. *Polarity* refers to whether the generality was expressed as an affirmative relation (e.g., *All snakes in the forest are reptiles*) or a negative one (*No snakes in the forest are mammals*). The polarity of the particular fact in each problem was opposite to that of the generality. Prior work showed that the polarity of the facts was irrelevant to the decisions (Revlis, 1974); however, the variable is included in the present study as a replication.

On every problem, the counterfactual assumption was positioned as the last statement in the problem to guarantee that the subjects would have acquainted themselves with all of the statements prior to considering the implication of the counterfactual. The sequence of problems within a booklet followed a single random order that was the same for each booklet.

Results

The preference for generalities is summarized in Table 2; it shows that students identified generalities as true 90.7% of the time, which is significantly more often than would be expected by chance alone ($p < .001$, binomial test). There was no reliable difference in preference for the generalities expressed either as class inclusion or as property assignment (91.3% and 90.0%, respectively). This preference for generalities was not affected by whether the general statement was ordered as the first or second mentioned in the problem. Nor did order interact with any of the other variables in this study.

Polarity exercised a reliable effect on preferences: Affirmative generalities were accepted more often than negative ones [94.1% and 87.2%, respectively: $F(1,55) = 15.9, p < .001$]. Polarity did not interact with relation.

Discussion

The present findings replicate earlier studies and indicate that when students reason from false assumptions, they accord generalities a priority over particular facts. Although the polarity of the statements influences the assignment of truth values, polarity per se cannot be the sole arbiter in making the decisions, because individually, both affirmative and negative generalities are accepted significantly more often than would be predicted by chance alone. Although the preferences exhibited by students when reasoning counterfactually suggest the importance of a *modal ordering* for assigning truth values, the modal ordering is not sufficient to explain all of the decisions. For example, affirmative and negative generalities should hold the same modal position, yet they are treated somewhat differently in this and earlier studies. In this way, the present findings are not completely consistent with the predictions of the modal logic account as embodied in the GCM. However, the model cannot be dismissed since negation may be a general cognitive effect, quite independent of counterfactual reasoning (see below).

These findings are consistent with natural deduction models in two ways. First, the preferred path is the one predicted on the basis of modus ponens rules. The natural deduction approach predicts that the path through the belief-

Table 2
Percent Acceptance of Generalities—Combining Counterfactual

	Affirmative Generalities				Negative Generalities			
	First Mention		Second Mention		First Mention		Second Mention	
Relation	%	*SD*	%	*SD*	%	*SD*	%	*SD*
Class inclusion	92.7	12.5	95.8	0.09	89.3	15.6	87.5	14.8
Property assignment	93.9	13.9	94.1	13.00	89.2	13.8	82.9	19.3

contravening problems will be consistent with *mental logic* rules, and, in the present condition, that would include modus ponens. Consequently, students are expected to identify the generality as true and the particular as false.

Second, the polarity of the generality, a syntactic element, is acknowledged to trigger extra processing in natural deduction accounts (e.g., Rips, 1994). For example, Evans, Clibbens, and Rood (1996) showed that when rule-like statements in a positional reasoning task contain a negative (i.e., *No P are Q*), reasoning accuracy is somewhat lower than when the statement does not contain a negative. This may be a consequence of difficulty in comprehending negation—especially due to change of focus (e.g., MacDonald & Just, 1989).

When reasoning counterfactually, the students in the present study treated generalities that expressed property-assignment relations similar to ones that expressed class-inclusion relations. This finding departs from previous studies of counterfactual reasoning in which the preference for property generalities was reliably less than for class inclusion, although the preference was still greater than chance (e.g., Revlis, 1974). It is also different from studies of inference, which show that when reasoning with property relations, students may need to be reminded of the superordinate class relations that are presupposed (Sloman, 1998). Some clarity on this variable may be gained from Experiment 2.

As we noted earlier, conjecturing about combining new elements is only one way to challenge current knowledge with counterfactual assumptions. By definition, positing new relations entails discarding old ones. That is certainly the case with the belief-contravening problems in Experiment 1, where reconciling the effects of the counterfactual assumption requires a denial of previously believed relations. Therefore, for a fuller treatment of counterfactual reasoning, one should also consider the direct impact of *uncoupling* the previously connected categories and properties—that is, *rending* rather than *combining*. To this end, the following experiment examines the preferences for generalities in the context of uncoupling instances from categories.

EXPERIMENT 2A

We refer to the problems in Experiment 1 as *combining problems* because a new relation is *added* to the set of beliefs by virtue of the counterfactual assumption (e.g., *This animal is now a whale*). In this experiment we examine *rending problems*, in which the counterfactual assumption *uncouples* a relation that already exists:

(5) All members of the Chicago Bulls are members of the NBA
This player is a member of the Chicago Bulls
This player is a member of the NBA
* This player is not a member of the NBA

If students evaluate counterfactual conditionals that uncouple an existing relationship just as they do for counterfactuals that create a connection, we would anticipate a pattern of acceptances similar to that shown for Experiment 1, which would argue for the importance of modal categories in these decisions. However, if the pattern of preferences is not consistent with that shown for the combining problems, then strong support for an NDS approach and other alternatives would be garnered (see summary in Table 1).

Method

Design and Procedure. Two groups of subjects were asked to solve 24 counterfactual conditionals that we referred to as *rending problems* (illustrated above). The details of the design and procedures were equivalent to those of Experiment 1.

