

Revising Causal Beliefs

Clare R. Walsh (Clare_Walsh@brown.edu)

Department of Cognitive & Linguistic Sciences, Box 1978,
Providence, RI 02912 USA

Steven A. Sloman (Steven_Sloman@brown.edu)

Department of Cognitive & Linguistic Sciences, Box 1978,
Providence, RI 02912 USA

Abstract

The aim of our studies is to examine how contradictions affect causal beliefs. For example, the discovery that your colleague Mark who has been following diet A suffers from iron deficiency may lead you to revise your belief that diet A provides a sufficient supply of iron. Would you also revise your belief that it causes you to lose weight? Experiment 1 shows that our belief that Mark will lose weight is reduced after encountering the contradiction. Experiment 2 shows that people are also less likely to believe that others will lose weight. The results suggest that people resolve contradictions by generating explanations that revise their causal model.

Belief Revision

As we go through life, we are constantly changing our beliefs. We give up old attitudes and we add new ones. When we discover credible information that contradicts our existing beliefs, then rationally we must revise our beliefs in order to restore consistency. Our aim is to examine how this is done.

For example, imagine you believe the following:

If the drink contains sugar, then it tastes sweet and you believe that in fact:

The drink on the table contains sugar.

But when you taste the drink you discover that it is not sweet leading you to withdraw your earlier inference. Much of our everyday reasoning is non-monotonic. People frequently overturn old conclusions in the light of new evidence. They readily suppress valid deductive inferences when they are presented with new information (Byrne, 1989; Byrne, Espino & Santamaria, 1999).

The discovery that the drink does not taste sweet may also lead us to revise our initial beliefs. Perhaps the drink does not contain sugar. Or it may be that a drink containing sugar doesn't necessarily cause it to taste sweet. For example, perhaps it contains a lot of lemon which suppresses the sweetness. Both possibilities are sufficient to resolve the inconsistency so one question is how to choose from among these possibilities. Logic provides no guidance (Revlín, Cate, & Rouss, 2001). The problem has been studied in philosophy (e.g., Harman, 1986) and in artificial intelligence (e.g. Gärdenfors, 1988). The focus there has been to develop formal principles to guide rational belief change (e.g., Alchourrón, Gärdenfors, & Makinson, 1985). The major principle underlying all existing theories of belief

revision is that we should minimize the amount of information that is lost when we revise our beliefs (e.g., Gärdenfors, 1988; Harman, 1986; James, 1907).

Despite the extensive research in developing formal models of belief revision, evidence on how people revise their beliefs is sparse. The way they do so may be very different from the formal systems developed in artificial intelligence (Legrenzi, Girotto, & Johnson-Laird, 2003). Work on attitude change does suggest that, in the face of new evidence, people will treat all contextually relevant beliefs as modifiable in order to increase consistency (Festinger, 1957; Simon & Holyoak, 2002; Thagard, 1989).

What should we do when we discover information that contradicts a causal belief? To the extent that causal relations describe law-like generalizations, the minimal change may be to retain the causal belief and to give up some of the factual information that led to the contradiction. Alternatively, when causal beliefs describe a theory, then evidence to the contrary is reason to dispense with the theory (Popper, 1959).

Studies show that people frequently focus on conditionals more than categorical facts when they revise their beliefs (e.g., Elio & Pelletier, 1997) although less so when they describe familiar causal than unfamiliar relations (Byrne & Walsh, in press, Walsh & Byrne, 2004) and the tendency to do so will depend on the initial degree of belief in the conditional (Diessaert, Schaeken, De Neys & d'Ydewalle, 2000). Furthermore, when people revise a causal statement, they rarely reject it outright (Byrne & Walsh, 2002). Instead, they may revise their interpretation of the relation (Johnson-Laird & Byrne, 2002). They frequently modify their causal belief by stating that the contradictory example is an exception to the rule or by imagining that possible disabling conditions are present (i.e., factors that prevent a cause from producing its usual effect; Byrne & Walsh, 2002). And people retain a higher degree of belief in a causal conditional when there are few available disabling conditions (Elio, 1997).

