Contradictions and Counterfactuals: Generating Belief Revisions in Conditional Inference

Ruth M.J. Byrne (rmbyrne@tcd.ie)
Psychology Department, University of Dublin,
Trinity College, Dublin, Ireland

Clare R. Walsh (cwalsh@tcd.ie)
Psychology Department, University of Dublin,
Trinity College, Dublin, Ireland

Abstract

Reasoners revise their beliefs in the premises when an inference they have made is contradicted. We describe the results of an experiment that shows that the belief they revise depends on the inference they have made. They revise their belief in a conditional (if A then B) when they make a modus tollens inference (from not-B to not-A) that is subsequently contradicted (A). But when they make a modus ponens inference (from A to B) that is contradicted (not-B) they revise their belief in the categorical assertion (A). The experiment shows that this *inference contradiction effect* occurs not only for factual conditionals but also for counterfactual conditionals. However, reasoners revise their beliefs in factual conditionals more than counterfactuals.

Belief Revision

Suppose you know the following well-established set of knowledge to be true:

If the car was out of petrol then it stalled.
The car was out of petrol.
What, if anything, follows?

You may conclude:

The car stalled.

But suppose additional knowledge comes to light at a later time and you discover the following is true:

The car did not stall.

What do you think you should believe to be true at this point?

New information can contradict previously held beliefs and inferences about the world. The ability to recognise inconsistency is a necessary step in revising beliefs (e.g., Legrenzi, Girotto, & Johnson-Laird, 2002). Once inconsistencies and contradictions are detected, they must be resolved (e.g., Elio & Pelletier, 1997). For example, you may decide to revise your belief in the conditional, and believe instead that the car being out of petrol does not necessarily mean that it stalled (it may be a diesel engine). Or you may revise your belief in the categorical, and believe instead that the car was not entirely out of petrol.

Dealing with contradictions is common not only in scientific discovery but also in everyday 'non-monotonic' inference. Which beliefs do people revise most readily? Conditionals can convey explanations, regularities or hypotheses about the world; categoricals can convey facts, data or observations (Elio, 1997). Revising the conditional or categorical is equally acceptable logically (Revlin, Cate, & Rouss, 2002). Yet most studies show that reasoners revise their belief in the conditional (Dieussaert, Schaeken, De Neys, & d'Ydewalle, 2000; Elio, 1997; Elio & Pelletier, 1997 Politzer & Carles, 2001; Revlin, et al, 2002).

Reasoners may prefer to revise some sorts of beliefs more than others because they accommodate new information by changing little of their existing beliefs (Harman, 1986). Minimal changes can be accomplished by altering beliefs that have the least explanatory power or informational content (Gardenfors, 1988). Categoricals convey more semantic information (they rule out more states of affairs as false) than conditionals (Johnson-Laird & Byrne, 1991); but conditionals and categoricals may differ in how entrenched they are in a belief system.

Intriguingly, some studies suggest that the belief reasoners revise depends on the inference contradicted. Consider a second problem:

If the car was out of petrol then it stalled.
The car did not stall.

You may conclude:

The car was not out of petrol..

But suppose the additional knowledge comes to light:

The car was out of petrol.

What do you think you should believe to be true? Once again you may decide to revise your belief in the conditional; or you may revise your belief in the categorical, and believe instead that the car did stall. The first example illustrates a modus ponens inference and the second illustrates a modus tollens inference, and Table 1 summarizes the structure of the two sorts of problem.

Table 1: Two types of belief revision problem

	Modus ponens	Modus tollens
1. Conditional	If A then B	If A then B
2. Categorical	A	Not-B
3. Conclusion	B	Not-A
4. Contradiction	Not-B	A

The Inference Contradiction Effect

Some studies show that reasoners revise their belief in the conditional more when a modus tollens inference has been contradicted, whereas they revise their belief in the categorical more when a modus ponens has been contradicted (Elio, 1997, experiment 1; Politzer & Carles, 2001). The possibility that the belief reasoners revise depends on the inference that has been contradicted, which we will call the *inference contradiction effect,* is puzzling. The contradiction establishes the same counterexample for both inferences, e.g., the car was out of petrol and it did not stall (A and not-B), yet the counterexample is accommodated differently in each case.

