On the epistemic entrenchment of different types of knowledge expressed as conditionals in belief revision tasks

Renée Elio

ree@cs.ualberta.ca

Technical Report TR96-16

August, 1996

Department of Computing Science

University of Alberta

Edmonton, Alberta , Canada T6G 2H1

Abstract

Some belief revision theories appeal to the notion of epistemic entrenchment as a guide to choosing among alternative ways of removing inconsistency that new information may cause with existing beliefs. While belief revision theorists may not be interested in natural language uses of conditionals *per se*, the appeal to epistemic entrenchment *because* certain kinds of knowledge (e.g., physical laws) are expressed in conditional form opens the door to a more careful consideration of whether the syntactic form itself serves as a useful cue, even in the mind of the researcher, for epistemic entrenchment principles. This study determines whether there is any empirical support for the notion that the *type* of knowledge expressed in a statement can serve as the basis for epistemic entrenchment principles. Four types of knowledge—promises, causal relationships, familiar definitions and unfamiliar definitions—were expressed in a common syntactic *if p then q* form. A belief revision task was given to people, in which these conditionals were used to define "initial belief" sentences, which were then followed by a "new information" sentence that created an inconsistency with the initial set of beliefs. The frequency with which people chose to disbelieve the conditional (or lower their degree of belief in it) as a way of resolving the inconsistency depended on the type of knowledge—causal, definitional, or promises—that the conditional expressed. The conditionals that expressed causal information were further analyzed according to the possible alternative causes and disabling factors associated with the causal relationship. These more subtle distinctions also affected how people revised the belief sets using causal scenarios. For normative belief revision models, such findings call into question the notion that conditionals ought to be more entrenched *by virtue of* their syntactic form. They also question whether the syntactic *if-then* form of conditionals can even serve as a useful cue for signaling the types of knowledge that it might be plausible to entrench, such as causal relationships. These results support higher-order epistemic entrenchment principles that distinguish among *types* of knowledge (regardless of the syntactic form in which they are expressed) and known necessity and sufficiency aspects of causality relationships in particular.

On the epistemic entrenchment of different types of knowledge expressed as conditionals in belief revision tasks

Belief change—the process by which a rational agent makes the transition from one belief state to another—is an important component for most intelligent activity done by epistemic agents, both human and artificial. When such agents learn new things about the world, they sometimes come to recognize that new information extends or conflicts with their existing belief state. In the latter case, rational reasoners would identify which of the old and new beliefs clash to create the inconsistency, decide whether in fact to accept the new information, and, if that is the choice, to eliminate certain old beliefs in favor of the new information. Alternatively, new information may not create any inconsistency with old information at all. In this case, the reasoner can simply add the new information to the current set of beliefs, along with whatever additional consequences this might entail.

Belief revision characterized in this manner is a widely-studied area within AI. Some of the seminal work in this area was done by Alchourrón, Gärdenfors, and Makinson (1985) (hereafter "AGM"), who proposed a set of rationality postulates for belief revision theories. These postulates specify certain types of logical consistencies that ought to hold between old and new belief states, regardless of what specific belief revision operators a particular model might propose. Even within a model that adheres to these general rationality postulates, there may be more than one way to revise a set of beliefs to reconcile the inconsistency that new information presents. The rationality postulates do not speak to how such a choice should be made. Several theorists in this arena have appealed to the general notion of *epistemic entrenchment* to impose a preference ordering on the possible changes (Gärdenfors, 1984, 1988; Gärdenfors & Makinson, 1988; Nebel, 1991; Willard & Yuan, 1990). While there are differences in how various researchers are now viewing epistemic entrenchment, it has generally been defined as a total pre-ordering relation on all the sentences of

the language, which obeys certain postulates within the AGM framework. More informally, the appeal to epistemic entrenchment is aimed at capturing the intuitive idea that some statements are more deserving of belief—for whatever reason—than other statements, in the face of contradiction.

This paper considers the issue of developing principles for epistemic entrenchment that are based on the type of knowledge a statement is expressing, regardless of its syntactic form, and on number of factors known to impact the truth condition of the statement. A series of experiments explored three types of knowledge expressed in conditional form—promises, contingent universals, and causal relationships. The results of these experiments indicated that people do make different belief revision decisions, depending on these high-level distinctions. To further test the idea that *types* of knowledge may serve as epistemic entrenchment principles, conditionals following the three major categories (definitions, promises, and causal) were written for an domain unfamiliar to our subjects (anthropology topics). Not only were people able to classify conditionals into these knowledge-type categories without any familiarity with the domain of discourse, but different patterns of belief revision decisions emerged as well. The rest of this report details the background motivation for this research, as well the experimental methodology and results.

On Epistemic Entrenchment

Much work has been done in specifying postulates for epistemic entrenchment that in turn support the definition of revision operations. The aim in such works is to assess whether such revision operations obey AGM-like rationality postulates. There has been somewhat less attention given to considering what the epistemic principles themselves might actually be. Certain approaches to database consistency (Elmasri & Navathe, 1994) and some syntactic theories of belief revision (Willard & Yuan, 1990; Foo & Rao, 1988) advocate the entrenchment of the conditional form $p-->q$ over non-conditional forms. The intuition driving the idea of entrenching $p-->q$ over other types of sentences is not because material implication *per se* is important, but because "law-like relations" are often expressed in sentences of this form. For example, Foo and

Rao (1988) assign the highest epistemic entrenchment to physical laws, a strategy which may be especially effective in reasoning about how a dynamic world can change.

Other perspectives on epistemic entrenchment (Doyle 1991; Gärdenfors, 1988; Harman,1986) implicitly or explicitly take a sort of utility viewpoint, proposing that some beliefs are less deserving of disbelief because they are more useful, e.g., have higher explanatory power. But as Gärdenfors (1993) notes, with the exception of "expected utility", there has been little cognitive research concerning how "expectations about knowledge" might determine which of several belief revision options gives the most plausible consistent belief set. The work presented here addresses this issue. In particular, it investigates whether the type of knowledge expressed in a conditional statement, and the number of factors known to "disable" the relation between the antecedent and consequent, influence the kinds of belief revision decisions that people chose to make. These meta-knowledge considerations can be viewed as expectations or assessments about the plausibility of alternative epistemic states. The results presented here can provide an empirical foundation for defining at least some epistemic principles. They also offer a way to characterize and formalize some of the extra-logical factors needed to render plausible belief revision decisions.

Contrary to the intuition that conditional statements might be more deserving of entrenchment possibly because of the kind of knowledge they express, Elio & Pelletier (in press; 1994) found that people readily abandoned conditional statements as a way to generate a consistent belief set, particularly when the belief sets involved natural language content rather than abstract symbols. That study identified, but did not address, the question of whether different types of knowledge expressed in conditional form might be differentially resistant to disbelief in the face of contradictory evidence. Evans et al. (1993) note that the material implication view of conditionals has proven inadequate from both a linguistic and psychological viewpoint. They identify a number of uses of the conditional in natural language, such as contingent universals ("If an animal is a fish, then it is cold-blooded."), temporal/causal ("If a glass is dropped, then it will break."), advice ("If you work hard, then you will do well in life"), promises ("If you do your homework, I will let you watch TV."), threats, and warnings. As they and other researchers (Grice, 1975 ; Geis & Zwicky,

1971) have noted, some uses of conditional expressions invite inferences that are not sanctioned by the material implication treatment of conditionals (e.g., If you don't do your homework, it's reasonable to conclude that I won't let you watch TV. However, such an inference is not sanctioned by standard logic.)