Subjects. Fifty-six students participated in this study to fulfill a course requirement. They were run in groups of up to 5 in sessions lasting approximately 30 min.

Materials. One group of 28 students solved problems where the critical propositions expressed class-inclusion relations (e.g., *All members of the Chicago Bulls are members of the NBA*), and the other group of 28 students solved problems where the critical propositions expressed property-assignment relations (e.g., *All members of the Chicago Bulls have athletic skills*). As in Experiment 1, the problems in each booklet represent the orthogonal contrast of two variables: order of appearance on the page and polarity.

On every problem, the counterfactual assumption was positioned as the last statement in the problem. The sequence of problems within a booklet followed a single random order that was the same for each booklet.

Results

The preference for generalities with both class-inclusion and property-assignment problems is summarized in Table 3; it shows that students identified generalities as true 48.5% of the time, which is *not* more often than would be expected by chance alone. Inspection of Table 3 reveals that the relation expressed in the statements was critical to the decisions: The students' pattern in assigning truth values to general and particular statements is different for class-inclusion and property-assignment statements [$F(1,55) = 12.1, p < .001$], and relation interacts with po-

Table 3
Percent Acceptance of Generalities—Rending Counterfactual

	Affirmative Generalities				Negative Generalities			
	First Mention		Second Mention		First Mention		Second Mention	
Relation	%	*SD*	%	*SD*	%	*SD*	%	*SD*
Class inclusion	65.2	36.8	49.2	33.8	69.6	34.9	64.6	30.0
Property assignment	39.5	44.1	28.9	36.5	34.9	32.5	26.9	33.7

larity [$F(1,55) = 7.1$, $p = .01$]. Consequently, we will examine the pattern for the two relations separately.

Class-inclusion statements. Table 3 shows that the preference for generalities, when they express class-inclusion in rending problems, is 62.5%, which is not reliably different from chance alone. In contrast with the pattern shown for combining problems, students showed a greater preference for reasoning with negative generalities (67.1%) than with affirmative ones [59.2%, $F(1,29) = 6.3$, $p < .01$], neither of which is reliably greater than chance. Order plays an important role in acceptance of these class-inclusion generalities: When the generality is the first mentioned, it is accepted more often than when it is mentioned second [$F(1,29) = 13.5$, $p < .001$]. However, this effect of order was more substantial for affirmative than for negative generalities—contributing to an order × polarity interaction [$F(1,29) = 6.8$, $p < .01$]. Recall that none of these effects reflect a preference for either generality or fact that exceeds chance expectations.

Property-assignment statements. When reasoning about rending problems, whose generalities express property-assignment relations, students showed a distinct preference for *rejecting* the generality and accepting the particular fact. Table 3 shows that the preference for generalities was 32.6% (or a rejection rate of 67.4%), which is reliably different from chance ($p = .05$, binomial). This preference for the particular facts rather than the generalities is independent of polarity or order of statements within a problem.

Discussion

In this experiment, students were asked to resolve inconsistencies resulting from a counterfactual assumption that *denied* an existing category structure. Overall, students did not show a preference for preserving (i.e., *entrenchment*) the most general statement. Indeed, when the generality expressed a property-assignment relationship (e.g., *All A have C*), students reliably preferred to reject the generality in favor of the particular.

Comparison of the preferences shown in Table 2 for combining problems and those shown in Table 3 for rending problems indicates that the logical context affects the decisions reached on belief-contravening problems. The preference for generalities was greater on combining problems than on rending ones, both when the generalities expressed class-inclusion relations [$F(1,48) = 23.5$, $p < .001$] and when the generalities expressed property-assignment relations [$F(1,48) = 69.7$, $p < .001$]. Of course in the latter case, the generalities in rending problems were actually rejected at a high level.

The reduced preference for generalities that express property assignment is in keeping with prior work on combining problems (e.g., Revlis & Hayes, 1972; Revlis et al., 1971), which is consistent with the notion that property relations are either more arbitrary than universally quantified class relations or more complex (e.g., when drawing inferences from property relations, the reasoner may need to be reminded about the class relations that are presupposed; Sloman, 1998).

EXPERIMENT 2B

In assessing the reasoning strategy employed by students on the rending problems, we sought to create statements that reflected everyday relationships (e.g., sports teams, amusement parks, etc.). These statements contrast with the materials previously used, such as those in Experiment 1, which were based on dictionary definitions of categories that reflect a combination of natural and artifactual kinds. It is possible that the relative preferences shown for combining and rending problems reflect this difference in the nature of the categories reasoned about. To test for this possibility, the class-inclusion statements from Experiment 1 (combining problems) were rephrased as rending problems and presented to students sampled from the same introductory psychology class as in the previous experiments.

Method

A group of 27 student volunteers were asked to solve 24 counterfactual conditionals. The procedure, instructions, and structural composition of the booklets were identical to those in Experiment 2A. However, the materials were based on the class-inclusion generalities shown to students in Experiment 1. For example, the combining problem shown in statement 1 was rewritten as statement 1′:

(1) All whales are mammals
This animal is not a whale
This animal is not a mammal
*Assume this animal is a whale

(1′) All whales are mammals
This animal is a whale
This animal is a mammal
*Assume that this animal is not a mammal

Results

Reasoners' preferences for generalities in resolving counterfactual conditionals are presented in Table 4, which shows that there was no reliable preference (compared with chance expectations) for general statements (or for particulars) either overall or separately for affirmative or negative generalities on these rending problems. Neither polarity nor order reached conventional levels of significance with these materials. Table 4 includes summary values for the same materials when they were part of combining problems. Clearly, reasoners accepted the same generalities more often when the statements were part of combining problems than when they were part of rending problems [$F(1,53) = 23.8$, $p < .001$]. Taken jointly, there was no overall effect of polarity or order or any interaction between these variables.