We address three questions which examine how a contradiction to a causal belief impacts on our belief system. The questions provide clues to the processes underlying belief change. In Experiment 1, we examine whether resolving a contradiction to a causal belief leads people to revise their judgment about that single belief or whether it has implications for other causal judgments. In attempting

to minimize the changes they make, people may alter a causal belief in a way that leaves other causal beliefs unchanged. Alternatively, if people introduce disabling conditions to modify a causal belief, this may lead to changes which resonate through the belief system. Support for this latter view comes from the finding that when people discover that a cause does not produce an expected effect, they may doubt whether other expected effects will follow (Walsh & Johnson-Laird, 2004).

People may mentally construct a causal model to represent the causal relations between events (e.g., Sloman & Lagnado, 2004). In Experiment 2, we examine whether people use their existing causal model to generate explanations about the situation in which the contradiction occurred or whether their explanations involve a change to the causal model itself. In addition, we examine whether people generate just one or several alternative hypotheses to explain the contradiction.

Experiment 1

We propose that when people encounter a contradiction to a causal belief they generate an explanation for why the cause occurred without the effect (Walsh & Johnson-Laird, 2004). Rather than giving up their belief in the causal relation, they tend to modify it (Byrne & Walsh, 2002) and they may do so by specifying certain conditions that will disable the relation. For example, imagine that despite your belief that exercise causes weight loss, you discover that Anne has exercised but didn't lose weight. Rather than inferring that exercise is not effective you may decide that it is only effective if it is cardiovascular or if you do not at the same time consume more calories. If our hypothesis is correct, then the explanation may influence other causal judgments, for example, whether the exercise increased Anne's fitness level. Our experiment was designed to test this hypothesis.

Method

We constructed six experimental problems. Each problem began by stating a pair of causal statements with a common antecedent, for example:

Jogging regularly causes a person to increase their fitness level.

Jogging regularly causes a person to lose weight.

To measure initial belief in the first statement, we asked the following question:

Tim jogged regularly. What is the probability that his fitness level increased?

We then introduced a contradiction to the second causal statement and we examined whether this influenced their belief in the first statement:

Sam jogged regularly but he did not lose weight. What is the probability that his fitness level increased?

The six problems were of the same form but with different causal materials.

We also constructed two control problems, which did not involve any contradiction and again we measured whether there was any belief change. For example:

Sam jogged regularly and he did lose weight. What is the probability that his fitness level increased?

Participants responded to the questions by giving a number between 0 and 100 (where 0 = definitely not and 100 = definitely).

The participants were 23 undergraduates of Brown University who took part in return for payment or course credit.

Results

Table 1 shows the mean probability ratings before and after the contradiction. The probability of the second consequent was rated as significantly lower after reading the contradiction (mean = 59%) than before reading the contradiction (mean = 77%; $t(df = 22) = 6.07, p < .001$). The pattern occurred for 21 out of 23 participants and the remaining two were tied. The pattern also occurred for each of the six types of semantic content and there was no significant difference in the amount of belief change between the different contents. In the control problems, there was no significant change in the probability of the second consequent when no contradiction was presented ($p > .5$).

Table 1: Mean probability ratings for experimental and control problems in Experiment 1

Problem format:	If A then B If A then C
<i>Experimental Problems</i>	
Given:	Probability of B (0-100)
A	77
A and not C	59
<i>Control Problems</i>	
A	81
A and C	82

The results show that when people receive information that contradicts one causal statement, they will be less confident that other expected consequents will follow from the same cause. One explanation for our finding is that people resolve the contradiction by introducing conditions which would disable the relation. These disabling conditions may also reduce the probability that other consequents will follow from the same cause. An alternative explanation is that people are generally less confident about what they've been told, perhaps because they consider the source less credible, so they reduce their judgments. In our second experiment, we compare these hypotheses.