As noted above, causal relations are often expressed in a conditional form. Reasoning about causal knowledge has itself been an active area of research by psychologists and is too extensive to review here in any detail. The work relevant to the present research is that done by Cummins and her colleagues (Cummins, Lubart, Alksnis & Rist, 1991; Cummins, 1995). They found that deductive reasoning about causality is affected by the number of alternative causes of the consequent and the number of disabling conditions — factors that prevent effects from occurring in the presence of viable causes. In particular, they examined four types of causal relations which were defined by crossing these two factors, namely causal conditionals that had (a) many alternative causes and many disabling conditions ("If the brake was depressed, then the car slowed down."); (b) many alternative causes and few disabling conditions ("If Alvin read without his glasses, then he got a headache."); (c) few alternative causes and many disabling conditions ("If the trigger was pulled, then the gun fired."); and (d) few alternative causes and few disabling conditions ("If Larry grasped the glass with his bare hands, then his fingerprints were on it."). Whether a particular causal statement fell into one or the other of these four categories was empirically determined by asking people to generate alternative causes and disablers for a collection of statements. When a different group of people rated the acceptability of conclusions generated from these statements falling into one of these four categories, Cummins et al. found that the acceptability ratings of modus ponens and modus tollens conclusions were influenced by the number of disabling conditions associated with a particular causal statement: the fewer the disabling conditions, the higher the acceptability of the inference (regardless of how many alternative causes there were). They also examined the acceptability ratings for two inferences that are not sanctioned under the material implication treatment of the conditional: denying the antecedent and affirming the consequent. The overall acceptability ratings for these "illegal"

inferences were generally quite low, but they were higher when a causal conditional had fewer alternative causes than if it had many alternative causes.

It seems plausible, as Cummins et al. (1991) note, that the influence of alternative-causes and alternative-disables is related to perceived causal necessity and sufficiency, which in turn determines deductive entailment relations. The possible import of these distinctions for belief revision and aspects of epistemic entrenchment seems clear: belief revision is required when that which is expected to be true (or false) is not. A consideration of disabling conditions is a knowledge-based, yet domain-independent, dimension of causal relations that could serve as the basis for epistemic entrenchment of causal conditionals. Such a notion in turn presumes that epistemic entrenchment is based on knowledge about *types* of knowledge.

The manner in which patterns of human reasoning are influenced by such natural language uses of conditionals, as well as by real world knowledge, is outside the scope of this report. So too is any argument about whether it would be plausible for even an artificial agent to make unsanctioned inferences for certain of these uses of the conditional. For present purposes, we simply need to acknowledge that a common syntactic form may express different types of knowledge, so that we can explore whether these different types of knowledge suggest different "rational" belief revision principles. While belief revision theorists may not be interested in natural language uses of conditionals *per se* when they construct normative models, the appeal to epistemic entrenchment *because* certain kinds of knowledge (e.g., physical laws) might expressed in conditional form opens the door to a more careful consideration of whether the syntactic form itself serves as a useful cue, even in the minds of the researcher, for epistemic entrenchment principles.

The experiments reported here use Elio and Pelletier's (in press; 1994) belief revision paradigm to examine the hypothesis that the tendency to "entrench" a conditional statement will be influenced by the type of knowledge expressed in the conditional form. The experiments presented people with a belief revision task, in which they were first given an initial "belief set" consisting of a conditional statement, a statement that permitted either a modus ponens or modus tollens inference, and the sanctioned inference. The conditional expressed either a promise, a familiar or

unfamiliar definition (what Evans, Bryne, & Newstead term "contingent universals"), or a causal relationship, with causal relationships further distinguished into subtypes defined by alternative causes and disabling factors, as per Cummins et al.'s studies. A "new piece of information" was then presented that contradicted the inference in the initial belief set. People indicated which of the initial beliefs—the conditional or the non-conditional statement that permitted the inference—they no longer believed, as a way of reconciling the contradiction. Previous work using this paradigm revealed that people were more likely to disbelieve the conditional than the non-conditional statement. The hypotheses examined here are (a) that the type of knowledge expressed in conditional statement will lead to different revision decisions, (b) that conditionals involving familiar definitions will be retained more often than those involving unfamiliar definitions, and (c) that the higher the number of disabling factors associated with a causal conditional, the more likely the causal conditional will be abandoned as a belief revision decision. Hypothesis (a) is merely a restatement of the general intuition guiding this study. Hypotheses (b) and (c) follow from a perspective that the agent involved in a belief revision decision considers the factors that can affect the truth status of a statement in a particular context: a familiar definition is more deserving of continued belief in the face of contradiction than an unfamiliar definition, because an agent cannot assess the relative likelihood that the unfamiliar definition is incorrect. A causal conditional that has few disabling factors may be more "entrenched" than one with many, because in the latter case, it is more likely that one of those disabling factors has come into play in a particular context. No specific hypotheses were formulated concerning the belief revision decisions involving promise conditionals; those decisions will be contrasted with the revision decisions made for definition conditionals and causal conditionals.

Experiment 1

Method

Stimuli Design. A set of 65 conditionals was constructed to include (a) the 16 causal conditionals used by Cummins (1995) and Cummins et al. (1991) and (b) other conditional statements that, in the experimenter's estimation, expressed familiar definitions, unfamiliar

definitions, and promises. To ensure that these intuitions matched those of subjects, this set of 65 statements was given to a group of 40 subjects, who classified each statement according to the type of knowledge it expressed.[1] These subjects received the statements in booklet form, with three statements appearing per page. Each statement appeared with a series of classification choices beneath it, in the following manner:

"If a substance is nitroglycerin, then its molecular structure is $C_3H_5(NO_3)_3$."

I would classify what this sentence is expressing as ....

a. a promise

b. a cause and effect

c. a prediction (not based on cause and effect)

d. a definition that I am familiar with

e. a definition that I am <u>not</u> familiar with

The choices included the category of "prediction" in case that subjects believed there was a temporal, but non-causal, contingency between a conditional's antecedent and consequent. An example prediction conditional might be "If the sky turns cloudy, then it will rain."

Appendix 1 gives the ratings of the 34 statements, used to construct belief revision problems, selected from the full set of 65 conditionals that were rated by subjects. This set of 34 statements consisted of (a) 16 causal conditionals, taken from Cummins et al., (b) 6 promises, (c) 6 familiar definitions, and (d) 6 unfamiliar definitions. As a side note, the ratings in Appendix 1 show that a few of the Cummins et al. causal conditionals were rated more frequently as predictions than as causal statements in the present norming study. These statements were still classified in the present studies as causal statements. Whether a statement expressed a prediction or a causal relation depends in part on what the person interpreting the statement knows: what seems

[1]The Cummins et al. (1991) causal conditionals were included in this set of 65 for completeness, even though these researchers had generated them through their own norming procedure as well.

like a (mere) temporal association to one person may be recognized as a statement of causality by another. The causal versus temporal association distinction is a critical one, because it reflects explanatory power in the former case that is absent in the latter, a feature that has been identified in other views of belief revision (Thagard, 1989). This distinction, while important, is not a focus of the present set of studies.