However, it is by no means clear whether an inference contradiction effect exists: some studies show the opposite pattern (Dieussaert, et al, 2000; Revlin, et al, 2002), and others show more revision of the conditional following a modus ponens inference, but equal revision of the categorical following modus ponens and tollens (Elio & Pelletier, 1997). The vagaries may arise because previous studies have asked participants to select from different sorts of - sometimes quite complex and constrained - options, e.g., to indicate denial or doubt about each of the statements, to choose one statement to reject, to rate degrees of belief, or to choose among various compound options such as 'disbelieve conditional and uncertain about categorical'. Our aim in the experiment we report is to establish whether an inference contradiction effect exists, and so we allowed participants to generate their own revisions, unfettered by pre-set selection options.

Previous studies have also presented a conclusion to participants prior to contradicting it, without requiring participants to indicate their evaluation of the inference. A participant who has not made the inference, or who does not agree that the presented inference is valid, may not need to engage in belief revision following the subsequent 'contradiction'. To guard against such a possibility, we allowed participants to generate the inferences they considered to follow from the premises, prior to presenting them with a contradiction, and in this way we ensured that their beliefs were genuinely contradicted.

Our conjecture is that an inference contradiction effect occurs because different cognitive processes are required to alter conditional and categorical beliefs following modus ponens and tollens inferences, and we return to this idea after we consider some new data.

Generation of Belief Revisions

We constructed a set of 8 problems, consisting of three modus ponens inferences, three modus tollens inferences, and two fillers based on quantifiers. The problems were based on a science fiction content about different aliens, their properties, living habits and so on (in other experiments we have examined causal and definitional contents, see Byrne and Walsh, 2002). The content and instructions were adapted from Elio & Pelletier (1997) and Politzer & Carles (2001). Participants were told they would be given 'an initial set of knowledge that was true and well established at the time you began exploring. There were no mistakes at that time'. They were given a set of premises on a page of a booklet (e.g., if A then B, A) and asked to write what, if anything, followed. On the next page, they were given the contradiction (e.g., not B). They were told this information was 'additional knowledge about the planet that has come to light at a later time. This knowledge is also true and well established. The world is still the same but what has happened is that knowledge about it has increased'. Their task was ' to try to reconcile the initial knowledge and the additional knowledge. You are to write down what you now believe to be true of all the knowledge you have at this point'.

The conditionals given to one group of participants were phrased in the indicative mood, e.g., 'If the ancient ruin was inhabited by Pings, then it had a force field surrounding it', and those given to a second group were in the subjunctive mood e.g., 'If the ancient ruin had been inhabited by Pings, then it would have had a force field surrounding it'. The participants were 28 undergraduates of the psychology department at the University of Dublin, Trinity College who participated for course credit

Belief Revision Responses

The sorts of revisions that reasoners spontaneously generated fall into three main categories:

1. Revisions or negations of the conditional. Reasoners indicated that the original interpretation of the conditional needed to be revised, saying, e.g., 'A does not mean must B', 'If A don't have to B', 'not all A's do B'. Or they denied its truth, e.g., 'that B if A is false', 'the original statement that A's B is incorrect'. Revisions of the interpretation were far more common than negations.

2. Revisions or negations of the categorical. Negating the categorical for modus ponens leads to the conclusion

'not-A', for modus tollens it leads to the conclusion 'B' via the double negation 'not not-B'. In other cases, reasoners deduced a new conclusion from the contradiction and the conditional. The contradiction for modus ponens is 'not B', and with the conditional leads (via modus tollens) to 'not-B and so not-A' (which is also the denial of the categorical). The contradiction for modus tollens is 'A', and with the conditional leads (via modus ponens) to ''B' (see also Elio & Pelletier, 1997). This tactic leads to the same conclusion as the previous one, but by a different process.

