Domino effects in causal contradictions

Sangeet Khemlani[1] and Phil Johnson-Laird[2,3]
sangeet.khemlani@nrl.navy.mil, phil@princeton.edu

[1]US Naval Research Laboratory, Washington, DC 20375 USA
[2]Princeton University, Princeton NJ 08540 USA
[3]New York University, New York, NY 10003, USA

Abstract

Inconsistencies call for reasoners to revise the information that yields them – but which information should they revise? A dominant view is that they should revise their beliefs in a minimal way. An alternative is that the primary task is to explain how the inconsistency arose. It implies that individuals are likely to violate minimalism in two ways: they should infer more information than is strictly necessary to establish consistency, and they should reject more information than is strictly necessary to establish consistency. Previous studies corroborate the former effect: reasoners use causal knowledge to build explanations that resolve inconsistencies. Here, we show that the latter effect is true too: as a consequence of their causal knowledge, reasoners reject more information than is strictly necessary to establish consistency. This hypothesis predicts domino effects: when a fact contradicts an element early in an inferred causal chain, reasoners should tend to reject each subsequent event in the chain too. Three studies corroborated domino effects and the causal knowledge hypothesis.

Keywords: inconsistency, bridging inferences, domino effects, causal reasoning, minimalism, mental models.

Introduction

When reasoners are confronted with conflicting information, their thinking has to change. One or more beliefs need to be modified or rejected to restore consistency to the information. Which beliefs should be revised, and how? The predominant view in philosophy is that people should attempt to preserve as much of the old information as they can (James, 1907; Quine, 1992; Harman, 1986; Levi, 1991; Gärdenfors, 1988). For example, Harman argues, "one tries to make a minimal change that eliminates the believed inconsistency, in other words, a minimal change that allows one to stop believing one's other beliefs are inconsistent" (1986, p. 57). Such "minimalist" accounts hold that reasoners should not change more beliefs than is necessary to accommodate the conflicting belief. Gärdenfors offers the following illustrative example:

> "...assume that you believe, firstly, that all persons in the bank after four o'clock yesterday were employees and, secondly, that all employees are honest. If you then obtain compelling evidence that money was stolen in the bank during this time, you must retract one of these beliefs. *It seems irrational, however, to retract both of these beliefs, since this would involve an unnecessary loss of information*" (1982, p. 137, italics added for emphasis).

In other words, minimalist accounts predict that individuals should refrain from rejecting $n + 1$ beliefs when the rejection of n beliefs yield a consistent set of assertions.

Recent evidence contravenes minimalism, however. When individuals encounter an inconsistency, their chief task appears to be to make sense of how it arose. They aim to explain it, and explanations can violate minimal revisions of beliefs (Legrenzi & Johnson-Laird, 2005; Walsh & Johnson-Laird, 2009). Likewise, individuals who have an explanation in mind are faster to revise their beliefs than those who do not (Khemlani & Johnson-Laird, 2013). These results support an alternative to minimalism, i.e., the causal knowledge hypothesis: individuals use their background causal knowledge to resolve inconsistencies, and their reliance on causal knowledge may produce non-minimal changes as a consequence (Johnson-Laird, Girotto, & Legrenzi, 2004).

There are two distinct ways in which reasoners can use causal knowledge to produce non-minimal changes. First, as in the aforementioned studies, they can introduce novel properties, relations, or entities into their beliefs in the form of a causal explanation. For example, Khemlani & Johnson-Laird (2011) report a study in which reasoners who were given the following problem

1. If a person is bitten by a viper then that person dies.
   Someone was bitten by a viper but did not die.
   What follows?

spontaneously drew conclusions such as (2) and (3),

2. The person received an antidote.
3. The person was wearing heavy clothing.

Note that (2) refutes the first premise and (3) refutes the second premise in (1). But they are not mere refutations: (2) introduces an entity (the antidote) and a relation (its reception), and (3) likewise introduces an entity (heavy clothing) and a relation (wearing it). The conclusions are striking given that more minimal alternatives exist, e.g., (4) and (5):

4. A viper's bite is not always deadly.
5. The person was not bitten by a viper.

Reasoners were less likely to make the latter two revisions even though they are direct refutations of the premises; and

when given an opportunity to compare the four options (2-5), they chose explanations over minimal changes. In sum, when reasoners assess inconsistencies they often introduce new information that violates minimal change. The new information has the side effect of eliminating or modifying prior beliefs.