Experiment 2

The aim of our second experiment was twofold. First, we wanted to know whether people's causal judgments depend on the explanation that was generated for the contradiction

and on that explanation alone. If their probability judgments depend on their explanations, then their judgments should be predictable from their explanation regardless of the contradicted fact. In contrast, if a contradiction just reduces confidence, then their probability judgments should vary with contradiction, and not with the explanation. We tested this by explicitly asking participants to generate an explanation for the contradiction before making a causal judgment, e.g.,

(1) Anne jogged regularly but she didn't lose weight.

Why?

What is the probability that her fitness level increased?

We then asked participants to use this explanation to make another causal judgment. For example, if a participant gave the explanation that Anne's appetite increased, then we asked them the following question:

(2) John jogged regularly and his appetite increased.

What is the probability that his fitness level increased?

If people use their stated explanation (and not the contradicted fact) to make the causal judgment in (1) and they don't consider any other hypotheses, then we expect the probability judgments in (1) and (2) to be equal. Previous research has shown that people frequently neglect to consider alternative hypotheses (e.g., Klayman & Ha, 1987). However, if reasoners do consider other explanations or if their causal judgments are reduced merely because they have less confidence in what they have been told, then we expect these judgments to differ.

The second question that we address in this study is whether people draw on information that already exists in their causal model to generate an explanation for a contradiction or whether resolving a contradiction leads people to revise the causal model itself. We did this by asking participants two further questions. Before reading the contradiction we asked them the following:

(3) Tom jogged regularly.

What is the probability that his fitness level increased?

And after reading the contradiction and generating the explanation, we asked them the following:

(4) Mary jogged regularly and you don't know if her appetite increased.

What is the probability that her fitness level increased?

If a reasoner's causal model already contains information about the relation between appetite and fitness level and they use this information in answering (3), then we expect their responses to questions (3) and (4) to be equal. But if they change their causal model when resolving the contradiction, we expect their answer to these two questions to be different. This study examines these two questions.

Method

We used the same six pairs of causal beliefs as used in the experimental problems in Experiment 1. Each pair was

followed by five questions as presented in Table 2. The questions were presented orally to the participants and the experimenter recorded their responses. The first question again measured participants' initial belief in the probability of the first conditional. Question 2 introduced a contradiction to the second conditional. This time we explicitly asked participants to generate an explanation for why the contradiction might have occurred before asking them to rate the probability that the consequent of the first conditional occurred.

The following three questions measured the probability of the consequent of the first conditional under different conditions, namely, when the explanation given in question 2 was either, unknown, absent, or present. For example, take the problem described in Table 2. If participants answered question 2a by saying that Kevin was taking sleeping pills, then in question 3 we told participants that Frank was worried but it is not known if he is taking sleeping pills and we asked for the probability that he had difficulty concentrating. In question 4, we asked for the same probability judgment given that Helen was not taking sleeping pills. And finally, in question 5, we asked for the probability given that Evelyn was taking sleeping pills.

Table 2: The format of the problems used in Experiment 2

	Worrying causes difficulty in concentrating.
	Worrying causes insomnia.
1.	Mark was worried. What is the probability that he had difficulty concentrating?
2.	a. Kevin was worried but he didn't have insomnia. Why?
	b. What is the probability that he had difficulty concentrating?
3.	Frank was worried but <i>you don't know if the explanation holds</i> . What is the probability that he had difficulty concentrating?
4.	Helen was worried and <i>you know that the explanation does not hold</i> . What is the probability that she had difficulty concentrating?
5.	Evelyn was worried and <i>you know that the explanation does hold</i> . What is the probability that she had difficulty concentrating?

The experiment allows us to examine whether the probability ratings depend on the single explanation that was generated for the contradiction. The experiment also allows us to test whether participants change their causal model before and after the contradiction.

The participants were 20 undergraduates of Brown University who took part in return for payment.