3. Reasoners affirmed the contradiction and the categorical, either in combination or separately. This tactic led to the conclusion 'A and not-B' (or equivalently, 'not-B and A'). Reasoners find it difficult to make the inference from 'not (if A then B)' to the conclusion 'A and not-B' (Handley, 1996), which supports the suggestion that the conclusion 'A and not-B' is reached by a different process from the conclusions in 1.

Revisions to Factual Conditionals

We report first the results for the participants who received indicative conditionals. They made the modus ponens and tollens inferences frequently (100% and 90% respectively) perhaps unsurprisingly given the content. Most participants generated one revision (81%) and we scored those who generated more than one by their first one (see Byrne & Walsh, 2002 for details).

Table 2: The percentages of revision types for modus ponens and tollens for indicative conditionals

	Modus Ponens	Modus Tollens	Mean
Revise conditional	33	54	44
Revise categorical	41	18	30
Affirm contradiction and/or categorical	8	18	13

Participants revised their belief in the conditional somewhat more than the categorical (44% versus 30%, binomial z = 1.32, 1-tailed p = .093). However, they revised their belief in the conditional more often when modus tollens was contradicted than when modus ponens was contradicted (54% versus 33%, Wilcoxon z = 1.94, p = .05), whereas they revised their belief in the categorical more often when modus ponens was contradicted than when modus tollens was contradicted (41% versus 18%, Wilcoxon z = 2.20, p = .03), as Table 2 shows. Some of their responses consisted of affirmations of the contradiction and the categorical (A and not-B) either together or separately (13%), as Table 2 shows. The remainder of responses consisted largely of explanations of the premises or contradiction, or assertions that none of the premises were true.

The results confirm earlier findings that reasoners revise their belief in the conditional more than the categorical; perhaps more importantly the results also confirm earlier findings of an inference contradiction effect, that is, the belief revised depends on the inference contradicted. In this experiment, the direction of the inference contradiction effect is that reasoners revise the conditional more following modus tollens and the categorical more following modus ponens (for similar results see Elio, 1997; Politzer & Carles, 2001).

The generated revisions show that reasoners do not revise their beliefs solely by rejecting or disbelieving one or both of the premises, nor by doubting or expressing uncertainty in one or both of them. Instead their revisions actively attempt to re-interpret the premises in a way that genuinely reconciles the conflicting information and resolves the contradiction, for example, by calling into question the necessity of the antecedent for the consequent. This sort of revision has not been identified in previous studies which relied on presented selections only. The generated revisions also show that reasoners do not confine themselves solely to revising their categorical or conditional beliefs; a third category of responses emerged which consisted of affirmations of (one or both of) the contradiction and the categorical. It is noteworthy that no participant generated a response which simply affirmed the conditional.

Revisions to Counterfactual Conditionals

The second group of participants received counterfactual conditionals in the subjunctive mood e.g., 'If the Spracks had had high-frequency sound sensor ears then they would have had tentacles'. A counterfactual seems to mean something different from its corresponding factual conditional (Costello & McCarthy, 1999; Ginsberg, 1986; Lewis, 1973; Stalnaker, 1968). It conveys the presupposition that the facts are 'Spracks do not have high-frequency sound sensor ears' and 'Spracks do not have tentacles'. When reasoners are given a surprise memory test for counterfactuals, they mistakenly identify that they were given these facts (Fillenbaum, 1974). They judge that someone uttering a counterfactual means to imply these facts (Thompson & Byrne, in press). They make the modus tollens inference more readily from a counterfactual than from a corresponding factual conditional (Byrne & Tasso, 1999). They make the modus ponens inference just as readily from both sorts of conditional.

