# How to Disbelieve p—>q: Resolving contradictions

**Renée Elio (ree@cs.ualberta.ca)**
Department of Computing Science
University of Alberta
Edmonton, Alberta, T6G 2H1 Canada

**Abstract**

This study discusses belief-change as the problem of deciding which previously-accepted belief, or premise, to abandon, when an inference from an initial belief set is subsequently contradicted. The data concern how "disbelieving" a previously-accepted conditional premise is realized as a particular modification to that premise. The types of revisions that are made are influenced by the kind of knowledge expressed in the conditional. The results and the broader issues of belief-revision are related to other concerns that have emerged in the literature on propositional inference, such as the reported reluctance of people to make simple valid modus ponens inferences in some circumstances and the general interest in incorporating subjective belief into accounts of deductive inference.

## Introduction

Consider the following occasion of common-sense belief-change. Suppose you believe (a) your colleague is planning on attending a seminar and (b) If he is attending the seminar, he will be leaving the office at 3:45. A conflict would become apparent when you fail to observe him leaving at 3:45. You might inquire "Aren't you going to the seminar?" and learn that he had changed his mind. Once you change the belief about his attendance, your resulting belief set is a consistent and accurate model of this particular situation. Alternatively, you might have questioned the belief *If my colleague is attending the seminar, then he will be leaving at 3:45*. Denying this conditional belief would also have served to eliminate the conflict with the new information *(colleague is not leaving at 3:45)*.

Given that there are these sorts of alternatives, how is it that a reasoner chooses among them, so as to identify a plausible new belief state as a model of a particular situation? That is, given that a reasoner is cognizant of some conflict, what are the principles guiding the belief-change decision, as characterized above, i.e., the principles by which a reasoner decides it is more plausible to abandon belief *i* rather than belief *j*, in order to arrive at a consistent model of some situation?

The real-life example outlined above has the feature that the reasoner could establish on-the-spot which beliefs were faulty, by asking a few questions. However, it is not to hard to imagine a variant in which the reasoner cannot immediately validate a new candidate belief-set (perhaps you observe your colleague through an office window across a courtyard and must nonetheless instead make a transition to a new belief-state without such validation). In any case, the issue still remains: what are the principles underlying how the reasoner chooses among several possible belief-revisions, thereby moving to some other belief-state?

Another way to look at the example above is that there is a need to resolve a contradiction arising between a valid inference derived from a set of accepted premises and some newly-arriving information. A reasoner can always refuse to accept the new information and not be so quick to assume that the premises were faulty. Indeed, there is a large literature that indicates that beliefs persevere even when they ought not. My interest in belief-change starts at the point where any inertia to accept the new information has been overcome, leaving the reasoner with the problem of deciding which accepted premise (belief) to no longer accept, in order to resolve the conflict.

Belief-change, prompted by the recognition of contradiction, has been studied as an element of scientific theory revision or formulation. But in addition to occurring in these grand-scale cases of constructing a theory of some domain, belief-change occurs on a much smaller scale, I believe, as a prevalent part of everyday reasoning by which we formulate and revise situational models of our world.

The work reported here is part of a larger research effort in understanding the principles that underlie how a reasoner, when faced with a contradiction, chooses to abandon one sort of belief (previously-accepted premise) over another (Elio, 1997; Elio & Pelletier, 1997). The data described in the present paper provide some insight into how "denying" a previously-accepted premise is realized as a particular modification to that premise, so that a contradiction-free belief set results. In particular, they provide some indication of a taxonomy of belief-revision operators, which are called upon to resolve contradictions in simple scenarios. The results and broader issues of belief-revision are related to other concerns that have emerged in the literature on propositional inference, namely the unwillingness of subjects to make simple valid modus ponens inferences in some circumstances (e.g., Byrne, 1989) and the interest in extending accounts of deductive inference with subjective inference (George, 1995; Johnson-Laird, 1994; Stevenson & Over, 1995).