Discussion

When presented with belief-contravening problems in which the counterfactual assumption explicitly disconnects an instance from its preexisting category, reasoners tend

Table 4
Percent Acceptance of Generality for Combining Versus Rending Counterfactuals With Identical Generalities

	Affirmative Generalities				Negative Generalities			
	First Mention		Second Mention		First Mention		Second Mention	
Logic	%	*SD*	%	*SD*	%	*SD*	%	*SD*
Combining	92.7	12.5	95.8	9.0	89.3	15.6	87.5	14.8
Rending	63.1	38.3	56.2	36.7	65.4	32.9	59.3	34.7

either to exhibit no systematic preference when the statements express class inclusion or deny the general statement when the statements express property assignment. This is the opposite of the entrenchment shown for combining counterfactuals that possessed identical generalities. The pattern for rending counterfactuals reflects somewhat the type of basic material reasoned about: Experiment 2A employed familiar common relationships (*All members of the Chicago Bulls are members of the NBA*), whereas generalities in Experiment 2B were familiar dictionary definitions (e.g., *All whales are mammals*). Students showed a preference for the particularist path for the former and no preference for the latter. However, a comparison of combining problems in Experiment 2 with rending problems in Experiment 2B that possessed identical generalities shows that the overall characteristic of the reasoning problem exercises a more powerful effect on the reasoners' decisions than does any incidental aspect of the statements themselves.

Which of the previously mentioned models of inference can account for the preferred paths taken on rending problems? The predictions from all of the approaches (in their present form) other than the NDS are the same on the rending as they are on the combining problems (Table 1). Consequently, only the NDS predictions fit the pattern on the rending problems, and so merits a more detailed description.

The prescriptions from the natural deduction approach for the combining problems might be as follows:

Rule and assumption are together as: $[p \rightarrow q, p]$
Particular fact: $[\sim q]$

Reasoning from the rule and the assumption allows the application of modus ponens, which produces $[q]$. This directly contradicts the fact $[\sim q]$, which is then rejected. This prediction was observed in Experiment 1 to occur more than 90% of the time. In contrast, on the rending problems, NDS specifies that

Rule and assumption together: $[p \rightarrow q, \sim q]$
Particular fact: $[p]$.

If modus tollens were directly available, it would produce $[\sim p]$, which directly contradicts the particular fact, which would be rejected. However, one characteristic of current NDS models is that they do *not* include the modus tollens rule directly in their models. To reason through situations calling for modus tollens requires the reasoner to employ multiple-step proofs (e.g., IF elimination as well as NOT introduction; see Rips, 1994). If reasoners in the present study are employing natural deduction operators in their efforts to revise the beliefs in the rending problems, they will have greater difficulty in working through the problems than in the combining problems and will likely show little preference for either path, which is what we found on class-inclusion problems.

EXPERIMENT 3A

The present experiments have thus far demonstrated that counterfactual reasoning proceeds in a systematic manner consistent with the reasoning context. For example, students doing belief revisioning will give a higher priority to a universally quantified relation of the form *All X are Y* in the context of a combining counterfactual than in the context of a rending counterfactual. There is a broader context, however, than the ones considered here: In the present study, all of the propositions can be said to have a real-world belief value. Are the preferences in belief revisioning dependent on the propositions having a real-world truth value, or is the pattern of accepting generalities based primarily on their role within the structure of the problem—independent of their believability? This issue is important to the credibility of the modal logic framework, which provides more than a syntactic analysis, requiring an ordering of propositions in terms of degrees of necessity, which can be computed only if the propositions have a place within the space of a real domain—a constraint not placed on the NDS. Experiment 3 addresses this question: If the generality is arbitrary (merely possessing a universal quantifier), would it be accorded the same priority as one with a real-world belief value? Moreover, under these conditions, the problems have only the superficial form of a counterfactual conditional, but are not truly belief contravening since there is no prior belief held about the statements. To evaluate this, we asked students to solve counterfactual conditionals with propositions that were sufficiently abstract that they had no real-world truth value, although they were understandable and contain real-world terms (e.g., *All animals in this forest are mammals*).

Method

Design and Procedure. Two groups of subjects were asked to solve 24 counterfactual conditionals. One group solved combining problems and the other solved rending problems. These problems were identical to the class-inclusi on problems studied in Experiments 1

and 2B except that the subjects of all propositions were abstract (e.g., *creature*, *organism*, *substance*) and the generality and counterfactual assumption contained a locative as illustrated in the following:

(6) All animals in the forest are mammals
This creature is not an animal in the forest
This creature is not a mammal
*Assume that this creature is an animal in the forest

In all other respects, the procedures were identical to those in Experiments 1 and 2. These materials were intended to have no a priori belief value. However, some of our subjects observed that the generality seemed slightly implausible, a fact that may have contributed to their decisions (see below).

Subjects. Fifty-six students participated in this study to fulfill a course requirement. They were selected from the same course as in previous experiments.