Since counterfactual conditionals convey both the facts 'Spracks do not have high-frequency sound sensor

ears', 'Spracks do not have tentacles', as well as the suppositions, 'Spracks have high-frequency sound sensor ears', 'Spracks have tentacles', we considered that reasoners would not revise their beliefs in counterfactual conditionals as often as factual conditionals.

Table 3: The percentages of revision types for modus ponens and tollens for counterfactual conditionals

	Modus Ponens	Modus Tollens	Mean
Revise conditional	16	38	27
Revise categorical	53	13	33
Affirm contradiction / categorical	18	29	24

The results support our conjecture, as Table 3 shows. Once again, participants made the modus ponens and tollens inferences frequently (96% and 87% respectively). A comparison of the means in both tables shows that reasoners revise a factual conditional more than a counterfactual conditional (44% vs 27%, chi^2= 5.29, $p < .05$). For a counterfactual conditional, they often affirmed the contradiction and the categorical (together or separately). The results also show the presence of an inference contradiction effect, and its direction is the same for counterfactual as for factual conditionals.

Cognitive Processes in Belief Revision

Different cognitive processes may be required to alter conditional and categorical beliefs following modus ponens and tollens inferences from a factual conditional. The different effects of the counterexample, A and not-B, on the way reasoners revise their beliefs may arise because of the mental representations they have constructed in the course of making an inference.

Reasoners may understand conditionals by keeping in mind different possibilities (Johnson-Laird & Byrne, 1991). The explicit set of models for the conditional 'if A then B are as follows:

A B
Not-A not-B
Not-A B

where in the diagram 'not' is a propositional-like tag to indicate negation. Reasoners who interpret the conditional as a biconditional will construct the first two models in the set only. Regardless of their interpretation, reasoners may construct an initial set of models that makes some information explicit, but leaves other information implicit, because of the constraints of working memory:

A B
. . .

where the three dots represent an implicit model (Johnson-Laird & Byrne, in press).

Reasoners can make the modus ponens inference from this initial set of models. The categorical, A, is consistent with the explicit model:

A B

which supports the parsimonious conclusion, B. To make the modus tollens inference, they must flesh out the initial set of models to be more explicit. The information, not-B, cannot be incorporated readily into the initial set of models. However, it can be incorporated into the fleshed-out models, it eliminates two of them and it leaves a single model:

not-A not-B

which supports the parsimonious conclusion, not-A.

The process by which the two inferences are made differs. The modus ponens inference is made from the initial set of models, but the modus tollens inference is made only after fleshing out the models to be explicit, eliminating models that are inconsistent. This difference in the process of making the inferences may affect the revision of beliefs.

Consider the contradiction to the modus ponens inference. Reasoners must incorporate the contradiction 'not-B'. They made the inference, 'B', based on the initial set of models, and so they have the option of returning to the initial set to flesh them out to be more explicit. Faced with the contradiction, they may decide they need to think more fully about the possibilities compatible with the conditional. People often return to earlier possibilities to think about what might have been (e.g., Byrne & McEleney, 2000). When they do so they can incorporate the contradiction 'not-B' into one of the fleshed out models:

Not-A not-B

The new model indicates that the belief to revise is the categorical, A.

For the modus tollens inference, reasoners must incorporate the contradiction, A. They made the modus tollens inference by fleshing out the models to be more explicit and eliminating all but the model:

Not-A not-B

They do not have the option of returning to flesh out the initial models again, since they have executed that option in the process of making the inference. In the course of making the inference they have considered several alternatives and eliminated them, 'cashing out' the possibilities to one single remaining possibility. The contradiction 'A' cannot be incorporated into the existing model and so the belief to revise is the conditional, if A then B.

The essential revision principle may be that a contradiction can be incorporated into one of the possibilities compatible with the conditional, if these

possibilities have not been thought about and eliminated already. In the case of modus ponens, the only possibility compatible with the contradiction and the conditional (not-A not-B) is incompatible with the categorical (A) and so the categorical must be revised. In the case of modus tollens, the possibilities have been exhausted already in the course of making the inference, and so the conditional itself must be revised. The inference contradiction effect, that the belief revised depends on the inference made before the contradiction, may arise because different inferences require people to keep in mind different possibilities, which subsequently limits their room for manuever in incorporating contradictory information.