## Previous work on belief-change

I have studied belief-change, as characterized above, by using a problem format that can be schematically described as follows: *Suppose you initially believe that p—>q is true, that p is true, and therefore, also that q is true.*

*Suppose you later discover that q is false. Given that the information about q being false is guaranteed to be accurate, indicate which of the following you regard to be the most plausible set of beliefs to have: (a) p—>q is true, p is false, q is false or (b) p—>q is false, p is true, q is false.*

In some experiments, subjects are given the option to claim both the initial premises are "uncertain" in their belief status; in other experiments, subjects are not forced to make this hard-and-fast decision about which premise to believe or disbelieve, and instead rate their degree of belief in the initial premises $p$—>$q$ and $p$, given that new information asserts $\sim q$. In brief, those studies found that (a) on problems where the antecedent and consequent are instantiated by nonsense phrases, subjects showed no preference in which initial premise they disbelieved, in resolving the contradiction; (b) when the problems involved natural-language cover stories about unfamiliar domain, subjects preferred to disbelieve the conditional; (c) when the problems used familiar, real-world content, subjects' preference for disbelieving $p$—>$q$ v. disbelieving $p$ depended upon the kind of knowledge expressed in the conditional (Elio & Pelletier, 1997; Elio, 1997). In the case of the last result, the conditionals expressed either causal relationships, promises, familiar definitions, and unfamiliar definitions. Following a manipulation used by Cummins and her colleagues (Cummins, Lubart, Alksnis, & Rist, 1991; Cummins, 1995), four types of causal conditionals were used, defined by whether there were many or few alternative causes for the consequent, and many or few 'disabling factors'—factors that would lead to the denial of the consequent even in the presence of the antecedent. The key result for belief- revision was the finding that causal conditionals with many associated disabling factors were more likely to be disbelieved, as a way to eliminate contradiction, than conditionals with few disabling factors. In the latter case, more subjects preferred to say, essentially, "It's more plausible to disbelieve the premise $p$" when $\sim q$ arrived as the new information. An account of these findings was based on the idea that the reasoner considers alternative candidate belief sets, each corresponding to assuming that some disabler might be in effect. It was supposed that, when a reasoner can identify many disabling factors that would prevent a conditional's consequent from occurring in the present of a conditional's antecedent, this is tantamount to identifying many belief sets in which the conditional is denied. When there are few such factors, the reasoner may regard it more likely that it is the non-conditional premise that is more worthy of disbelief, to eliminate a contradiction. This account did not consider the plausibility of candidate belief sets. I return to this matter in the Discussion section.

The kind of conditionals used in the belief-change experiments include ones such as: *If the ignition key is turned, then the car starts; If Larry grasped the glass with his bare hands, then his fingerprints will be on it; If Susan completes the report by Friday, her boss will give her a day off next week; If a mineral is a diamond, then it is made of compressed carbon.* When subjects indicate that, to resolve contradiction, a plausible belief set would be one in which one of those conditionals is "disbelieved", what might this mean? In what sense is a conditional "denied"?

To obtain some insight on "disbelief" or "denial" in the context of belief revision, subjects were given the belief-change scenarios used in previous belief revision studies (Figure 1a). The present experiment asked them to provide open-ended information, specifically to indicate what changes they would make to one or the other of the initial beliefs, in order to resolve the contradiction.

Suppose you initially belief the following:
If Joe cut his finger, then it bled
Joe cut his finger.
Therefore, you believe his finger bled.
You do some additional study and discover this is true:
Joe's finger did not bleed.
(a)

Causal-few disabler:
If Mary jumps in the pool, then she gets wet.
Causal-many disabler
If John studies hard, then he does well on the test.
Promise
If Susan completes the report by the weekend, then her boss will give her a day off next week.
Familiar Definition
If a mineral is a diamond, then it is made of compressed carbon.
Unfamiliar Definition
If a plant is an equisetium, then it spreads by creeping horizontal root stems.
(b)

Figure 1: Sample belief-revision problem (a) and illustrative examples of each conditional type (b)

## Experiment

### Stimuli and Design

Twenty-eight belief-revision problems, used in previous belief-revision studies, were adapted for this task. This set consisted of (a) 16 causal conditionals (8 many disabler, 8 few disabler), (b) four promise conditionals, (d) four familiar definition conditionals, and (e) four unfamiliar definitions. The data reported here concerns modus ponens belief-change problems, like the one illustrated in Figure 1, in which the initial belief set presents a modus-ponens inference that is contradicted by the new information. Figure 1 also gives examples of each type of conditional. Cummins et al. (1991) and Elio (1997) identified items as exemplars for these conditional types through norming studies.