Belief Revision Problems. The conditional statements given in Appendix 1 were used to construct belief revision problems. Each belief revision problem had three main parts. First, a set of three statements appeared as an *initial belief set*. Second, a single *update sentence* was presented with the leading remark "You later discover that...." The update sentence always introduced a contradiction with an inference that was part of the initial belief set. Finally, three *revision choices* — ways of changing belief in the initial sentences so that they were logically consistent with the update sentence — were given. Table 1 shows two examples of belief revision problems as they appeared to subjects. The next paragraphs elaborate on these three parts of each belief revision problem: the initial belief sentences, the update sentence, and the revision choices.

The initial belief sentences took one of two forms. In the modus ponens form, the initial belief set consisted of a conditional statement $p -> q$, a statement asserting the antecedent p, and a statement asserting q as a consequence that follows from the other two statements. In the modus tollens form, the initial belief set consisted of $p -> q$, a statement asserting $\sim q$, and a statement asserting $\sim p$ as a consequence following from the other two statements. As in previous studies, the modus ponens and modus tollens inferences were explicitly included in the initial belief set. This is because the focus of this work is not on whether the inferences could be reliably drawn, but rather on patterns of retaining or abandoning the initial premises that supported the inference.

The update sentence always contradicted the inference in the initial belief set. In the modus ponens case, the update sentence was $\sim q$. In the modus tollens case, the update sentence was p.

Following the update sentence, three revision choices were listed. Each choice presented sentences marked as "Believe", "Disbelieve" or "Uncertain." In what will be referred to as the *disbelieve data/keep conditional* choice, the initial $p -> q$ conditional was marked "Believe" and

the initial non-conditional sentence (hereafter called the *data statement*) supporting the contradicted inference was marked "Disbelieve." In the modus ponens case, this amounted to disbelieving *p*; in the modus tollens case, this meant disbelieving ~*q*. In the *disbelieve conditional/keep data* choice, the initial *p*—>*q* conditional was marked as "Disbelieve" and the initial data sentence supporting the contradicted inference was marked "Believe." In the *Uncertain* choice, both the conditional and the initial data sentence supporting the contradicted inference were marked as "Uncertain." Table 1 shows examples of each of these choices.

Design. Each of the 34 conditional statements in Appendix 1 was used to construct both a modus ponens and a modus tollens belief revision problem. Subjects saw any given conditional in only one or the other of these forms. Feedback from a small set of pilot subjects was indicated that the full set of 34 problems was too long to hold subjects' attention through completion. In one booklet type, each of 38 subjects saw (a) all 16 causal conditionals, half in modus ponens form and half in modus tollens form, and (b) all 6 promise conditionals, three in modus ponens form and three in modus tollens form. In the other booklet type, another set of 35 subjects saw (a) all six familiar definitions, half in modus ponens form and half in modus tollens form; (b) all six unfamiliar definitions, half in each syntactic form, and (c) all 6 promise conditionals, half in each syntactic form. This allowed causal type and definition-type to serve as a repeated factors, for these two groups, respectively. Analyses of the promise problems used data from both these two groups combined, to increase the sample size. The order of the problems for a subject was randomized and the order of the revision choices for each problem was also randomized.

Subjects. A total of 73 subjects were recruited from the Department of Psychology's subject pool, and received credit for experiment participation as required by an introductory psychology course.

Procedure. The problems were presented in booklet form, one per page. Subjects were run in group sessions. The experimenter reviewed the written instructions at the beginning of the session and subjects worked at their own pace through the booklets. The instructions indicated that each problem consisted of a set of initial beliefs, followed by new information that contradicted the

initial beliefs, and then a set of choices that differed in how the new information could be reconciled with the initial beliefs. The instructions then presented an example modus-ponens belief revision problem, *If Fred goes to the dance, then he'll be happy, Fred went to the dance, Therefore, you also believe Fred will be happy,* with the new information *You later discover Fred is not happy*. The instructions then outlined the three revision choices and indicated how any one of them could be a plausible new set of beliefs, given the update information. The instructions emphasized that there were no right "logical" answers to problems, and that the study's goal was to discover what "common sense" choices people made across a variety of simple scenarios.

Results

The dependent measure was the frequency of particular revision choices on the problem set as a function of the independent variables. An analysis of variance assumes that data are normally distributed, an assumption that does not hold for frequency data. Therefore, a chi square statistic was used to assess simple main effects of the independent variables and possible interaction effects. This approach essentially evaluates whether frequencies of subjects' responses deviate from the expected frequencies across the three possible revision choices, given particular combinations of values on the independent variables.

Causal Conditionals. Chi-square statistics indicated that the number of disabling factors and the form of the initial belief set—modus ponens v. modus tollens—each affected subjects' preferred revision, after considering the contradictory information. There was no influence of alternative causes, and neither the disabler factor nor the alternative causes factor differentially affected revision choices on modus ponens and modus tollens belief sets.

The nature of the disabling-factors effect can be seen in Table 2, which presents the percentages for each revision option (disbelieve the causal conditional, disbelieve the causal data statement, or decide both are uncertain), as a function of few v. many disabling factors. When there were many disabling factors for a causal relationship, there was a slight preference to disbelieve the causal conditional over disbelieving the causal data statement (41% v. 37%).

However, when there were few disabling factors, retaining the causal conditional and disbelieving the causal data was the clear preference (51%) over disbelieving the causal conditional (31%) (χ^2=13.57, df=2, p =.001).

Table 2 also shows the percentages for each revision choice for modus ponens and modus ponens belief sets. For modus ponens belief sets, the preferred revision was to believe $p->q$ and disbelieve p (48%), rather than disbelieve $p->q$ (30%); there was little difference between these two options for modus tollens problems (39% for disbelieving $p->q$ and retaining $\sim q$, and 42% for disbelieving $\sim q$ and retaining $p->q$). This pattern is consistent with previous studies (Elio & Pelletier, 1994; Elio & Pelletier, *in press*), which found that subjects prefer to disbelieve the conditional and continue to believe the data sentence on modus ponens belief sets; on modus tollens problems, there was more "uncertain" labels chosen for the initially-believed sentences. Elsewhere, Elio & Pelletier (*in press*) have discussed various model and proof-theoretic accounts of these differences in revision choices for modus ponens and modus tollens belief sets. While their analysis will not be discussed in detail here, one account of this difference is that subjects may not accord full belief to the initial modus tollens belief sets in the first place. This would be consistent with the generally-accepted difficulty people have with generating or validating modus-tollens inferences.

Definition Conditionals. Table 3 presents the frequencies of revision choice for definitions. Chi-square statistics indicated that both problem form and familiarity affected revision choice; these factors did not interact. The effect of familiarity was in the direction expected: familiar definition conditionals were retained more often after contradiction than were unfamiliar ones (47% v. 39%; χ^2= 6.34, df=2, p=.041) but the effect was not as strong as might be expected, since a 37% of the revisions to familiar-definition belief sets involved disbelieving the familiar definition. When the definition was unfamiliar, there was not the expected preference to disbelieve the definition, but instead more choices for labeling both the unfamiliar definition and the definition data statement uncertain.

The effect of problem form follows what was reported above for causal conditionals: a clear preference to retain belief in $p->q$ and disbelieve p following contradiction on modus-ponens belief sets, and with a more equivocal pattern of revision choices for modus-tollens problems (χ^2= 7.51, df=2, p=.02).