Results

Students' preferences for identifying generalities as true are presented in Table 5 for both combining and rending problems. It shows that students' preference for generalities varies with the logic of the problem [$F(1,55) = 42.9$, $p < .001$]. Students show a modest, nonreliable preference for generalities when they are expressed in combining problems (54.7%) and a distinct rejection of the same generalities when they are expressed in rending problems (12.6%, $p < .001$, binomial).

Table 5 shows that although there was no overall effect of polarity, polarity did interact with the logic of the problem [polarity × logic: $F(1,55) = 17.8$, $p < .001$] and the order in which generalities appeared in the problem [polarity × logic × order: $F(1,55) = 5.9$, $p = .01$]. These interactions require a separate analysis for each of the two logic problems.

Combining problems. When solving abstract combining problems, students accept affirmative generalities more often than negative ones [$F(1,33) = 7.6$, $p < .01$], with no effect of order. In contrast with Experiment 1, only affirmative generalities, in the second position on the page, were accepted more often than chance (affirmative: 63%, $p = .05$, binomial test). Also in contrast with Experiment 1, negative generalities, in the second position on the page, were *rejected* at greater than chance (negative: 44%, $p = .01$, binomial test). Clearly, the preference for generalities shown in previous combining problems was substantially reduced when the believability of propositions was neutralized and students were unsystematic in their preferences.

Rending problems. Students were presented abstract rending problems in the same logical form as in Experiment 2. Every student showed a distinct preference for *rejecting* the generality (acceptance = 12.6%, equivalent to a rejection rate of 87.4%, $p < .001$, binomial test). In belief revisioning with rending problems, students reject affirmative generalities more often than negative ones [$F(1,27) = 17.5$, $p < .001$]—in contrast with combining counterfactuals. There was a greater preference (less rejection) for negative generalities when they were mentioned second on the page, contributing to a polarity × order interaction [$F(1,27) = 6.9$, $p = .01$].

These findings demonstrate that reasoners do not view entrenchment of generalities as a viable basis for making decisions on rending counterfactuals when the material is abstract, and, in these cases, students prefer reasoning with particular statements.

Discussion

This experiment addressed whether the psychological processes for resolving counterfactual conditions are syntactically based or are sensitive to prior believability of the statements reasoned about. Students were asked to reason with statements that have no prior believability (e.g., *All creatures in this forest are reptiles*). For counterfactuals that combine categories, students showed no reliable preference for one resolution path over another. In contrast, for counterfactuals that rend categories, students showed a reliable preference for the *particularist* path—retaining particular statements of "fact" and rejecting general ones.

These preferences are distinctly different from those shown in Experiments 1 and 2. In reasoning counterfactually about assumptions that combine instances with categories, the preference for generalities was significantly greater when they were expressed as real statements with a prior positive believability than when they were arbitrary and abstract [$F(1,60) = 41.6$, $p < .001$]. When reasoning counterfactually about assumptions that rend categories, students showed a substantially lower tendency to reason with generalities when they had no prior believability than when they did [$F(1,51) = 52.6$, $p < .001$]. Recall that when rending categories did have some prior believability, students showed no reliable preference for one path over another. The implication of the present findings for the three primary reasoning approaches will be taken up as part of the General Discussion.

EXPERIMENT 3B

Glancing back over this sequence of experiments, we notice that only the decisions for combining counterfac-

Table 5
Percent Acceptance of Generalities for Combining Versus Rending Counterfactuals With Abstract Propositions

	Affirmative Generalities				Negative Generalities			
	First Mention		Second Mention		First Mention		Second Mention	
Logic	%	*SD*	%	*SD*	%	*SD*	%	*SD*
Combining	60.9	38.5	63.8	34.3	50.2	32.7	46.0	35.4
Rending	6.7	15.8	4.3	10.3	15.7	18.4	23.6	22.3
Combining replication	90.7	14.9	85.3	20.5	87.5	19.8	85.9	17.9

tuals in Experiment 1 show a reliable preference for generalities. Since this finding is critical to the tests of the various accounts of belief revision, it might be useful to replicate those results before drawing any conclusions about the importance of concreteness of the materials.

Method

Design and Procedure. Subjects were given problems and instructions identical to those in Experiment 1.

Subjects. Twenty-seven undergraduate volunteers from introductory psychology classes were asked to participate to fulfill a course requirement.

Results

The results for this group are presented in Table 5. A comparison with the preferences exhibited in Experiment 1 shows no main effects or interactions. Overall, students preferred to retain generalities that were believable when reasoning with combining counterfactuals (87.3%, $p < .001$, binomial). We compared this replication group with the students from Experiment 3A, who solved combination problems that contained arbitrary relations (Table 5) and again found that the preference for reasoning with arbitrary generalities was significantly less than that for reasoning with believable generalities [$F(1,64) = 33.7$, $p < .001$].

Discussion

Once again we have demonstrated that when reasoning with combining counterfactuals, students show a preference for reconciling inconsistencies by retaining the most general statement and rejecting the particular fact. In contrast, when the generalities and facts are of such abstractness so as to not possess prior believability, students exhibit no distinct preference in counterfactual reasoning.