### Subjects and Method.

To ensure the task was completed in a reasonable amount of time, the problem set was divided into two sets, one comprised of belief-revision problems using only the causal conditionals and another using only the promises and definitions. Two groups of 21 subjects each received one or the other of the sets. Problems were presented in

random orders to subjects in booklet form. For each problem, subjects were first asked which of the two initial beliefs they believed it was more plausible to disbelieve, given that the new information was accurate. They were then asked to indicate the revisions they would make to the belief they targetted for denial (either $p$ or $p$—>$q$) so that it became consistent with the new information, ~$q$. Subjects were drawn from the University of Alberta Department of Psychology subject pool.

## Results

The major interest is in the kinds of modifications subjects proposed to the initial belief set, so that contradiction created by the new information is eliminated. Some descriptive data on which belief they targeted for revision is useful, however, to show consistency with previous findings. Table 1 presents these data as the frequencies with

Table 1: Percentage of choices disbelieving $p$—>$q$ v. $p$ to resolve contradiction with ~$q$

| | Disbelieve $p$—>$q$ | Disbelieve $p$ |
|---|---|---|
| Causals | | |
| Few Disablers | 57% | 43% |
| Many Disablers | 74% | 26% |
| Promises | 84% | 16% |
| Familiar Defs. | 39% | 61% |
| Unfamiliar Defs | 45% | 55% |

which subjects modified the conditional or the non-conditional premise, as a function of the type of knowledge expressed in the conditional form. Consistent with previous studies, the key factor is the role of disabling factors for causal conditionals. More subjects marked the $p$—>$q$ premise for disbelief when the causal conditional had many disablers than when it had few (74% v. 57%). The trends for promises and definitions are also consistent with previous studies: For contradicted promises, the preference is to disbelieve the conditional premise; for familiar definitions, subjects prefer to disbelieve the non-conditional premise $p$. This preference occurred to a lesser degree in the unfamiliar definitions.

The primary focus of this study was a descriptive characterization of what might be meant when a belief is labelled for "disbelief" or "denial", in the context of identifying a plausible consistent belief set. Seven categories were used to describe the modifications that subjects proposed to the conditional belief, when it was the conditional belief that they indicated ought to be "disbelieved" to resolve contradiction. These categories and the percentage of responses falling into each of them, for each type of conditional belief, are given in Table 2.

The demote-to-default category covered responses of the sort "Usually p—>q " or "p only increases the likelihood of q," or "p —>q, but there are exceptions". Of course, subjects expressed these notions in the context of a particular problem, such as "If the apples are ripe, then they *often* they fall from the tree" (few-disabler causal) or "If the ignition is turned, then the car *should* start" (many-disabler causal). Category 2—missing enabler—covers responses such as "If Susan finished the report by the weekend *and the report was good enough*, then her boss would give her a day off next week" (promise); "If Joe cut his finger and *the cut was deep enough*, then it would bleed" (few-disabler causal). These are cases in which subjects expressed a necessity condition and often (although not always) indicated that the condition was not holding, and hence the inference $q$ could not be made, thus eliminating the contradiction with ~$q$.

The third category in Table 6— *present disabling factor* — covers responses in which subjects expressed the presence of an additional antecedent proposition that makes ~$q$ a consistent inference. Examples include "If the trigger was pulled *and the gun had no bullets*, then the gun would not fire" (causal: many disablers) and "If Chris signs up additional students for the art course *and the students are not of the right type*, then Chris's instructor will not give her a discount on art supplies"(promise). It may be tempting to collapse categories 2 and 3, since necessary and disabling factors are related: an absent necessary condition can be viewed as disabling the relationship. But from an inference viewpoint, they are not quite the same: a present disabler allows one to conclude ~$q$ and an absent necessary condition just blocks $q$ from being inferred. The subsequent belief state is different in these two cases.