Table 3 also presents the frequencies of each belief revision choice for contradicted promise conditionals. Unlike their revision of causal and definition belief sets, subjects clearly preferred (over 50% of the time) to disbelieve a promise conditional as a way of resolving contradiction, regardless of whether the belief sets involved modus ponens and modus tollens inferences. This pattern of revision choices on contradicted promises is quite distinct from the other patterns shown in Table 3, and is significantly different from a chance distribution of revision choices (χ^2= 83.44, df=2, p < .0001)

Experiment 2

Three results concerning the impact of knowledge type on resolving contradictory information emerged from Experiment 1: (a) when there were many disabling conditions associated with a causal conditional, there was a preference to disbelieve the conditional to reconcile contradictory information; with few disabling factors, the preference was to retain the causal conditional and disbelieve the initial data-statement; (b) contradictions involving promise belief sets are resolved quite differently than either definitional or causal belief sets: for promises, there was a very strong preference to disbelieve the promise conditional as a way to reconcile contradictory information regardless of whether the belief set was in modus ponens or modus tollens form; (3) there was a greater tendency to retain belief in familiar definitions than in unfamiliar definitions in the face of contradictory information.

Experiment 1's selection task forces subjects to chose among pre-generated revision choices, each of which would be a logically consistent way to reconcile the new information with the initial beliefs. A benefit of this approach is that it forces subjects to make a definite choice about whether the conditional or the data statement is to be abandoned. There are, however, three

drawbacks to this paradigm. The first is that the "uncertain" option gives subjects an "out" from the clear dichotomous choice of choosing between belief in a conditional v. a non-conditional statement. The second drawback is related to the task's benefit: by asking subjects to make absolute choices, the selection task cannot measure changes in *levels* of belief. Changing the task to allow subjects to specify a degree of belief in each of the initial beliefs, after contradictory information arrives, may offer a more sensitive measure of whether different types of *if-then* knowledge are differentially entrenched. A third drawback to the selection task is that the frequency data are not amenable to more familiar methods of analysis, such as analysis of variance. A degree-of-belief task alleviates this drawback as well. Experiment 2 used the same problems and design described for Experiment 1. The difference was that subjects specified what they thought a new degree of belief should be for each of the two initial beliefs—the conditional and the data statement— in light of the contradiction that the update information introduced.

Method

Stimuli and Design. Experiment 2 used the same problems and design as Experiment 1. The initial belief sets and the update sentences were the same. Instead of the three revision options, each problem presented the following questions:

> Assuming the new information is true, what do you think the degree of belief should be for.....
>
> *<statement-1>*
>
> Disbelieve Believe
>
> A———B———C———D———E———F———G
>
> Uncertain
>
> and for....
>
> *<statement-2>*
>
> Disbelieve Believe
>
> A———B———C———D———E———F———G
>
> Uncertain

where *statement-1* and *statement-2* were the conditional statement and the data statement from the initial belief set. The order in which these two statements were presented for rating was randomly determined across subjects and problems.

Subjects and Procedure. A total of 68 subjects were recruited from the Department of Psychology's subject pool, and received credit for experiment participation as required by an introductory psychology course. Thirty-four subjects received the 22-problem set that contained the causal conditionals and promise conditionals (as described for Experiment 1), and thirty-four received the 18-problem set that contained the promise conditionals and the definition conditionals (again as described for Experiment 1). The procedure was the same as described for Experiment 1.

Results

Causal Conditionals. Table 4 presents the degree-of-belief ratings for problems for causal belief-revision problems. Recall that for each problem, a subject gave a pair of judgments after considering the contradiction introduced by the update: a new degree of belief in the conditional and a new degree of belief in the causal data statement. The larger the absolute difference is between these two judgments, the greater the degree to which subjects are calling one of the statements into question over the other, given the contradiction.

A 2 (problem form:MP v. MT) X 2 (alternative causes) X 2 (alternative disablers) X 2 (item rated: conditional statement, data statement) analysis of variance, with repeated measures on all three factors revealed a main effect for problem-form, and interactions between alternative-disablers and item-rated ($F(1,33) = 14.63$, $p = .001$), alternative-causes and item-rated ($F(1,33) = 10.30$, $p = .003$), and inference-rule and item-rated ($F(1,33) = 7.42$, $p = .01$. Although the patterns of responses appeared different for MP and MT problems as a function of causal-knowledge types, the three-way interaction did not approach significance.

The effect of problem form is consistent with results from previous studies on these types of problems: when considering the contradiction that the update information presented, subjects

gave somewhat higher belief ratings to causal conditionals than to causal data-statements on the modus ponens belief sets, and higher ratings to causal data-statements than to causal conditionals in the modus tollens belief sets.

The key effect is the role of disabling factors. Essentially, few-disabler conditionals were "more entrenched" given contradictory evidence than the initially-believed data statements, and many-disabler conditionals were "less entrenched" than the initially-believed data statements. The degree-of-belief ratings (after contradiction) for many-disabler conditionals and causal data-statements were 3.9 and 4.8, respectively. For few-disabler conditionals and the associated causal data-statements, the mean ratings were 4.6 and 3.9, respectively.

This effect of disabling factors was most pronounced for the modus-ponens problems in the few alternative causes/few disabling conditions case, where subjects' rated the conditional at 5.1 and the data statement, 3.3. This few-alternatives/few-disablers case might be viewed as the "tightest" possible coupling of the antecedent and consequent, one that is almost definitional in nature. (Indeed, subjects' revision decisions on familiar definitions, presented, shows a similar pattern.)

The number of alternative causes had an effect on post-contradiction belief-ratings as well, although the magnitude of the differences is rather small. With few alternative causes of the consequent, the initially-believed conditional received a lower rating than the initially-believed data statement (4.4 v. 4.2, respectively); with many alternative causes of the consequent, the data statement received a higher belief rating than the initial conditional 4.6 v. 4.0 respectively), after the contradiction was considered.

Definition Conditionals. Table 5 gives the belief ratings for conditional and data statements on the definition problems. A 2 (problem form) X 2 (familiarity) X item-rated (conditional, data statement) analysis of variance indicated main effects for definition type and item, and interactions between problem form and item rated ($F(1,33)=13.15$, $p=.001$) and between definition type and item rated ($F(1,33)=5.61$, $p = .024$). When belief sets using familiar definitions were subsequently contradicted, subjects gave a higher belief rating to the definition-conditional (5.0) than to the

definition data-statement (3.9). When belief sets using unfamiliar definitions were contradicted, subjects rated both the conditional and data statements around the "uncertain" level on the belief scale. As before, there was a higher-belief rating given to the conditional (5.0) than to the data statement (3.5) for modus ponens belief sets that were contradicted; the ratings for the definition vs. the definition data statement were essentially the same for contradicted modus tollens belief sets.

Promises. Table 5 also gives the belief-ratings assigned to promise conditionals and the associated data statements, after contradiction. The rating task data are consistent with the selection task data: belief in the promise data was considerably higher than in the promise conditional, following contradiction. This difference was greater for modus tollens than for modus ponens belief sets ($F(1,67)=4.85$, $p = .031$).

Discussion

Neither Experiment 1's selection task nor Experiment 2's belief-rating task directly measured the change in a subject's acceptance of particular sentences that generated a subsequently-contradicted inference. However, I think the data do indicate which of the two statements that comprised the initial belief sets— the conditional and the so-called data statement—was more "suspect" when an inference following from them was contradicted. The data from Experiments 1 and 2 clearly indicate that it is the type of knowledge expressed in a conditional form, rather than the conditional form itself, that influences what we might regard as plausible belief revision decisions.