GENERAL DISCUSSION

The belief-contravening problems examined in this study were designed to be paradigmatic of the kind of counterfactual reasoning found in scientific inference and hypothesis testing (see, e.g., Farris & Revlin, 1989, 1991); in *reductio ad absurdum* arguments in mathematics; and belief-revision in AI models of database management (e.g., Elio & Pelletier, 1997). The importance of counterfactual reasoning is not restricted to these technological domains but may also constrain our ability to pretend (Vygotsky, 1953) and provide the basis for our reassessment of past actions and our planning for the future (Roese & Olson, 1995). Counterfactual reasoning potentially allows us to gain new understandings and new knowledge and an opportunity to adjust our system of beliefs. In spite of the importance and ubiquity of this kind of thinking, there are no guidelines—at least not from standard logic—that tell us how to accomplish it. This is true, in part, because in truth-functional logic, if the premise is false, the validity of the conclusion cannot be guaranteed (see review of such methods by Schneider, 1952).

The present study employed a paradigm in which reasoners had to choose between competing alternatives to make consistent what a counterfactual conditional disrupts. Although it is an empirical question whether such binary choices occur as part of counterfactual reasoning in natural settings, the present procedure is in keeping with philosophical treatments of counterfactuals (e.g., Rescher, 1964) and the ubiquity of the dictum *ceterus paribus* (e.g., Simon & Rescher, 1966), which presupposes binary choices.

The present study evaluated the predictive accuracy of three different approaches to understanding human counterfactual reasoning: predicate calculus, modal logic, and NDS. Table 6 presents a summary of predictions and observations of the three approaches.

The Three Approaches

Standard predicate calculus. This approach makes no psychological claims about resolution of belief-contravening problems since it acknowledges only that no valid conclusion can be specified when the premises are false—the quintessential case of counterfactual reasoning. It is reasonable to assume, therefore, that if predicate calculus rules were embedded in a minimal process model, that model would predict preferences at chance levels for the generalist or particularist solution paths, a prediction that holds only on rending counterfactuals with substantive content in Experiment 2B. Table 6 shows that predicate calculus fails to account for distinct preferences for the generalist path on combining counterfactuals in Experiments 1 and 3B and for the content effect shown across the experiments of the study. This is noteworthy since predicate calculus is syntactically based and would anticipate identical preferences for both types of counterfactuals independent of content. Consequently, predicate calculus does not offer a viable approach to counterfactual reasoning—an observation long noted by philosophers and an import factor leading to the application of modal logic to the analysis of counterfactual conditions, to which we now turn.

Modal logic. A psychological model that embodies the principles of a modal logic analysis is the GCM (Revlis,

Table 6
Predictive Accuracy of Three Approaches to Counterfactual Reasoning

	Combining Counterfactuals		Rending Counterfactuals	
Approach	Believed	Not Believed	Believed	Not Believed
Predicate calculus	—	—	✓	—
Modal logic (GCM)	✓	?	✓	—
Natural deduction	✓	✓	✓	—

Note—GCM, generality coding model. Check marks indicate that the models' prediction accord with the data.

1974; Revlis & Hayes, 1972), which has previously been shown to provide a good account of reasoning with combining problems of the sort given in Experiments 1 and 3B (e.g., Revlis, 1974). But until now it has not been tested with rending counterfactuals.

The GCM asserts that counterfactual inference proceeds through three stages. First, the reasoner is said to construct a possible world in which the counterfactual assumption can be true. Second, the reasoner orders the relevant available descriptions about that world in terms of his/her ability to structure the domain—that is, in terms of modal categories (Rescher, 1961, 1964). Third, the reasoner seeks to reconcile inconsistencies among beliefs in order to reinstate a truth-based (i.e., alethic) homeostasis. Where a conflict exists among relevant statements, the reasoner retains the one with the lowest modal status (the most necessary proposition). For combining problems examined in Experiment 1 (and replicated in Experiment 3B), the contrast in modal status between the key conflicting statements is readily discriminable—the generality versus the particular statement. Therefore, the GCM predicts that the reasoner will retain the generality and reject the contrasting particular statement. This is what students do an overwhelming proportion of the time when solving combining counterfactuals with believed generalities.

The GCM was designed to account only for combining counterfactuals, and it does not address rending counterfactuals. However, it is capable of making predictions for rending counterfactuals if it is allowed to reevaluate the modal status of the contrasting particular propositions. To wit: In the rending case, the generality again has the highest degree of necessity (e.g., *All whales are mammals*). However, the contrasting particular has lost only its distant category membership by virtue of the counterfactual assumption (e.g., *This animal is no longer a mammal*); it still possesses its original, immediate category membership (e.g., *This animal is a type of whale*). In this case, the contrast is between two immediate category memberships: the generality, which assigns all members of one category into another superordinate, and the particular statement, which maintains the assignment of an instance to an immediate superordinate. Clearly, this choice is less discriminable to a modal-sensitive inference engine than in the combining case. The difficulty of the choice is reflected in the modest preference shown by students for accepting the generality and rejecting the particular in Experiment 2B—in keeping with the predictions of the GCM.

How can the GCM predict the reasoning pattern shown in Experiment 3A, where the generality and particulars had no prior believability and where the generality might even be implausible? Let's consider the combining counterfactuals first. Here, the generality is superficially lawlike, but unlike in the other experiments, the generalities in this experiment do not hold across time and space, and at best are accidental (*All animals in this forest are reptiles* or like the property-assignment generalities in Experiment 2A: *All TV talk shows have hosts*) and in the case of 3A, may actually be implausible. These generalities do not possess the degree of necessity expected of scientific laws, nor do the contrasting particulars (e.g., *This creature—an animal in the forest—is not a reptile*). Indeed the counterfactual assumption may establish an "accidental" universe of discourse where generalities cannot organize plausible constraints, so that the modal status of all statements is roughly equivalent. In this situation, the GCM would predict no discernible preference for one revision path over another. Table 5 shows that this is close to the observed preferences, where only affirmative generalities were reliably retained.