The rejection of prior beliefs could also violate minimalism. Reasoners could reject more information than is strictly necessary to restore consistency. No studies that we know of have tested this hypothesis. But, it follows from the causal knowledge hypothesis, which postulates that individuals use their knowledge to infer causal relations, which are missing in descriptions. Our goal in this paper is to examine whether individuals make such non-minimal changes.

To test the divergent predictions of minimalism and causal knowledge, we carried out three experiments to examine whether a *domino effect* occurs in causal reasoning, i.e., when the facts contradict an earlier event in a causal chain, participants lose confidence in the occurrence of subsequent events. Experiments 1 and 2 gave participants descriptions of a sequence of three events that they should either relate to one another in a causal chain or not. A fact then contradicted one of the events, and their task was to judge each of three assertions in the sequence as true or false. Experiment 3 called instead for participants to estimate the likelihood of those assertions in the descriptions other than the ones that the facts contradicted.

Experiment 1

In previous studies, participants often failed to detect inconsistencies (Otero & Kintsch, 1992), and so the experiment used simple descriptions and an obvious contradiction of one of them. Participants received two sorts of simple description. For the first sort, the "causal" descriptions, they should infer a causal chain, e.g.:

1. David put a book on the shelf.
2. The shelf collapsed.
3. The vase broke.

The description was followed with a fourth assertion that contradicted one of the three preceding assertions, e.g.:

In fact, David did not put a book on the shelf.

For the second sort of description, the "control" descriptions, the participants should not infer a causal chain, e.g.:

1. Robert heard a creak in the hall closet.
2. The faucet dripped.
3. The lawn sprinklers started.

The description was also followed with a fourth assertion that contradicted one of the three preceding assertions, e.g.,

In fact, Robert did not hear a creak in the hall closet.

Neither sort of description asserts any explicit causal relations, but individuals should make 'bridging' inferences that establish a causal chain for the causal descriptions (Clark, 1975).

How should the contradiction affect individuals' belief in the assertions in the descriptions? According to minimalism, there should be no difference between the two sorts of description. Individuals should cease to believe only the assertion that the facts contradict. For the causal example, above, the fact that David did not put a book on the shelf should not cast doubt on the shelf collapsing. It could collapse for other reasons. To cease to believe this assertion is to violate minimalism. In contrast, according to the causal knowledge hypothesis, a domino effect should occur for the causal descriptions. Reasoners should be less likely to believe that the shelf collapsed, or that the vase broke. Domino effects should tend to propagate forwards in time, not backwards. Hence, reasoners should be more likely to abandon events occurring after the contradicted one than those occurring before it. The two hypotheses, however, concur that in the absence of an inferred causal chain, as in the control descriptions, reasoners should cease to believe only those assertions that the facts directly contradict. Experiment 1 was designed to test the contrasting predictions of minimalism and causal knowledge. The participants' task on a given trial was to decide whether one of the three assertions in the description was true or false in the light of the contradiction. Thus, they had to evaluate a contradicted assertion or an uncontradicted assertion, e.g.,

Did Robert hear a creak in the hall closet?	(contradicted)
Did the faucet drip?	(uncontradicted)
Did the lawn sprinklers start?	(uncontradicted)

The causal knowledge hypothesis predicts that for causal descriptions the contradiction of the first assertion should initiate a reevaluation of the whole chain, whereas the contradiction of the second or third assertion should initiate reevaluations of only those events that took place after the assertion. But, no such trend should occur for control descriptions.

Method

Participants. Thirty-two participants were recruited through an online platform hosted through Amazon Mechanical Turk, and they completed the study for monetary compensation. Participation was restricted to United States residents, and repeat participation, both within and across experiments, was impossible.