First, it appears that there is a preference to suspect a promise conditional subjects prefer to abandon promise conditionals and retain a non-conditional statement as a way resolving contradiction; their treatment of promise conditionals and promise data is quite distinct from their revision patterns for belief sets involving causal or definitional information. Second, familiar definitions expressed in conditional form are more often retained (or, alternatively, have a higher degree of belief) than unfamiliar definitional conditionals, in light of contradictory information. Third, causal relationships expressed in condition form display different revision patterns, as a

function of Cummins et al.'s many v. few disabler distinction and, to some extent, the many v. few alternative causes distinction. In general, subjects were more likely to revise their belief set by abandoning (or have a lowered degree of belief in) an initially-believed "data" statement, when the initially-believed causal conditional had few disabling factors, than when it had many disabling factors.

Clearly, the conditional form itself is too crude to serve even as a heuristic upon which to hang principles of epistemic entrenchment. And so too is knowledge-type distinction itself, given the influence of the few-disablers distinction on how subjects revised belief sets involving causal conditionals. These results offer some insight into what the "extra-logical" preferences guiding belief revision might be, and suggest that they can be formalized in a somewhat domain independent fashion if we view a belief revision decision as based in part on considering alternative explanations for a particular situation. When there are few factors that can disable the causal relation $p -> q$, then the contradiction of an inference casts suspicion on the antecedent (i.e., what's termed the data statement here) actually holding. Put another way, this interpretation suggests a view of epistemic entrenchment modeled as the *result* of assessing the likelihood of the alternative possible worlds that correspond to different ways of accounting for a contradiction. When there are few-disablers for $p -> q$, then there are few(er) possible worlds in which a disabler could be in effect. It is less likely, then, that current world (to be modeled by a belief state) is one of those worlds in which a disabler is also true. In this case, it may be a more plausible belief-revision decision to retain belief in the conditional statement and question the validity of the data statement. Put another way, the heuristic that "data has priority" may not always lead to a plausible belief revision decision, given a reasoner's background knowledge.

The few-alternative-causes/few-disablers case can be interpreted as a much tighter, almost definitional relation between p and q: there are few other explanations for p when q is true, and few other explanations for why q would fail to hold when p is true. The pattern of data for few-causes/few disabler causals looks remarkably similar to the pattern of data found for contradicted belief-sets involving familiar definitions. On modus ponens problems, the post-contradiction

belief-rating for few-disablers/few-causes conditionals was 5.1 versus 3.3 for the causal data statement; the post-contradiction belief rating was 5.2 for the familiar-definition conditional versus 3.6 for the familiar-definition data statement. Cummins (1995) also noted that a few-disablers causal relation is quasi-definitional in nature.

It would be parsimonious to extend the notion of disablers (and hence the possible-worlds account) to the treatment of contradicted promises. For example, our understanding of a promise "*If you do x, then you will get y*" might include an understanding that many factors outside the control of the promiser might derail the promise. Thus, promises might be subsumed under a "many-disabling factors" category of relations. There is some support for this idea from the present data: the degree-of-belief ratings for promise-conditionals and promise-categorical statements are most similar to those given to causal-conditionals in the many-disabler cases, at least for the MP case.

Experiment 3: Knowledge Types and Unfamiliar Domains

A previous study (Elio & Pelletier, 1994) used belief revision problems with science-fiction cover stories and contrasted the revision choices on these problems with those made on problems that used nonsense syllables and symbols. Subjects were more likely to abandon the conditional in the science fiction problems than in the symbolic problems. There was no background knowledge subjects could bring to that task—the science fiction problems mentioned strange alien creatures, locations, and behaviors. The interpretation offered for those results conjectured that conditionals were viewed as hypothesized regularities or theories, and subsequent contradiction presented by the update information showed that the hypothesized regularity was flawed.

The question that follows naturally from both those findings and the knowledge-type results of Experiments 1 and 2 in the present study is whether different types of knowledge expressed in *if-then* form are differentially revised, even when the domain of discourse is unfamiliar to a reasoner.

To explore this hypothesis, Experiment 3 used concepts from a domain that is less obviously fictitious than the science-fiction stories, but presumably is nonetheless unfamiliar: anthropological facts and conditionals about obscure cultures and societies. The concepts were culled (and modified where necessary) from actual anthropological journals. As before, an independent group of subjects categorized each statement according to whether they thought it expressed a causal relation, a prediction, a definition, or a promise. If people's plausible belief revision decisions are guided at a meta-level consideration of knowledge types, then there should be different patterns of revision choices made for these knowledge types, even when the domain is unfamiliar.

Method

Stimuli. A set of fifteen conditional statements using anthropological concepts was constructed. An example anthropological conditional (hereafter "anthro conditional") is *If there is a death in a Meorian tribe, then the tribe relocates its camp*. The same subjects who categorized the conditionals according to knowledge type for the previous studies also categorized the anthro conditionals. As a prelude to categorizing the anthro conditionals, the subjects were told that the statements concerned topics taken from anthropology studies of cultures in various foreign places. The instructions acknowledged that they were likely to have no first-hand knowledge about the topics, but that their task was still to categorize the kind of knowledge that each statement was expressing. The instructions indicated that the classification categories would include only definition, promise, cause-and-effect, and prediction. The familiar vs. unfamiliar distinction for definitions was dropped, on the presumption that all the anthropology statements would be unfamiliar.

Of the fifteen conditionals, none were clearly identified as promises and four had no clear majority votes for any category. Appendix 2 gives the rating data for the eleven remaining statements, grouped into three sets: four causal statements, four statements that were classed nearly equally often as predictions and definitions, and three statements that were clearly classed as

predictions. It should not be surprising that the subjects giving these classification choices were less consistent than they were in giving categories to the stimuli used in Experiments 1 and 2, since the anthropology content was unfamiliar. Still, these ratings allowed a preliminary study of how three general classes of knowledge types might impact belief revision decisions.

Design. Problem sets based on these eleven anthro conditionals were created in the following way. Each problem set used (a) two of the four causal statements to create problems in modus ponens form and the remaining two for problems in modus tollens form; (b) two of the four predict/define statements for modus ponens problems and the remaining two for modus tollens problems; and (c) the three prediction statements to create either modus ponens or modus tollens problems. This design permits two types of analysis. The first analyzes belief ratings for modus ponens problems, where knowledge type (causal v. predictive/definition v. predictive) is a repeated measure. The second analysis focuses on modus tollens problems, again with knowledge type as a repeated measure.

Each particular conditional appeared equally often in each of the two possible forms (modus ponens vs. modus tollens) across all the problem sets designed in this manner. This study used the degree-of-belief rating task paradigm described for Experiment 2.

Subjects and Procedure. Thirty-nine subjects participated in the study. The procedure was the same as that used in Experiments 1 and 2; the instructions were the same as those given for the rating task in Experiment 2, modified slightly to inform subjects that the topic of the problems concerned statements from anthropological studies of unfamiliar cultures.

Results

Table 7 gives the mean belief ratings for the anthro conditional and data statements as a function of knowledge type. First consider the data on modus-ponens belief sets in columns 1 and 2. A 2 (form) X 2 (causal vs. prediction/definition) X 2 (item rated) ANOVA with repeated measures on all factors revealed main effects for knowledge type and item rated, a form by item-rated interaction, and a three way interaction of all factors ($F(1,38)=6.79$, $p=.013$). Consistent

with previous findings using the science-fiction topic, there was a tendency for subjects to retain greater belief in the data statement, and to have lowered belief in the conditional statement, when confronted with contradictory information.