Rending counterfactuals in Experiment 3A, whose generalities are not believed, pose a problem for the GCM. If particulars are more plausible than generalities (as our results suggest), then the GCM would predict that students should prefer them—as they do on rending problems. However, the same prediction should apply to not-believed combining counterfactuals in Experiment 3A. But, rather than rejecting the generalities wholesale, reasoners still exhibited a slight preference for affirmative generalities and no reliable preference for particulars. Consequently, the GCM cannot account for the decisions on *both* the combining and the rending problems in Experiment 3A. Hence, there is a question mark placed in the appropriate cell of Table 6 for modal logic, which shows that the GCM accounts for only 2+ of the four conditions in this study.

Natural deduction system. Using this approach, combining problems might be represented as the conjunction of the counterfactual assumption [*p* ("X is P")] and the generality, [*If P then q* ("All P are Q")]. Application of the equivalent of a modus ponens rule (e.g., forward IF elimination in Rips, 1994) would generate the proposition [*q* ("X is Q")]. This would require the rejection of the particular, [~*q* ("X is not Q")]. If the reasoner took the particularist path and combined the assumption with the particular [*p* and ~*q*], then an effort to prove the generality would fail (backward *if* elimination). If forward rules are applied prior to backward ones in systems such as PSYCOP, then reasoners would follow the generalist path in combining counterfactuals. As such, Table 6 shows that NDS predicts all of the preferences in combining counterfactuals in Experiment 1.

Rending counterfactuals should prove more difficult to resolve than combining problems for NDS of the sort proposed by Rips (1994), since it does not provide a direct proof of modus tollens arguments. Hence, NDS does not predict a preference for either path on rending counterfactuals, which is consistent with the findings of Experiment 2.

The preference for different solution paths when the materials are abstract pose difficulties for the natural deduction model, but not insurmountable ones. This is a case where the believability of the individual propositions may intrude on the proof-theoretic processing and focus the reasoner's attention on the more plausible particular statements and create a preference to reason from those state-

ments. The two most prominent psycho–logic proposals in psychology—Braine and O'Brien (1998) and Rips (1994)—have built-in flexibility in the ordering of rules/operators and, in the case of Braine and O'Brien, include a set of reasoning heuristics that come into play when belief is particularly salient and can adroitly be used to account for the knowledge-based processing in the selection task (e.g., Cheng & Holyoak, 1985; Cosmides, 1989). For example, on combining problems, where there is a competition between modus ponens based generalist path and believability, only a marginal preference for the generalist path will be observed. For rending problems, where natural deduction shows no a priori preferred strategy, believability takes precedence, and students show a preference for the particularist path. In this way, the natural deduction approach can sensibly distinguish between rending and combining counterfactuals and between believed and not-believed statements.

The choice between the GCM and the NDS is not easily made. The former is more sensitive to the semantics of the statements, the latter to the formal structure of the reasoning task. However, since natural deduction can be made to accommodate prior believability, it holds great promise for a model of counterfactual reasoning.

Alternative Accounts

Belief strength. A simple model of belief revisioning would have the reasoner retain those statements that are most believable, either by a direct access of the information from semantic memory or on the basis of plausibility by associated stored facts—with the caveat that the counterfactual assumption, which is disbelieved, must be retained. This approach would predict that the same statements that are retained when reasoning with combining counterfactuals would be retained when reasoning with rending counterfactuals. A comparison of Experiment 1 with Experiment 2B shows that strongly believed generalities (based on dictionary definitions) are accepted significantly more often than chance when they are part of combining counterfactuals and accepted only at chance levels when they are part of rending counterfactuals. This difference in entrenchment pattern in the two types of counterfactuals is also seen when the propositions have no particular believability. For example, Experiment 3A contained generalities that either were of no particular believability or may not have been totally plausible (e.g., *All creatures in this forest are reptiles*). These statements are retained at chance levels within combining counterfactuals and are rejected at significantly greater than chance levels when part of rending counterfactuals. Note that believability may exercise some influence on the entrenchment of a statement since over all conditions, believed generalities are treated with greater deference than are generalities with no prior believability (e.g., contrast rending counterfactuals in Experiments 2B with those in Experiment 3A and combining counterfactuals in Experiment 3A with with those in Experiment 3B). However, taken together, these findings imply that the believability of the generalities is insufficient to account for the entire pattern of acceptances.

Alternately, it is possible that the *relative* believability of statements within a belief-contravening problem dictates which ones will be retained and which will be sacrificed for consistency. Within the combining counterfactuals, the particulars in Experiments 1 and 3B are clearly less believable than the generalities (e.g., *This animal—which is conjectured to be a whale—is not a mammal* vs. *All whales are mammals*). So, on the basis of relative believability, one would anticipate that generalities would be preferred within the context of combining counterfactuals. Within rending problems, the contrast is similar to that for the combining problems (*This animal—which is conjectured to be not a mammal— is a whale* vs. *All whales are mammals*). Although the contrast in believability may not be identical between combining and rending problems, the substantial difference between reasoners' preferences on the two types of contexts could not sensibly be attributed to the minor differences in belief contrast.[3] Indeed, the particular statements for the rending counterfactuals in Experiment 2A are accepted at the same rate as the generalities in those problems; yet, the particulars have no prior believability (*This player is a member of the Chicago Bulls*), whereas the generalities are highly believable (*All members of the Chicago Bulls are members of the NBA*). In sum, though absolute believability or contrastive believability may be important to the counterfactual judgments, the reasoning context appears to be more important than the belief value of the statements per se.