Design, materials, and procedure. Participants received sets of three assertions in causal or control descriptions (see Appendix for materials). A factual assertion contradicted one of the assertions in the description. On each trial, the participants received a question of the form, "Did X

happen?", where X referred to the first, second, or third event in the initial description, e.g., "Did Robert hear a creak in the hall closet?" The participants acted as their own controls and carried out 18 trials based on whether the description was causal or control (2), whether the fact contradicted the first, second, or third assertion (3), and whether the question referred to an event in the first, second, or third assertion (3). For each causal problem, there was a corresponding control problem containing assertions with the same number of syllables. The contents of the problems were rotated over the eighteen sorts of trial so that each content occurred equally often in each sort of trial in the experiment as a whole. To answer a question, a participant pressed a button marked 'yes' or 'no'.

Results and discussion

Table 1 presents the percentages of "yes" responses to the contradicted and uncontradicted assertions, i.e., the proportion of trials on which participants endorsed the truth of an assertion when that assertion had been contradicted or not. The bold cells across the diagonals show the cases in which participants evaluated assertions that the facts directly contradicted. As the table shows, participants answered sensibly: they accepted uncontradicted assertions more often than contradicted ones (69% vs. 13%, Wilcoxon test, z = 4.63, p < .0001, Cliff's δ = .80). They also accepted assertions more often for control descriptions than for causal assertions (66% vs. 38%, Wilcoxon test, z = 3.49, p = .0005, Cliff's δ = .58). The table also reveals that the acceptance of uncontradicted assertions varies depending on whether or not the assertions occurred in a causal chain. In particular, the results corroborated the trend predicted by the causal knowledge hypothesis. For causal descriptions, participants accepted uncontradicted assertions 25% of the time when the first assertion was contradicted, 55% when the second was contradicted, and 83% of the time when the third was contradicted (Page's trend test, L = 410.5, z = 3.31, p = .0005). The trend shows that the earlier in the causal chain that the contradiction took place, the less likely reasoners were to accept subsequent, uncontradicted assertions. However, for control descriptions, which did not elicit causal chains, the participants accepted uncontradicted assertions uniformly, i.e., 89% when the first assertion was contradicted, 79% when the second assertion was contradicted, and 89% when the third assertion was contradicted (Page's trend test, L = 381.0, z = .38, p = .64). This pattern of results yielded the predicted interaction (Page's trend test, L = 407.5, z = 2.93, p = .002).

Participants' responses revealed that they understood the pragmatics of the task. They rejected assertions that were directly contradicted. For instance, if they received the following problem:

1. Tony pressed the accelerator.
2. The car lurched forward.
3. The fender slammed into a tree.

	Assertion that the facts contradicted		
	Assertn. 1	*Assertn. 2*	*Assertn. 3*
Causal descriptions			
Did event in assertion 1 occur?	*13*	84	85
Did event in assertion 2 occur?	**21**	*4*	77
Did event in assertion 3 occur?	**30**	**26**	*13*
Control descriptions			
Did event in assertion 1 occur?	*20*	86	83
Did event in assertion 2 occur?	80	*19*	85
Did event in assertion 3 occur?	100	73	*15*

Table 1. The percentages of "yes" answers in Experiment 1 to questions about the occurrence of the events in the first, second, or third assertion depending on whether the description was causal or control, and whether the facts contradicted the first, second, or third assertion in the description. Grey italicized cells denote answers about assertions that the facts contradicted, and bold cells highlight domino effects.

and were told that in fact, Tony did not press the accelerator, they registered the factual information and answered accordingly, i.e., they no longer believed that he pressed the accelerator.

The results corroborated the causal knowledge hypothesis and contravened minimalism. Minimalism, by definition, put a premium on the preservation of knowledge. It therefore predicts that reasoners should never reject information unless it is necessary to do so in order to preserve consistency. However, participants rejected assertions that were not directly contradicted, i.e., they rejected assertions even when it was logically possible to preserve them. But, in accordance with the causal knowledge hypothesis, they rejected uncontradicted assertions more often when the facts contradicted an earlier event in the causal chain, i.e., when the first assertion was contradicted.