The new result is that the magnitude of the differences depended on the knowledge type expressed in the conditional statement. The difference between the conditional belief rating and the data statement belief rating was -.86 for the causal statements and -2.15 for the prediction/definitional statements. The pattern for modus tollens problems were somewhat different: on causal problems, the difference between belief ratings given to conditionals and data statements was not as great on causal problems (-.58) and was negligible on prediction/definition problems (.22). Indeed, on prediction/definition problems, the mean belief ratings given to both statements (4.22 and 4.00) corresponds to claiming them both to be equally uncertain, given the contradictory update information.

These data support the assertion that a distinction between causal v. predictive knowledge, even in unfamiliar domains, leads to belief revision decisions. The small difference in the belief ratings given to conditional v. data statements in causal belief sets (-0.8) could plausibly be interpreted as indicating an uncertainty about which statement was more suspect. The larger difference in belief ratings for the predictive statements (-2.1) is consistent with the idea that conditionals expressing a loose regularity or association are less deserving of entrenchment than are data statements, in the face of contradictory information. This is consistent with previous findings using science-fiction stimuli (Elio & Pelletier, 1994) and the explanation offered for those results seems applicable here as well. Namely, the unfamiliarity of the domain and the task itself invite the reasoner to regard the conditionals as hypothesized regularities in a poorly-understood domain. When new information conflicts with the hypothesis, the hypothesis rather than the data is suspected. This makes sense at least in unfamiliar domains, where the relative merit of the hypothesis or its track record in accounting for is unknown to the reasoner.

General Discussion

The results from all three studies supported the hypothesis that the type of knowledge expressed in a conditional form influences the degree to which the conditional is 'entrenched' in the face of contradictory information. Further, Experiment 3's results showed that (a) subjects could impose a knowledge-type classification on conditional statements about an unfamiliar domain and (b) the knowledge-type classification lead to different patterns of belief revision decisions.

Taken together, these results support epistemic entrenchment principles based on (a) domain-specific knowledge about possible alternative causes or disabling conditions, when making belief revision decisions about a causal scenario; and (b) domain-independent knowledge about the *type* of knowledge being expressed in conditional form and its inherent believability in the face of contradiction (e.g., promises v. causals). Knowing whether there are many or few disabling factors associated with a causal relationship is an important piece of meta-knowledge that can determine whether it is more plausible to disbelieve the conditional statement or more plausible to disbelieve the non-conditional statement. Elsewhere (Elio & Pelletier, *in press*), the notion that $p \rightarrow q$ expresses an hypothesized "regularity" about the world and *p* expresses "data" was part of an account for people's reluctance to entrench conditionals on tasks like this. This work clarifies that the type of regularity or expectation that the conditional statement expresses is crucial.

The results call into question the notion that syntactic form can even serve as a plausible heuristic for entrenchment: people were more prepared to abandon conditionals that expressed promises than they were to disbelieve conditionals expressing causality. But this knowledge-type distinction itself is too gross, as demonstrated by influence of disabling factors on whether or not the causal conditional was disbelieved to create a consistent belief-state change. These data indicate that plausible epistemic entrenchment principles are a function of the *type* of knowledge expressed in if-then form, not a function of the syntactic form itself.

It would be useful to expand the many v. few disablers distinction to a larger class of relations than those typically viewed as causal. For example, our understanding of a promise "*If*

you do x, then y will occur" might include an understanding that, generally speaking, many factors outside the control of the promiser can derail the promise. In this fashion, promises might be subsumed under a "many-disabling factors" category of relations. There is some support for this idea from the present data: the pattern of belief-revision data on promises most closely mirrors that of the many-disabler causal conditionals, at least for the degree-of-belief ratings (compare belief ratings for many-disabler belief sets, Table 4, with those for promise belief sets, Table 5).

Besides presenting empirical evidence for certain meta-knowledge factors serving as entrenchment principles, what else does this work say about constructing normative models of belief revision? One implication is that epistemic entrenchment principles are not best conceptualized as a pre-specified priority ordering on sentences in the language, which has the unfortunate feature of being "outside" the belief revision model itself. This has been noted by other researchers, such as Rott (1993), who proposes a view of epistemic entrenchment as a product, rather than a guide, of belief change. The process perspective offered here is that the agent constructs *evidence* that belief *p* is less deserving of entrenchment than belief *q* in the face of contradiction, *because* the agent can generate more possible worlds in which belief *p* easier to disbelieve than belief *q*. Thus, the many-disabling factors for some particular conditional statement may crudely correspond to the many possible worlds in which that relationship does not hold; the more such worlds, the more likely it is that the epistemic state the agent is considering is one of them. Finally, we can view both the causal and promise conditionals as expectations about the current world that turn out not to hold. Thus the agent could be seen as considering "*If p were to happen, then q would have happened, but it didn't*". From this perspective, the relation of this work to deontics and reasoning about counterfactuals (Lewis, 1975; Boutilier, 1994) warrants further study.

The consideration of few v. many disablers seems more consistent with an ordering of belief *sets* rather than *sentences*. This is because an agent must somehow generate those disablers as truths that could co-exist with the conditional under consideration, if they are to have an influence on the plausibility assigned to continued belief in the conditional. It may be equivalent

from a formalization viewpoint to say that the presence of many v. few disablers impacts a probability that a particular belief is corrigible (Gärdenfors, 1988).

The few v. many disablers for an "expectation" (be it based on a causal, predictive, or promise conditional) may be related to the idea of a strong v. weak default rule. The common default-rule example *If x is a bird, then x has the ability to fly*. One could argue that there are actually very few common and realistic disabling conditions for this relation, when *x* is instantiated as a particular bird that is not a known exception. But consider another plausible default rule: *If x is a student, then x carries a full course load*. There seems to be many disabling conditions for this rule (e.g., a student's age, employment demands, program of study, financial considerations, previous academic performance, and so forth).The point is that this is not a set of known exceptions, but rather a set of factors that are known to weaken or disable the predictive quality of the default rule. Again, the perspective here is that epistemic entrenchment is the by-product of generating and assessing alternative accounts of an unexpected contradiction. The formalization of this process may ultimately involve elements of a probabilistic approach, which too has been applied to default reasoning (e.g., Bacchus *et al*., 1992) as well as to belief revision (Dubois & Prade, 1991).

When a conditional has many known disablers, one way to account for the contradicted inference is to appeal to the existence of those disablers and reject the idea that the rule is holding, *at least in a particular case*. Generally speaking, belief revision theories have developed using propositional calculus. But a careful consideration of the conditionals used here leads to the question as to whether or not *any* "propositionally-stated" conditional ought to be understood as a particular instantiation of a universally-quantified relation. For example, the statement *If Larry grasps the glass, then his fingerprints will be on it* seems, introspectively, to be a particular instantiation of the general rule *If x grasps a glass, then x's fingerprints will be on it*. It is difficult to consider any alternative cause or disabling factor here, regardless of how the variable is instantiated. For other statements, it is not so clear. Take for example a universally-quantified conditional about an unfamiliar topic, e.g., *If a Koemetian tribe migrates to a new hunting ground,*

then it re-elects its leaders. The Koemetian tribe south of the river migrated to a new hunting ground. Therefore, it re-elected its leader. When this inference is contradicted, one could elect to disbelieve the second, non-conditional statement, or the conditional as stated, *or* this particular instantiation of the conditional. Logically speaking, the latter option amounts to disbelieving the conditional as universally quantified. But another way to disbelieve a conditional is to demote it to the status of a default rule (hence, it doesn't apply to all cases) or to modify the conditional so that it would never again apply to the particular case involved in the contradiction. This distinction between a propositional or universally-quantified perspective of a belief set is important, if the process of belief revision is grounded on the generation of alternative explanations that accommodate the update information: these explanations must be bound to some particular individuals that instantiate the relations under consideration.