Minimal distance metrics. One approach for revising beliefs is to employ methods that minimize the overall change in the belief network. This technique is frequently applied in the AI literature (see review by Elio & Pelletier, 1997), in which a resolution procedure organizes the set of options (here two paths) based on which disrupts the fewest number of initial beliefs. This approach is untestable in the present paradigm since either solution path disrupts the same number of statements. More broadly, minimal distance approaches need to give a reasoned basis for estimating how much weight a single piece of data possesses. How much disruption in the network of beliefs should be accorded a single disconfirming instance? Without such reasoned bases for estimating disruption, minimal distance is not a viable alternative.

Mental models. This approach claims that the logical decisions reached by reasoners are derived from semantic interpretations of sentences (called "models") and the mental manipulation of the elements of those interpretations (e.g., Johnson-Laird, 1983). Research in this tradition has not focused on the kinds of problems presented here but may be well suited to this form of inference. To test for this possibility, we took it upon ourselves to develop the following mental models account of counterfactual reasoning. We assume that the reasoner constructs models of possible worlds as part of the process of coun-

terfactual reasoning. Upon reading the counterfactual assumption, the reasoner begins with initializing conditions that allow the integration of incoming information with existing knowledge to construct a coherent representation (Byrne et al., 2000). Statement 7 shows an ordering of propositions for combining counterfactuals, though admittedly not in the formalism followed by Johnson-Laird and his associates.

(7) All X are Y
This A is not Y
*This A is an X

The reasoner's interpretation of the propositions orders them with the lawlike generality at the top because, by virtue of being universally quantified, these statements describe the most common state of affairs; hence they express the *standard*, or normative, situation. The counterfactual assumption directs the reasoner's attention to any model that has the X term so that it can be linked with a model that contains the A term. In searching for the X-model, it finds the generality at the top of the stack (the order in which these statements are physically presented in the problems is not relevant here—but see Girotto et. al., 1997). For most reasoners, the generality provides a sufficient basis for organizing the hypothetical universe of discourse. The particular is rejected because it is implicitly in conflict with the generality. Hence the high degree of entrenchment of generalities in combining counterfactuals with believable content.

Rending counterfactuals with believed generalities have the following model configurations:

(8) All X are Y
This A is an X
*This A is not Y

The search to bond the not-Y relation fails in statement 8. The student is left with no matching propositions and, because the generality is at the top of the configuration, may be anticipated to show a modest preference for the generality (the *standard* model), which is characteristic of the choices made in Experiment 2.

The foregoing is illustrative that a mental models approach with simple configurations of propositions (the models) and a goal can account for the general regularities in the preferences shown on combining and rending counterfactuals. Note that the arrangement of models corresponds closely to that for the modal logic analysis. This should not be a surprise since Bell and Johnson-Laird (1998) have demonstrated that mental models can offer an account of modal reasoning. Viewed in this context, the present study can be interpreted to show how students prefer to organize their mental models in a more formal counterfactual reasoning environment.

Conceptual integration—blending. The description of counterfactual reasoning embedded within the GCM bears a similarity to the more ambitious treatment of conceptual integration networks by Fauconnier and Turner (1998). "Blending" is said to be a general cognitive operation that constructs networks of connected spaces. In that system, common information is extracted from diverse input sentences under the guidance of a *generic space*, which in turn marks these elements to be combined within a *blended space.* It is these blended spaces that allow for inferences, including counterfactually based ones. In a sense, the blended spaces correspond to the possible worlds described as part of the first stage in counterfactual inference in the GCM. The second stage in the reasoning process involves ordering the propositions in terms of alethic categories (degrees of necessity). Although such an ordering is not to our knowledge strictly specified within the conceptual integration network, reasoners are certainly capable of distinguishing between *necessity* and *possibility* (e.g., Bell & Johnson-Laird, 1998; Osherson, 1976), and there is no impediment to the inclusion of such ordering within the broader system described by Fauconnier and Turner (see also Fauconnier, 1997) since the conceptual projections are sensitive to causal relations and ontological categories.

The third stage in the GCM necessitates the selection of statements to be retained and others to be rejected under the impetus of the counterfactual assumption. We assume that such a process may be said to occur within the network as an effort to satisfy optimality constraints by creating an integrated blended space (Fauconnier & Turner, 1998). In sum, the GCM could be construed as a specific instantiation of the integration and blending processes that are generally described as part of the conceptual integration network.

The foregoing treatment suggests that although the best account of counterfactual reasoning is provided by the NDS, it may also be adequately treated by either natural mental models or modal logic analyses. In the latter case, counterfactual reasoning would be an example of conceptual blending.