One shortcoming of the present experiment is that it asked participants to evaluate premises that had been directly contradicted. Individuals learned about an event, then they were told that the event did not occur, and finally they had to evaluate whether or not the event had occurred. They did not balk at the task, but it may have been artificial enough to bias their reasoning. That is, the presence of direct contradictions may have decreased their confidence in the fidelity of the information in a description. Likewise, the experiment forced participants to confront contradictions, and this experience may have increased their discomfort or the subjective difficulty of the task. This difficulty, in turn, could have forced them to engage in analytic reasoning (e.g., Alter, Oppenheimer, Epley, & Eyre, 2007). Experiment 2 eliminated this alternative account by asking participants to reevaluate only those assertions that the final assertion did not directly contradict.

Experiment 2

Experiment 2 used the same materials and task as the previous study. Unlike the previous study, participants evaluated only those events in assertions that the facts did not contradict. For instance, given the following problem:

1. Harry pulled the trigger
2. The gun fired
3. The bullet shattered a window

In fact, Harry did not pull the trigger

participants were not asked to answer a question about the event in the first assertion. They answered questions only about events in assertions that the facts did not contradict, e.g., on separate trials:

Did the gun fire?
Did the bullet shatter a window?

Method

Participants. Twenty participants were recruited through the same online platform used in the previous study, and they completed the study for monetary compensation. None of the participants had received any training in logic.

Design, msyrtisld and procedure. Participants received causal and control descriptions, i.e., 12 sets of three assertions describing events that could be related causally or else were independent of one another (see Appendix). Each description had a following factual assertion that contradicted one of the assertions in the description. The participants' task, as before, was to respond to the question, "Did X happen?", where X referred to an event described in the first, second, or third assertion in the description provided that the assertion was not the one that the facts contradicted. Hence, when the facts contradicted, say, the second assertion in the description, the participants were asked on one trial whether the event in the first assertion occurred, and, on another trial, whether the event in the third assertion occurred. This design yielded six separate conditions. To answer a question, each participant pressed a button marked 'yes' or 'no'. The contents of the descriptions and facts were rotated across the participants and conditions to ensure that each participant received every condition, and that contents occurred equally often over the experiment as a whole.

Results and discussion

Table 2 presents the percentages of trials in Experiment 2 on which participants responded "yes" to the question, "Did X happen?" when factual information contradicted one of the other assertions. The results replicated those in Experiment 1. Participants were asked to evaluate only those assertions that were not contradicted, and so minimalist accounts predict that they should have accepted the causal assertions without reservation. Nevertheless, they accepted such premises only 49% of the time. For control descriptions, they accepted the premises at ceiling (94%, Wilcoxon test, $z = 3.64$, $p = .0003$, Cliff's $\delta = .86$). The results replicated the trend that the causal knowledge hypothesis predicts. For causal descriptions, participants accepted uncontradicted premises 18% when the first assertion was contradicted, 50% when the second assertion was contradicted, and 80% when the third assertion was contradicted (Page's trend test, $L = 267$, $z = 4.26$, $p < .0001$). As in the previous experiment, the trend suggests that contradictions that take place earlier in the causal chain cause participants to reject subsequent effects However, with control descriptions, which did not imply a causal chain, the participants accepted uncontradicted premises uniformly (Page's trend test, $L = 243.0$, $z = .47$, $p = .32$). This pattern of results yielded the predicted interaction (Page's trend test, $L = 265.5$, $z = 4.03$, $p < .0001$).

	Assertion that the facts contradicted		
	Assertn. 1	*Assertn. 2*	*Assertn. 3*
Causal descriptions			
Did event in assertion 1 occur?	--	80	90
Did event in assertion 2 occur?	**15**	--	70
Did event in assertion 3 occur?	**20**	**20**	--
Control descriptions			
Did event in assertion 1 occur?	--	100	95
Did event in assertion 2 occur?	95	--	100
Did event in assertion 3 occur?	95	80	--

Table 2. The percentages of "yes" answers in Experiment 2 to questions about the occurrence of the events in the first, second, or third assertion depending on whether the description was causal or control, and whether the facts contradicted the first, second, or third assertion in the description. The bold cells highlight domino effects.