Results

The mean responses for each question are presented in Table 3. The results replicate the finding of Experiment 1. The probability of the second consequent was rated as significantly higher in question 1 before reading the contradiction (mean = 85%) than in question 2 after reading the contradiction (mean = 63%; $t = 5.03, p < .001$).

Table 3: Mean probability ratings for each of the five questions in Experiment 2

Problem format:	If A then B If A then C
Given:	Probability of B (0-100)
1. A	85
2. A and not C	63
3. A and explanation unknown	71
4. A and explanation absent	85
5. A and explanation present	62

Our second finding was that responses to question 2 and question 5 did not differ significantly ($t = 0.60, p > .5$) and this pattern occurred for all six types of problem content. For problems in which the contradiction reduced the judged probability of B (response to question 2 was lower than to question 1), participants gave the same answer to question 2 and 5 for 53% of problems. We would expect greater variety if participants were considering multiple hypotheses. Hence the results are consistent with the view that in many cases, people consider just the one hypothesis given in their explanation and they fail to consider other possibilities. They allow this hypothesis to mediate their later causal judgments without considering the possibility that they are wrong (see also Shaklee & Fischhoff, 1982).

Finally, our results suggest that people resolve contradictions by making a change to their causal model. Ratings for question 1 were significantly higher than for question 3 when the explanation was unknown (mean = 71; $t = 5.37, p < .001$). People do not merely change their causal judgments about the specific case in which the contradiction occurred. They extend these changes to new situations. Responses to question 1 did not differ significantly from responses to question 4 (mean = 85; $t = 0.1, p > .8$). People do not generally resolve contradictions by drawing on events that they have already represented in their causal model.

We also examined the nature of explanations given for the inconsistency. The most common explanation was to introduce a disabling condition which would prevent the cause from producing its usual effect. 74% of responses were of this type. In many cases, the conditions disabled the cause from both consequences. For example, the fact that worry did not lead to insomnia may be explained by the fact

that the person did relaxation exercises. This in turn may reduce the probability that worry will lead to a difficulty in concentrating. The next most common type of response was to suggest that the level or amount of the cause was not sufficient to produce the effect, for example, there was not enough sugar in the drink or the person was not very worried. 18% of responses were of this type. In both cases, the pattern of responses and significance ratings for the probability of B were the same as for the overall ratings.

Discussion

The results of our experiments confirm previous findings that people prefer to modify than to give up a causal belief when they encounter a contradiction. The results also give us insight into how those modifications alter other causal judgments. In Experiment 1 we showed that when we discover a situation where a cause does not produce an expected consequence, we become less certain whether the cause will lead to other expected consequences in this instance. The results of Experiment 2 show that we also become uncertain about whether other expected consequences will follow in a situation involving a different agent.

The findings suggest that discovering a contradiction can lead us to change the information that we use to make causal judgments. The contradiction makes salient or forces us to imagine conditions that may impact on these judgments. Hence the basis for making our judgments has changed.

The results reaffirm the view that monotonic logic systems are inadequate for understanding how people reason. In most cases, when people reason from cause to effect the conclusion is indeterminate. It is rarely possible to state all of the conditions in which the cause will necessarily produce an effect (e.g., Johnson-Laird & Byrne, 2002). One approach used by artificial intelligence researchers is to make the default assumption that all of the necessary conditions are present unless there is information to the contrary (Minsky, 1975). Similarly, people may mentally construct models that do not represent all of the information explicitly (Johnson-Laird & Byrne, 1991) although they may consider additional factors if they come to mind easily (Cummins, Lubart, Alksnis & Rist, 1991).

An alternative way to approach these problems is to assume that judgments are probabilistic. Probabilistic judgment does not require specification of all of the conditions that prevent a cause from having its usual effect; the judgment merely reflects the likelihood that this occurs.