Category 4 —"generalize q" — occurred only in a few particular items which seemed to invite this sort of revision. For definitions, almost all occurrences involved the item "If Amanda is a cardiologist, then she specializes

Table 2: Categories of modifications proposed to p—>q , when p—>q was disbelieved to resolve contradiction, and percentages of each type

| | Causals | | Promises | Definitions | |
|---|---|---|---|---|---|
| | Few Disablers | Many Disablers | | Familiar | Unfamiliar |
| 1. demote p→q to default | 30% | 27% | 41% | 63% | 81% |
| 2. p & enabler →q | 39% | 43% | 23% | 0 | 7% |
| 3. p & disabler → ~q | 21% | 23% | 8% | 3% | 0 |
| 4. generalize q | 0 | 0 | 13% | 25% | 0 |
| 5. p→q invalid/incorrect | 1% | 1% | 0 | 6% | 10% |
| 6. exceptional instance | 5% | 6% | 12% | 3% | 2% |
| 7.time, intervening events | 4% | 0% | 3% | 0 | 0 |
| (total N) | (114) | (142) | (75) | (33) | (42) |

in diseases of the heart." Here, the revisions of this type included "*Cardiologists do more things than specialize in diseases of the heart....If she's a cardiologist, then she is concerned with general aspects of the heart.*" The promise conditional that invited this sort of revision was "If Jeremy mows their lawn, the Robinsons will give him $15". The amount of payment was dropped or specified more generally ("...then they will pay him something"). Similarly, the case of Susan completing a report (mentioned above) invited a generalization of the consequent to "... then her boss will give her a day off *sometime*." Despite the infrequent use of this belief-revision option, which I attribute to the nature of a few of the items used here, it is an interesting revision operator because it is a small semantic fix to the sense of the conditional belief: the notion is domain independent, but requires a domain-dependent understanding of how a variable can be generalized.

Category 5—the conditional as a whole is deemed incorrect or generally invalid — subsumes cases such as, for the Susan promise, "Susan's boss lied" or for the Harry promise "It was a misunderstanding that, if Harry found someone to job-share with him, his boss would approve it." This revision seems motivated by some notion of epistemic uncertainty of the information. Category 6—*exceptional instance*—covered cases such as "Joe was not human" (for *If Joe cut his finger, then it bled* ) and "The match was wet" (for *If the match is struck, then it lit).* These responses can be viewed as alternative expressions of demoting $p \rightarrow q$ to a default rule which has exceptions, which in turn is a less-specified account of missing necessary or present disabling conditions (e.g., "If *the match is struck and the match is not wet* ....")

Category 7—intervening events or an appeal to the passage of time—corresponded only to a few cases that are nonetheless interesting for belief revision and for endorsement of a conditional to make an inference in the first place. Examples of responses that initiated this category include *"Larry wiped his fingerprints off the glass"* for "If Larry picks up a glass with his bare hands, then his fingerprints are on it." (causal: few disablers); "The apples will *eventually* fall off the tree" (i.e., just not now) for "If apples are ripe then they fall off the tree";*"Alvin got a headache and then it went away"* for "If Alvin reads the newspaper without his glasses, then he has a headache."

Having gone through the meaning of these categories, what are we now in a position to observe? Firstly, these data offer some insight into a taxonomy of belief-revision operators that people have in their repertoire, and draw upon, to resolve contradiction. There is a clear connection to accounts of why subjects may not draw or fully-believe modus ponens inferences that appeal to notions of entailment (e.g., George, 1995). These data, from belief-revision perspective, underscore role of abduction in both the explaining aspect (why didn't something occur—the revision case) and the predictive aspect (what do I expect to be true—the inference case) of plausible inference. Secondly, these data indicate that, even when the choice is to deny $p \rightarrow q$, the type of revision proposed is related to the type of knowledge expressed in the conditional; or perhaps more precisely, to the type of knowledge the reasoner can bring to bear to plausibly deny the conditional so as to resolve the contradiction. So it is not surprising that the most frequent denial of unfamiliar definitions takes the form of demoting them to a default—the reasoner has no other knowledge (presumably) for generating specific possible necessity or disabling factors that might be at play. It is also a "quick-and-dirty" way to get rid of contradiction, when the reasoner does not find it easy or necessary to identify more specific accounts of a contradiction. Unlike definitions, causal conditionals are "disbelieved" through the appeal to necessary and disabling conditions, as well as to the simple demoting to default status.