On each trial, participants in Experiment 2 answered a question about whether or not an event occurred. However, contradictions may affect, not just this binary decision, but also the subjective probability of the various events. Indeed, recent evidence shows that binary choices can obscure more nuanced inferences and judgments (Zhao & Oppenheimer, under review). To extend the results to a task that did not elicit binary choices, Experiment 3 therefore used the same design, but its participants had to estimate the likelihood of events in the descriptions.

Experiment 3

Experiment 3 used the same materials and design as Experiment 2, but the participants estimated the likelihood of events. For instance, participants saw the following problem:

1. Robert heard a creak in the hall closet.
2. The faucet dripped.
3. The lawn sprinklers started.
   In fact, Robert did not hear a creak in the hall closet.

They were asked to evaluate (on a seven-point scale) the likelihood that a particular event took place, e.g., on separate trials:

How likely is it that the faucet dripped?
How likely is it that the lawn sprinklers started?

Participants selected one of seven different degrees of likelihood, ranging from *very likely* to *very unlikely*. These estimates serve as a proxy to reasoners' subjective probabilities, and so they allowed participants to make a more refined response than merely accepting or rejecting the occurrence of events.

Method

Participants. Twenty-four participants were recruited through the same online platform used in the previous studies, and they completed the study for monetary compensation. None of the participants reported that they had received any training in logic.

Design and procedure. The study used the same design and procedure as Experiment 2. Participants registered their response using a Likert scale that ranged from +3 (very likely) through a midpoint of 0 to -3 (very unlikely).

Results and discussion

The study extended the findings to estimates of likelihood, and corroborated the causal knowledge hypothesis. As in the previous study, participants were asked to evaluate only those assertions that the final assertion did not contradict, and so minimalism predicts that they should have judged all assertions to be highly likely. However, their mean estimates of the likelihood of events in causal chains were relatively low (M = -0.64). The difference between the mean likelihoods for the two sorts of description was highly reliable (M = 1.53, Wilcoxon test, z = 3.86, p = .0001, Cliff's δ = .67). Table 3 reports mean estimates of likelihood.

The likelihood estimates once again replicated the trend predicted by the causal model hypothesis. For causal problems, participants' mean estimates likelihood were -1.31, -0.79, and 0.19 when the first, second, and third premises were contradicted respectively (Page's trend test, L = 305.5, z = 2.53, p = .006). The trend suggests that contrad ictions of events earlier in the causal chain cause participants to attribute a lower likelihood to subsequent effects. However, with control descriptions, participants judged the events to have uniformly high likelihoods (Friedman analysis of variance, χ^2 = .23, p = .89). This pattern of results yielded the predicted interaction (Page's trend test, L = 303, z = 2.16, p = .02). Judgments of likelihood reflected the same non-minimal changes observed when individuals evaluated whether or not events occurred. We conclude that the tendency to make non-minimal changes by reevaluating causal premises is robust.

	Assertion that the facts contradicted		
	Assertn. 1	*Assertn. 2*	*Assertn. 3*
Causal descriptions			
How likely is it that the event in assertion 1 occurred?	--	0.33	0.42
How likely is it that the event in assertion 2 occurred?	**-1.77**	--	-0.04
How likely is it that the event in assertion 3 occurred?	**-1.46**	**-1.92**	--
Control descriptions			
How likely is it that the event in assertion 1 occurred?	--	1.5	1.75
How likely is it that the event in assertion 2 occurred?	1.58	--	1.38
How likely is it that the event in assertion 3 occurred?	1.46	1.50	--

Table 3. Mean estimates of likelihood for the first, second, or third assertions in Experiment 3 depending on whether the final assertion in the scenario contradicted the first, second, or third assertion. Estimates of likelihood ranged from +3 (very likely) to -3 (very unlikely). Bold cells highlight domino effects.

General discussion

Reasoners in three experiments made non-minimal changes when they answered questions about descriptions containing an assertion that a fact contradicted. In particular, all three studies replicated a trend predicted by the causal model hypothesis: the revision of one cause in an chain should initiate a *domino effect*, i.e., a tendency to reject, or to reduce the probability of subsequent effects. The studies corroborated this prediction. Experiments 1 and 2 found that people rejected events subsequent to a contradiction in a causal chain, but not in control descriptions.. Experiment 3 generalized this finding to estimates of the likelihoods of events. The results of the three studies revealed a pronounced domino effect, and therefore violated the predictions of the minimalist hypothesis, which states that individuals deal with an inconsistent set of information by making as few changes to their information as possible. Instead, the participants reliably chose to change more than what was absolutely necessary to maintain the consistency of the information they were given.