Nebel (1989) notes that "extra-logical pragmatic preferences are necessary to guide the revision process." This work offers some insight into what those extra-logical preferences might be, and suggests that they can be formalized in a somewhat domain-independent fashion. They are also consistent with a view of epistemic entrenchment modeled as a result of assessing the likelihood of the alternative possible worlds corresponding to these candidate belief states. The reported influence of knowledge *types* and disabling factors of conditional relations on revision decisions offers builders of normative models an empirical foundation for characterizing epistemic entrenchment as the result of assessing the plausibility of the alternative belief states that correspond to particular revision or update decisions. This would be useful for belief revision systems that interact with people and need to infer belief-state changes of human agents.

Acknowledgments

This work was supported by NSERC Research Grant A00089. Thanks go to Mike Johnston for data collection and coding, to the Department of Psychology, for it continued cooperation in allowing access to its subject pool for these studies, and to Jeff Pelletier for comments on an earlier version of this manuscript.

References

Alchourrón, C., P. Gärdenfors, D. Makinson (1985). On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. Journal of Symbolic Logic, 50, 510-530.

Bacchus, F., Grove, A., Halpern, J.Y., & Koller, D. (1992). From statistics to belief. In Proceedings of the Tenth National Conference on Artificial Intelligence, (pp. 602-608). Cambridge, MA: MIT Press.

Boutilier, C. Conditional logics of normality: A modal approach, Artificial Intelligence, 68, 87-154, 1994.

Cummins, D. D. (1995). Naive theories and causal deduction. Memory & Cognition, 23, 646-658.

Cummins, D.D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. Memory & Cognition, 19, 274-282.

Doyle, J. Rational belief revision: Preliminary Report. In Fikes, R.E. & Sandewell, E. (Eds.) Proceedings of the second conference on knowledge representation and reasoning. San Mateo, CA, 1991.

Dubois, D. and Prade, H. Epistemic entrenchment and possibility logic, Artificial Intelligence, 50, 223-239.

Elio, R., & Pelletier, F. J. (in press). Belief revision as propositional update. Cognitive Science.

Elio, R., & Pelletier, F. J. (1994). The effect of syntactic form on simple belief revisions and updates. In Proceedings of the 16th Annual Conference of the Cognitive Science Society. (pp. 260-265). Hillsdale, NJ: Lawrence Erlbaum.

Elmasri, R. & Navathe, S. (1994). Fundamentals of database systems, 2nd Edition. Redwood City, CA: Benjamin/Cummins.

Evans, J. St.T., Newstead, S. E., & Byrne, R. M. J. (1993). Human reasoning. Hillsdale, NJ: Lawrence Erlbaum.

Foo, N.Y., & Rao, A.S. (1988). Belief revision is a microworld (Tech. Rep. No. 325). Sydney: University of Sidney, Basser Department of Computer Science.

Gärdenfors, P. (1984). Epistemic importance and minimal changes of belief. Australasian Journal of Philosophy, 62, 137-157.

Gärdenfors, P. (1988). Knowledge in flux: Modeling the dynamics of epistemic states. Cambridge, MA: MIT Press.

Gärdenfors, P. Expansions and revisions of belief revision theory. Proceedings of the workshop on belief revision, Melbourne, pages 1-17, Australian Joint Conference on Artificial Intelligence, 1993.

Gärdenfors, P. (1990). Belief revision and nonmonotonic logic: Two sides of the same coin? In L. Aiello (Ed.) Proceedings of the Ninth European Conference on Artificial Intelligence, Stockholm, pp. 768-773.

Gärdenfors, P., & Makinson, D. (1988). Revisions of knowledge systems using epistemic entrenchment. In Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge, (pp. 83-95). Los Altos, Calif.: Morgan Kaufmann.

Geis, M.C. & Zwicky, A. M. (1971). On invited inference. Linguistic Inquiry, 2, 561-566.

Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Studies in syntax, Vol. 3: Speech Acts. New York: Academic Press.

Lewis, D. Counterfactuals. Basil Blackwell, Oxford, 1975.

Makinson, D., & Gärdenfors, P. (1991) Relations between the logic of theory change and nonmonotonic logic. In A. Fuhrmann & M. Morreau (Eds.) The logic of theory change. Vol. 465 of Lecture Notes in Computer Science. Berlin: Springer-Verlag.

Nebel, B. (1991). Belief revision and default reasoning: Syntax-based approaches. In Proceedings of the Second Conference on Knowledge Representation, (pp. 417-428) San Mateo, Calif.: Morgan Kaufmann.

Rott. H. Coherent choice and epistemic entrenchment (preliminary report). In Proc. of the Workshop on Belief Revision, Melbourne, pp 18-37, Australian Joint Conference on Artificial Intelligence, 1993.

Thagard, P. (1989). Explanatory coherence. Behavioral and Brain Sciences, 12, 435-50.

Touretzky, D., Horty,J., & Thomason, R. (1987). A clash of intuitions: The current state of nonmonotonic multiple inheritance systems. Proceedings IJCAI-87, pp. 476-482.

Willard, L., & Yuan, L. (1990). The revised Gärdenfors postulates and update semantics, In S. Abiteboul & P. Konellakis (Eds.) Proceedings of the International Conference on Database Theory, (pp. 409-421). Volume 470 of Lecture Notes in Computer Science. Berlin: Springer-Verlag.2.

Table 1

Example Belief Revision Problem: Familiar Definition in Modus Ponens Form

---

This is what you initially believe:

If Amanda is a cardiologist, then she specializes in diseases of the heart.

Amanda is a cardiologist.

From this, you believe she specializes in diseases of the heart.

You do further investigation and discover:

Amanda does <u>not</u> specialize in diseases of the heart.

The new information contradicts the initial beliefs. Assume that this new information is true and indicate what you think the best revised theory would claim about each of the beliefs:

A	Believe	If Amanda is a cardiologist, then she specializes in diseases of the heart.
	Disbelieve	Amanda is a cardiologist.
B	Believe	Amanda is a cardiologist.
	Disbelieve	If Amanda is a cardiologist, then she specializes in diseases of the heart.
C	Uncertain	If Amanda is a cardiologist, then she specializes in diseases of the heart.
	Uncertain	Amanda is a cardiologist.

Table 2

Percentage of revision choices for causal belief sets, after contradiction

	Revision Option		
	Believe p —>q Disbelieve data	Believe data Disbelieve p—>q	Both Uncertain
Disablers			
Many Disablers of p—>q	.37	.41	.23
Few Disablers of p—>q	.51	.31	.17
Problem Form			
Modus Ponens	.48	.30	.22
Modus Tollens	.39	.42	.18

Note: The data statement is *p* for modus ponens belief sets, and *~q* for modus tollens belief sets.