The remaining insights about how a previously-accepted belief is "disbelieved" to resolve contradiction come from the considering the other type of revision, namely "disbelieving" the non-conditional premise $p$ in order to obtain a consistent belief set with $p \rightarrow q$ and $\sim q$ (see again Table 1). In most cases, subjects who targeted this premise for disbelieve merely indicated that $\sim p$ was holding, e.g., "Larry did *not* pick up the glass with his bare hands." or "Joe did *not* cut his finger." This kind of flat denial was most prevalent in the causal/few disabler case. In contrast, revision to the premise $p$ had a different flavor for definitions. Here, the disbelief in $p$ was often expressed as doubt about the validity of $p$ as an observation. That is, there were frequent appeals such as "It only *appeared* that the mineral was a diamond" or "It *was not yet firmly established* the plant under investigation was eugenolic" (offered for a contradicted, unfamiliar definition). What seems interesting is that there was no appeal to such misleading appearances with the premise $p$ for the causal belief-revision problems. When subjects opted to disbelieve the non-conditional belief paired with a causal conditional, their disbelief always took the flat-denial form and never "It only *appeared* that Joe cut his finger." Whether there is something more, or less, to this reading of the data is a question for closer study.

## Discussion

This goal of this investigation was to obtain some insight into how "disbelief" might be operationalized, when subjects resolve a contradiction by targeting either a conditional or non-conditional premise from an initial belief set as the most plausible one to deny. The broader significance of the results reported here is as follows. First, these data give some insight into the range of belief-revision operators that people have in their repertoire for resolving contradictions involving simple, everyday knowledge. Secondly, it underscores the abductive component to the resolution of contradiction. Thirdly, these belief-revision issues are, I contend, intimately tied to views on "belief-based" inference and the current interest in probabilistic extensions to models of deductive inference.

The unwillingness of subjects to make modus ponens inferences in certain circumstances has contributed to the

interest in considering probabilistic accounts of human deductive inference. Some researchers appeal to the notion that a conditional may not be fully "endorsed", and so even a simple modus ponens inference based on it may not be forthcoming (George, 1995; Politzer & Braine, 1991). It seems that "endorsement" and "entrenchment" of a conditional are opposite sides of the same coin: the "coin" of default reasoning. When we start imagining cases where an antecedent may not be sufficient for a consequent, then we have entered the realm of default reasoning. Default reasoning is also called non-monotonic reasoning, because the set of accepted beliefs does not grow monotonically. Initially, our background knowledge plus a set of premises may entail conclusion *s*. Upon later learning statement *r* is true, our background knowledge and premises combined with *r* may no longer entail statement *s*. Unlike the operators of standard logic, default reasoning requires a "retraction" operator to remove *s* from the set of accepted beliefs.

The "suppression of valid inferences" reported by Byrne (1989) can be viewed in this manner. Politzer and Braine (1991) argue that "suppression" of the modus ponens inference *p* given the premise set $\{p \rightarrow q,\ r \rightarrow q,\ p\}$ occurs when the *r* proposition primes a consideration of whether *p* is sufficient to conclude *q* (e.g., *"If Mary meets her friend, she will go to the play. If she has enough money, she will go to the play. She meets her friend).* This suppression does not occur when the *r* proposition does not cast suspicion on the sufficiency of *p* for *q*, e.g., when *r* is instantiated as *If she meets her family.* Bryne's interpretation of these results is that, in the former case, the conditional is re-interpreted as $p \ \& \ r \rightarrow q$ and in the latter case, as $p \ \vee \ r \rightarrow q$. Whether subjects make a conjunction of *p* and *r* in the antecedent of the conditional, and that is why they do not make the inference, or whether they no longer think *p* is sufficient to conclude *q*, reduces, in my view, to the same matter: $p \rightarrow q$ is interpreted as a default rule. The antecedent might be a good *predictor* of the consequent, but it is not logically sufficient, for the consequent may be retracted upon learning something else (e.g., *Mary does not have enough money for the movies*).