Conclusion

Counterfactual situations occur when we are asked to consider the implications of a proposition that we are entertaining "for the sake of argument" though we believe it is false. Such reasoning requires at least a temporary revising of our beliefs. The present study examined some of the psychological processes involved in belief revising by employing a reasoning environment called belief-contravening problems. When students were asked to consider their beliefs in the face of a counterfactual assumption that combined instances and categories in new ways, they tended to protect the most general, lawlike propositions and sacrifice more particular, less necessary propositions in an effort to maintain consistency. In contrast, when students had to reason from counterfactual assumptions that directly dislodged instances from their superordinate categories, they tended either to show no preference for maintaining one proposition over another or to show a distinct preference to reason from particulars.

The purpose of the present study was not to decisively eliminate one theoretical perspective or another, but to establish a task domain that characterizes counterfactual reasoning and to illuminate how the different approaches may frame our understanding of how we reason about situations we believe to be false. The study shows that when we conjecture about what "might have been" or what "may be," our decisions are rational consequences of having imagined a possible world of propositions, models, or spaces.

REFERENCES

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science*, **22**, 25-51.

Braine, M. S., & O'Brien, D. P. (Eds.) (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Byrne, R. M., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000). The temporality effect in counterfactual thinking about what might have been. *Memory & Cognition*, **28**, 264-281.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, **17**, 391-416.

Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind*, **55**, 289-307.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, **31**, 187-276.

Elio, R., & Pelletier, F. (1997). Belief change as propositional update. *Cognitive Science*, **21**, 419-460.

Evans, J., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory & Language*, **35**, 392-409.

Farris, H., & Revlin, R. (1989). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory & Cognition*, **17**, 221-232.

Farris, H., & Revlin, R. (1991). Rule discovery strategies: Falsification without disconfirmation (reply to Gorman). *Social Studies of Science*, **21**, 565-567.

Fauconnier, G. (1997). *Mappings in thought and language*. Cambridge: Cambridge University Press.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, **22**, 133-187.

Girotto, V., Mazzocco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition*, **63**, 1-28.

Goodman, N. (1952). The problem of counterfactual conditionals. In L. Linsky (Ed.), *Semantics and the philosophy of language* (pp. 231-246). Urbana: University of Illinois Press.

Hewstone, M., Hopkins, N., & Routh, D. A. (1992). Cognitive models of stereotype change: I. Generalization and subtyping in young people's views of the police. *European Journal of Social Psychology*, **22**, 219-234.

Hewstone, M., Macrae, C. N., Griffiths, R., & Milne, A. B. (1994). Cognitive models of stereotype change: V. Measurement, development, and consequences of subtyping. *Journal of Experimental Social Psychology*, **30**, 505-526.

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

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, **10**, 64-99.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.) Chicago: University of Chicago Press.

MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 633-642.

Osherson, D. N. (1976). *Logical abilities in children: Reasoning and concepts*. Hillsdale, NJ: Erlbaum.

Peirce, C. S. (1877). The fixation of belief. Reprinted in Charles Hartshorne & Paul Weiss (Eds.), *The collected papers of Charles Sanders Peirce* (pp. 223-247). Cambridge, MA: Harvard University Press.

Redding-Stewart, D., & Revlin, R. (l978). Hypothetical inference and category structure. *Bulletin of the Psychonomic Society*, **12**, 465-467.

Rescher, N. (1961). Belief-contravening suppositions. *Philosophical Review*, **70**, 179-195.

Rescher, N. (1964). *Hypothetical reasoning*. Amsterdam: Elsevier, North-Holland.

Revlis, R. (1974). Prevarication: Reasoning from false assumptions. *Memory & Cognition*, **2**, 87-95.

Revlis, R., & Hayes, J. R. (1972). The primacy of generalities in hypothetical reasoning. *Cognitive Psychology*, **3**, 268-290.

Revlis, R., Lipkin, S., & Hayes, J. R. (1971). The importance of universal quantifiers in a hypothetical reasoning task. *Journal of Verbal Learning & Verbal Behavior*, **10**, 86-91.

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

Roese, N., & Olson, J. (1995). Counterfactual thinking. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 1-55). Hillsdale, NJ: Erlbaum.

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145-181). Hillsdale, NJ: Erlbaum.

Ryle, G. (1949). *The concept of mind*. New York: Barnes & Noble.

Schneider, E. (1952). Recent discussions of subjunctive conditionals. *Review of Metaphysics*, **6**, 623-647.

Simon, H. A., & Rescher, N. (1966). Cause and counterfactual. *Philosophy of Science*, **33**, 323-340.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, **35**, 1-33.

Sternberg, R. J., & Gastel, G. (1989). If dancers ate their shoes: Inductive reasoning with factual and counterfactual premises. *Memory & Cognition*, **17**, 1-10.

Thagard, P. (1989). Explanatory coherence. *Behavioral & Brain Sciences*, **12**, 435-502.

Vygotsky, L. (1953). *Thought and language*. (G. Anscombe, Trans.). New York: Macmillan.

NOTES

1. It should be noted that counterfactual reasoning may be partially syntactic since students show entrenchment of nonsensical (amphigorous) generalities (e.g., *All frups are toves*) when they are contrasted with similar particulars (Revlis & Hayes, 1972).

2. Modus tollens is present in mental logic (Braine & O'Brien, 1998), but is relegated to the set of pragmatic inference rules, which are an adjunct to the basic system.

3. Although believability ratings are possible for the generalities in this study (they were based on dictionary definitions), believability ratings for the particular facts cannot be gathered without creating an appropriate context for the statements, which would be equivalent to presenting the reasoning problem in question.

(Manuscript received September 27, 2000;
revision accepted for publication June 12, 2001.)