Reasoners can rely on background knowledge to add or eliminate possibilities (Johnson-Laird & Byrne, in press). As a result, when they have relied on knowledge to add or eliminate possibilities, their revisions may *not* be influenced by the inferences they have made (Byrne and Walsh, 2002). For example, given a causal conditional, 'if water was thrown on the campfire then it went out', and 'the fire did not go out', reasoners make the modus tollens inference, 'water was not thrown on the campfire'. But when the inference is contradicted 'water was thrown on the campfire' they can incorporate it by saying, for example, 'not enough water was thrown on the campfire'. Reasoners may even short-cut the process by 'matching' various models (Legrenzi, Girotto, & Johnson-Laird, 2002). The inference contradiction effect may be a feature of certain kinds of content.

Modifying or Abandoning Beliefs?

Previous studies have focused on what beliefs people disbelieve, deny or reject, doubt or are uncertain about. However, a contradiction can call for a revision to the original *interpretation* of the premises. A putative counterexample, A and not-B, does not necessarily mean that a conditional, if A then B, is false. Our participants generated revisions to the interpretation (e.g., 'A's do not necessarily have B's', 'Some other variable affected B, e.g., C') more often than they indicated disbelief, denial, rejection, doubt or uncertainty about the conditional's truth.

Reasoners may reach many different interpretations of a conditional (Johnson-Laird & Byrne, in press). One interpretation of 'if A then B' is that A is sufficient but not necessary for B, and a second is that A is both necessary and sufficient for B. These 'conditional' and 'biconditional' interpretations are inconsistent with the counterexample, A and not-B. But other interpretations are consistent with it, for example, that A is necessary but not sufficient for B. This 'reverse conditional' interpretation may be common in everyday reasoning (Byrne, Espino, & Santamaria, 1999).

The reverse conditional interpretation can occur when an additional requirement is made explicit, e.g., 'if the ruin was inhabited by Pings then it had a force field, if they had to protect their habitations then it had a force field'. The modus ponens inference from, e.g., 'the ruin was inhabited by Pings' is suppressed (Byrne, 1989). Reasoners say there is not enough information, or they incorporate the second requirement e.g., 'The ruin had a force field if the Pings had to protect their habitations' (Byrne, et al, 1999). They select options that refer to the second requirement (e.g. Diuessaert, Schaeken, Schroyens, & d'Ydewalle, 2000) and they judge the requirements to be conjoint (Byrne & Johnson-Laird, 1992). When both requirements are affirmed they readily make the inferences (Byrne, 1989; 1991), and they can be enhanced when reasoners know the additional requirements have been satisfied (Manktelow & Fairley, 2000). The suppression is increased when the additional requirement is emphasized, by phrasing it as a biconditional, 'if and only if the Pings had to protect their habitations…' (Byrne, et al, 1999), by relying on familiar content (Chan & Chua, 1994; see also Bonnefon & Hilton, in press), by qualifying its satisfaction (Stevenson & Over, 1995), or by specifying that the requirement was uttered by an expert rather than a novice (Stevenson & Over, 2001). Conditionals with many additional requirements lead to more suppression (Cummins, Lubart, Alksnis, & Rist. 1991; see also Elio, 1997). The conditions in which a reverse conditional are true can be specified with as much certainty as the truth conditions of a conditional or biconditional (but see Politzer & Braine, 1991; Stevenson & Over, 1995; 2001).

Our results show that revising belief in a conditional can mean modifying the original interpretation, for example changing from a conditional interpretation to a reverse conditional one, rather than abandoning belief in the truth of the conditional. In everyday life, just as in scientific thought, it may be rare to abandon entirely either a theory or a fact, upon discovery of another contradictory fact; instead reasoners may progress by attempting to modify their interpretation to restore consistency.