Table 3

Percentage of revision choices on definition and promise belief sets, after contradiction

	Revision Option		
	Believe p —>q Disbelieve p	Believe p Disbelieve p—>q	Both Uncertain
Definitions-Familiarity			
Familiar	.46	.37	.17
Unfamiliar	.39	.34	.27
Definitions-Form			
Modus Ponens	.49	.33	.18
Modus Tollens	.36	.39	.25
Promises-Form			
Modus Ponens	.25	.52	.23
Modus Tollens	.19	.56	.25

Table 4

Mean belief revision ratings for sentences in causal belief sets, after contradiction

	After contradiction, the rated degree of belief in ...		
	p—>q	data sentence	difference
Disablers			
Many Disablers of p—>q	3.9	4.8	-0.9
Few Disablers of p—>q	4.6	4.0	0.6
Causes			
Many Causes of q	4.1	4.6	-0.5
Few Causes of q	4.4	4.2	0.2
Problem Form			
Modus Ponens Belief Set	4.4	4.1	0.3
Modus Tollens Belief Set	4.1	4.7	-0.6

Note: A seven point rating scale was used, where 1 meant disbelieve and 7 meant believe. The data statement is *p* for modus ponens belief sets, and *~q* for modus tollens belief sets.

Table 5

Mean belief ratings for sentences in definition and promise belief sets, after contradiction

	After contradiction, the rated degree of belief in ...		
	p —>q	data statement	difference
Definition Belief Sets			
Familiar	5.0	3.9	1.1
Unfamiliar	4.3	4.0	0.3
Modus Ponens	5.0	3.5	1.5
Modus Tollens	4.3	4.3	0.0
Promise Belief Sets			
Modus Ponens Form	3.5	4.5	-1.0
Modus Tollens Form	3.4	5.2	-1.8

Note: The data statement is *p* for modus ponens belief sets, and *~q* for modus tollens belief sets.

Table 6

Mean belief ratings for sentences in anthropological belief sets, after contradiction

Modus Ponens Form

After the contradiction ~*q*, the rated degree of belief in ...

	p —>q	p	Difference
Causal	3.9	4.7	-0.8
Prediction/Definition	3.2	5.3	-2.2
Prediction	3.6	5.7	-2.1

Modus Tollens Form

After the contradiction *p*, the rated degree of belief in ...

	p—>q	~q	Difference
Causal	4.2	4.7	-0.5
Prediction/Definition	4.2	4.0	0.2
Prediction	3.2	5.0	-1.8

Appendix 1

Conditional statements used for belief revision tasks listed according to knowledge type

The numbers in parentheses following each statement indicate the frequency with which 40 subjects classified the statement according to the category to which it was assigned. A second frequency for a different classification is provided for those cases in which there was a clear, closest alternative classification. The causal stimuli are taken from Cummins (1995). See text for a discussion of these items.

Causal (Many Alternative Causes/Many Disabling Conditions)

If fertilizer was put on the plants, then they grew quickly. (19; prediction = 18) [a]

If the brake was depressed, then the car slowed down. (35)

If John studied hard, then he did well on the test. (16; prediction = 24) [a]

If Jenny turned on the air conditioner, then she felt cool. (26; prediction = 14)

Causal (Many Alternative Causes/Few Disabling Conditions)

If Alvin read without his glasses, then he got a headache. (26; prediction =14)

If Mary jumped into the swimming pool, then she got wet. (30; prediction = 8)

If the apples were ripe, then they fell from the tree. (14; prediction = 25)[1]

If water was poured on the campfire, then the campfire went out. (35)

Causal (Few Alternative Causes/Many Disabling Conditions)

If the trigger was pulled, then the gun fired. (34)

If the porch switch was flipped, then the porch light went on. (37)

If the ignition key was turned, then the car started. (33)

If the match was struck, then it lit. (33)

Appendix 1 continued

Causal (Few Alternative Causes/Many Disabling Conditions)

If Joe cut his finger, then it bled. (37)

If Larry grasped the glass with his bare hands, then his fingerprints were on it.(25; prediction = 11)

If the gong was struck, then it sounded. (32; prediction = 7)

If the doorbell was pushed, then it rang. (36)

Promises

If Jeremy mows the lawn, then the Robinsons will give him $15.(34)

If Susan completes the report before the weekend, then her boss will give her a day off next month.(37)

If Shiela serves as the director for the committee, then her Department Head will give her a salary bonu (38)

If Harry finds someone willing to job-share with him, then his boss will support Harry's job sharin proposal. (35)

If Chris signs up an additional 15 people for the art class, then her instructor will give her a discount o supplies. (38)

If Lorna teaches an extra chemistry class for her department this term, then her department will give her course off next year. (38).

Appendix 1 continued.

Unfamiliar Definitions

If a plant is an equisetium, then it spreads by creeping rhizomes (horizontal root stems). (33)

If an animal is an epithelium, then it has a membranous cellular tissue lining the stomach cavity. (29; familiar definition = 8).

If a tree is eugenolic, then it produces a liquid phenol $C_{10}H_{12}O_2$. (30)

If a mineral is a rhyolite, then it has the same crystalline structure as granite. (29; prediction = 5)

If an organism is toxophasmic, then it is a parasite of vertebrate animals. (34)

If a subatomic particle is a thermion, then it is emitted by an incandescent substance. (33)

Familiar Definitions

If a flower is an annual, then it dies after one year of blooming. (24; prediction = 8)

If a person is a cardiologist, then she specializes in diseases of the heart. (35)

If a mineral is a diamond, then it is made of compressed carbon. (28)

If a tree is deciduous, then it loses its leaves every autumn. (25; unfamiliar = 8)

If an animal is a reptile, then it gives birth by laying eggs. (30)

If a plant is capable of photosynthesis, then its leaves contain chlorophyll. (34)

Appendix 2

Anthropological conditionals used in Experiment 3.

The letters and numbers in parentheses indicated the top two categories—causal ("c"), predictive ("p"), promise ("prom") or definitional ("d")— and the frequencies with which the statement was placed in that category.

Causal

1. If there is a death in a Meorian tribe, then the tribe relocates its camp. (c=25; p=10)
2. If a new title is conferred on a high chief, then he re-announces his successors. (c=24; p=8)
3. If it is the rainy season, then the Sh'doma occupy their campsites for at least four weeks. (c=24;p=12)
4. If a Koemetian tribe migrates to a new hunting ground, then it re-elects its leaders. (c=25; p=9)

Prediction/Definition

5. If different families speak the same dialect of S'wara, then they belong to the same sharing camp. (p=15; d=21)
6. If two people have a legal dispute, then the judge has the same social rank as the disagreeing parties. (p=15; d=17)
7. If the walls of a hut are decorated with images of birds, then the hut functions as a hospital for sick tribe members. (p=15; d=20)
8. If a Gheolian tribe leader gives a ceremonial speech, then the speech begins with a list of the village's assets. (p=18; d=14).
9. If a sacred ruin has a ring of boulders around it, then it is inhabited by tribal doctors called Shamen. (p=18; d=15)

Appendix continued next page

Appendix 2 continued

Prediction

10. If two villages are meeting for food exchanges, then the hosting village is represented by a female. (p=22; d=9; prom=9)
11. If a Karn'sha village settles near a cave, then that cave has underground water. (p=30).
12. If a tribe lives near the main village, then that tribe favors cooperation with government representatives (p=26; c=7; d=6)