Once we enter the realm of default or probabilistic reasoning, it is unclear that our propositional models will be appropriate. Making this step moves us toward quantification, for how can even levels of belief or acceptability in a statement be derived without the notion that some variable is instantiated with some probability distribution? In Stevenson and Over's (1995) task, *If John goes fishing, he will have a fish dinner. John is (always, often, sometimes, rarely) lucky when he goes fishing. John goes fishing*, subjects' likelihood of concluding *John will have a fish dinner* was a function of the content of the frequency level mentioned in the second, syntactically unrelated premise. Stevenson and Over propose that the frequency-of-luck manipulation tells subjects something about the proportion of worlds in which antecedent and consequent co-occur vs. those in which they do not. Johnson-Laird (1994) has made similar suggestions, with a notion of extending a mental models framework to have probabilities attached alternative models. We should note three points that are relevant to these views. First, these proposals are consistent with rendering $p \rightarrow q$ as a default rule, one that (at best) is true only most of the time. Second, another way of understanding the effect of the frequency-of-luck manipulation is to say that it is a clue about the frequency with which disabling events and conditions might be present, or necessary events and conditions might be absent, such that *p* is not sufficient to conclude *q*. Both these points are related to a last one, namely that to apply this perspective we must interpret these conditionals as quantifying over some variable, as in *For all events in which John goes fishing, there is another event in which John has a fish dinner.* This is one of those kind of conditionals that seems to have hidden variables in it, and the frequency-of-luck manipulation says something about the probability distributions of those variables. Items used here can be interpreted similarly. Intuitively, it seems our understanding (indeed, our endorsement) of the conditional *If Larry picks up a glass with bare hands, then his fingerprints are on it* stems from our belief in something like *For all events e and for all persons p, if there is a glass-picking-up event (e) done by person (p) with bare hands, then person(p)'s fingerprints will be on the glass.* One might still wish to argue that we reason about the world "propositionally"—by constructing concrete models of atomic sentences with assigned truth values. However, it does not seem we can identify a level of belief or acceptability in many sorts of conditional statements relevant to real-world situations without there being a representation of corresponding universally-quantified forms. And these universally-quantified statements, in turn, may be defeasible. The suppression of logically valid inferences may be best understood as an expression of defeasible reasoning; the different experimental manipulations that obtain this suppression effect may, in turn, be understood as indicating that the statement being reasoned about is defeasible. Different populations of reasoners may not all demonstrate the same patterns of suppression effects on the real-world conditional statements, because their different background knowledge informs them differently as to whether or not such statements are defeasible (see Chan & Chua, 1994).

Since it seems that there are very few simple conditionals that accurately describe the real world without the addition of complex qualifications, aspects of both endorsement and entrenchment may be better viewed as assessing how plausible it is that these qualifications come into play: "If p and *unstated assumptions are holding*, then q, otherwise ~q." This perspective, however, presents at least two problems as we consider a process model for the sort of reasoning required here. One problem is what I'll call the Problem of Infinite Regress: the process that gathers evidence for assessing whether *p* is sufficient for *q* in a particular situation needs a "stopping rule." A second problem is specifying an evaluation function that returns one of several candidate epistemic states as the most plausible one to make a transition to. I have developed these ideas more fully in Elio (1998) and here I will make remarks only about the second issue here. Suppose that I have generated a set of examples and counterexamples that