Our results suggest that when people encounter a contradiction they generate explanations. The most common type of explanation is to describe a condition that disables the cause from its effect. These disabling conditions may often be ones that people haven't previously considered and as a result they introduce these new conditions into their causal model. These new conditions may have the effect of disabling the cause from other possible consequences.

Introducing a new disabling condition into a causal model could have two possible results. One is that it could explain

why the effect does not always follow from the cause but the probability judgment may remain unchanged. A second possibility is to decide that this new condition should lower the probability that the effect will follow from the cause. Our results suggest that people tend to use the second approach. When our participants considered new conditions, they used these to reduce their probability judgment further.

Acknowledgments

We thank Phil Johnson-Laird for discussions on this topic. This research is supported by NASA grant NCC2-1217 to Steven Sloman.

References

- Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the Logic of Theory Change. *Journal of Symbolic Logic*, 50, 510-530.
- Byrne, R.M.J. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31, 61-83.
- Byrne, R.M.J., Espino, O. and Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, 40, 347-373.
- Byrne, R. M. J. & Walsh C. R. (2002). Contradictions and Counterfactuals: Generating Belief Revisions in Conditional Inference. In W. Gray & C. Schunn (Eds.) *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates (pp. 160 - 165).
- Byrne, R. M. J. & Walsh C. R. (in press). Resolving Contradictions. To appear in: Johnson-Laird, P.N. & Girotto, V. (Eds.). *The shape of reason: essays in honour of Paolo Legrenzi*.
- Cummins, D.D., Lubart, T., Alksnis, O. & Rist. (1991). Conditional reasoning and causation. *Memory and Cognition*, 19, 274-282.
- Diussaert, K, Schaeken, W., De Neys, W. & d'Ydewalle, G. (2000). Initial belief state as a predictor of belief revision. *Current Psychology of Cognition*, 19, 277-288.
- Elio R. (1997). What to believe when inferences are contradicted. In M. Shafto & P.Langley (Eds). *Proceedings of the 19th Conference of the Cognitive Science Society*. Hillsdale: Erlbaum (pp. 211-216).
- Elio, R. & Pelletier, F.J. (1997). Belief change as propositional update. *Cognitive Science*, 21, 419-460.
- Festinger, L. (1957). A theory of cognitive dissonance. Oxford: Row, Peterson.
- Gärdenfors, P. (1988). *Knowledge in flux*. Cambridge, MA: MIT Press.
- Harman, G. (1986). *Change in view*. Cambridge, MA: MIT Press.
- James, W. (1907). *Pragmatism – A New Name for Some Old Ways of Thinking*. New York: Longmans, Green & Co.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Erlbaum.
- Johnson-Laird, P.N. and Byrne, R.M.J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646-678.
- Klayman, J. & Ha, Y. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, 94, 211-228.
- Legrenzi, P., Girotto V., & Johnson-Laird, P.N. (2003). Models of consistency. *Psychological Science*, 14, 131-137.
- Minsky, M.L. (1975). Frame-system theory. In Schank, R.C., and Nash-Webber, B.L. (Eds.) *Theoretical Issues in Natural Language Processing*. Preprints of a Conference at MIT, Cambridge, MA.
- Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchinson and Co.
- Revlín, R., Cate, C.L., & Rouss, T.S. (2001). Reasoning counterfactually: combining and rendering. *Memory and Cognition*, 29, 1196-1208.
- Shaklee, H. & Fischhoff, B. (1982). Strategies in information search in causal analysis. *Memory & Cognition*, 10, 520-530.
- Simon, D., & Holyoak, K.J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review*, 6, 283-294.
- Sloman, S.A. & Lagnado, D.A. (2004). Do We “do”? *Manuscript in submission*.
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12, 435-502.
- Walsh, C.R. & Byrne, R.M.J. (2004). Belief revision, the inference contradiction effect and counterfactual conditionals. *Manuscript in preparation*.
- Walsh, C.R. & Johnson-Laird, P.N. (2004). Changing your mind. *Manuscript in submission*.