Conclusions

In everyday reasoning, the conclusions to inferences can be readily withdrawn in the light of subsequent information, that is, they are non-monotonic. An important task in everyday inference is the revision of beliefs in the light of contradictions. The results also show an inference contradiction effect, that is, reasoners revise a categorical belief when a modus ponens inference is contradicted and they revise a conditional belief when a modus tollens inference is contradicted. Our results show that reasoners revise their belief in a factual conditional more than a counterfactual

conditional. Our novel revision generation task allowed us to capture some of the rich re-interpretations that people produce to resolve contradictions through modifying rather than abandoning beliefs.

Acknowledgements

We thank Jean-Francois Bonnefon, Renee Elio, Uri Hasson, Phil Johnson-Laird, Mark Keane, and Guy Politzer for helpful comments on an earlier draft and Michelle Cowley and Michelle Flood for help with the experiment. The research was supported by the Dublin University Arts and Social Sciences Benefactions Fund.

References

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

Byrne, R.M.J. (1991). Can valid inferences be suppressed? *Cognition*, 39, 71-78.

Byrne, R.M.J. & Johnson-Laird, P.N. (1992). The spontaneous use of propositional connectives. *Quarterly Journal of Experimental Psychology,* 45A, 89-110.

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. & Tasso, A. (1999). Deductive reasoning with factual, possible and counterfactual conditionals. *Memory and Cognition,* 27, 726-740.

Byrne, R.M.J. & McEleney, A. (2000) Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1318-1331.

Byrne, R.M.J. & Walsh, C.R. (2002). Belief revision, the inference contradiction effect and counterfactual conditionals. *Manuscript in preparation.*

Bonnefon, J-P, and Hilton, D. (in press). The suppression of modus ponens as a case of pragmatic preconditional reasoning, *Thinking and Reasoning*.

Costello, T., & McCarthy, J. (1999). Useful Counterfactuals. *Electronic Transactions on the Web, 3,* 51-76.

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

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

Diuessaert, 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.

Diuessaert, K, Schaeken, W., Schroyens, W. & d'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking and Reasoning,* 6, 125-160.

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.

Fillenbaum, S. (1974). Information amplified: memory for counterfactual conditionals. *Journal of Experimental Psychology,* 102, 44-49.

Gardenfors, P. (1988). *Knowledge in flux.* Cambridge, MA: MIT Press.

Ginsberg, M. L. (1986). Counterfactuals. *Artificial Intelligence*, 30, 35-79.

Harman, G. (1986). *Change in view.* Cambridge, MA: MIT Press.

Handley, S. (1996). Explicit negation. *Phd thesis, University of Wales.*

Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Erlbaum.

Johnson-Laird, P. N. & Byrne, R. M. J. (in press). Conditionals: a theory of meaning, inference, and pragmatics. *Psychological Review*.

Legrenzi, P. , Girotto V., & Johnson-Laird, P.N. (2002). Models of consistency. *Manuscript.*

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

Manktelow, K.I. and Fairley, N. (2000). Superordinate principles in reasoning with causal and deontic conditionals. *Thinking and Reasoning*, 6, 41-65.

Politzer, G. and Braine, M.D.S. (1991). Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition*, 38, 103-108.

Politzer, G. and Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking and Reasoning*, 7, 217-234.

Revlin, R., Cate, C.L., & Rouss, T.S. (2002). Reasoning counterfactually: combining and rending. *Memory and Cognition.*

Stalnaker, R.C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory*. Oxford: Basil Blackwell.

Stevenson, R.J. and Over, D.E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology,* 48A, 613-643.

Stevenson, R.J. and Over, D.E. (2001). Reasoning from uncertain premises: effects of expertise and conversational context. *Thinking and Reasoning,* 7, 367-390.

Thompson, V. & Byrne, R.M.J. (in press). Making inferences about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* .