are relevant deciding whether $p$ is sufficient to conclude $q$ in the current situation (i.e., deny $p$—>$q$), and another set of examples relevant to whether I ought just to change my mind about believing $p$ is holding in the current situation (deny $p$). These imagined situations are only input for some evaluation function that must still assign a metric to each candidate epistemic state, by which one emerges as the "most plausible." One approach is to assign a degree-of-belief to each possible contender, and the formal semantics for deriving degrees-of-belief from probabilities developed by Bacchus et al. (1992) are relevant here. They consider the problem of what prior probability distribution might characterize this set of imagined situations, and they note that as long as there is *some* probability distribution, degrees-of-belief can be generated from statistical information using Bayesian conditioning. Their simplest case—assuming a uniform distribution—corresponds to the intuitions offered by Johnson-Laird (1994) and Stevenson and Over (1995), who suggest that a degree-of-belief for $p \rightarrow q$ can be computed from the proportion of the imagined situations in which $p$ and $q$ are both true v. those in which $p$ and $\neg q$ are true.

However, we should not lose sight of the fact that these imagined situations *include other many predicates* besides $p$ and $q$, namely the predicates whose truth value we conjectured in order to consider whether to change our minds about accepting *$p$ is sufficient for $q$* v. *$p$ is true.* Thus, we have to wonder about holistically assessing the plausibility of *each entire model* that we generate as an example or counterexample situation, because each one was generated using abductive inference. There are always very many such imaginary situations that can serve as examples and counterexamples to a set of formulas. Not all are equally plausible. Ultimately, the reasoner must settle on some particular model of a situation, so that further inferences can be drawn or actions can be justified.

This underscores the need for some kind of evaluation function on the candidate *models* being considered. Here too a few possibilities have been proposed. Both Thagard (1989) and Ng and Mooney (1990) offer coherence metrics that may serve this function. Another implication of these considerations is that epistemic entrenchment and endorsement might not be best conceptualized as a feature of individual sentences, premises, or beliefs. Instead, the consideration of entailment in these realms seems more consistent with a holistic ordering of belief *sets* rather than *sentences*. This is because an agent must generate something like "disablers" or "enablers" as truth conditions that could co-exist with the antecedent under consideration, if they are to have an influence on the plausibility assigned to continued belief in some conclusion. Even if we assert that the set of possible situational models a reasoner generates has a uniform probability distribution, each of those entire models—and not just a single sentence under consideration—must satisfy the reasoner's background knowledge. Thus we may wish to think of both endorsement as well as entrenchment as a holistic property assigned to belief *sets*, and not to individual sentences.

It is useful to remind ourselves that notions like entailment and derivability are monotonic and are properties of logic. Everyday reasoning is likely to be non-monotonic. We need retraction operators or processes for withdrawing or modifying previously accepted beliefs as part of a broad theory of everyday human inference. Exploring belief revision is an important avenue for specifying the scope and content of such a theory.

## Acknowledgments

This work was supported by NSERC Research Grant A0089 to Renée Elio. Thanks to Geneva Lui for serving as a research assistant, to Jeff Pelletier for comments on this work, and to the University of Alberta Department of Psychology, for access to their subject pool. This paper was prepared while the author was on sabbatical at the Center for the Study of Language and Information, Stanford University, and support of that environment is gratefully acknowledged.

## References

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

Byrne, R. (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61-83.

Chan, D. & Chua, F. (1994). Suppression of valid inferences: syntactic views, mental models, and relative salience. Cognition, 53, 217-238.

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.

Elio, R. (1998). Belief revision and plausible inference. Unpublished manuscript under review.

Elio, R. (1997). What to believe when inferences are contradicted. The impact of knowledge type and inference rule. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society,* 211-216. Hillsdale, NJ: Lawrence Erlbaum Associates.

Elio, R. & Pelletier, F.J. (1997). Belief revision as propositional update. *Cognitive Science, 4,* 419-460.

George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86*, 93-111.

Johnson-Laird, P N. (1994). Mental models and probabilistic thinking. *Cognition,* 50, 189-209.

Politzer, G. & Braine, M.D. (1991). Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition, 33* 103-108.

Ng., H. T., & Mooney, R. J. (1990). On the role of coherence in abductive explanation. *Proceedings of the Eighth National Conference on Artificial Intelligence.* (pp. 337-342). Boston, MA: Morgan Kaufmann.

Stevenson, R.J. & Over, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology, 484*, 613-643.

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