One limitation of the present studies is that the premises given to participants were artificial because the descriptions were so sparse. We used such problems because they ensured that participants would notice the contradictions, and to show that they would use 'bridging' inferences to

establish causal chains (Clark, 1975). Another motivation for sparse descriptions was that previous studies revealed that reasoners lost track of contradictions in even short descriptions (Otero & Kintsch, 1992). Nevertheless, future studies should investigate whether domino effects occur when participants are given more naturalistic descriptions.

The studies suggest that to resolve a contradiction, individuals are likely to establish a consistent interpretation of the facts of the matter and the original premises by modifying the premises to construct a consistent interpretation with explanatory plausibility. That is, individuals would not have needed to reevaluate subsequent effects unless they sought to construct a consistent causal chain that incorporated the facts in a plausible way.

Acknowledgements

This research was supported by a National Research Council Research Associateship awarded to SK and by NSF Grant No. SES 0844851 to PJL to study deductive and probabilistic reasoning. We are grateful to Ruth Byrne, Monica Bucciarelli, Sam Glucksberg, Adele Goldberg, Geoff Goodwin, and Marco Ragni, for their helpful comments and criticisms.

References

Alter, A., Oppenheimer, D., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytical thought. *Journal of Experimental Psychology: General, 136,* 569-576.

Clark, H. H. (1975). Bridging. In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. New York: Association for Computing Machinery.

Gärdenfors, P. (1982). Epistemic importance and minimal changes of belief. *Australasian Journal of Philosophy, 62,* 136-157.

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

Harman, G. H. (1986). *Change in view: Principles of reasoning.* Bradford, MA: Bradford Books.

James, W. (1907). *Pragmatism – A new name for some old ways of thinking*. New York: Longmans.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111,* 640-661.

Khemlani, S., & Johnson-Laird, P.N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, *64*, 276-88.

Khemlani, S., & Johnson-Laird, P. N. (2013). Cognitive changes from explanations. *Journal of Cognitive Psychology, 25,* 139–146.

Legrenzi, P., & Johnson-Laird, P. N. (2005). The evaluation of diagnostic explanations for inconsistencies. *Psychologica Belgica, 45,* 19-28.

Levi, I. (1991). *The fixation of belief and its undoing*. Cambridge, MA: Cambridge University Press.

Otero, J. & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 3,* 229-235.

Quine, W. V. O. (1992). *Pursuit of truth.* Cambridge, MA: Harvard University Press.

Walsh, C., & Johnson-Laird, P.N. (2009). Changing your mind. *Memory & Cognition, 37*, 624-631.

Zhao, J. & Oppenheimer, D.M. (under review). Beyond binary: Limitations of the binary choice paradigm for studying judgment heuristics. Manuscript under review.

Appendix. The materials used in Experiments 1-3. The first, second, and third premises were provided to participants, and one of them was contradicted.

First premise	Second premise	Third premise
Causal		
David put a book on the shelf	The shelf collapsed	The vase broke
Sammy pushed a button on his cell phone	The phone dialed	The answering machine began
Fred threw a water balloon at George	The balloon hit him	He was drenching wet
Harry pulled the trigger	The gun fired	The bullet shattered a window
Sarah turned on the kitchen light	The bulb burst	Glass fell on the kitchen counter
Tony pressed the accelerator	The car lurched forward	The fender slammed into a tree
Non-causal		
Molly laughed at a man on TV	The dog barked	The door opened
Robert heard a creak in the hall closet	The faucet dripped	The lawn sprinklers started
Frank gave a telephone book to Ron	The clock struck five	Ron was running late
Peter mowed the lawn	The mailman arrived	The dog chased a car
Katie switched off the washing machine	The cat meowed	The children came home from school
Walter lit a candle	The radio was blaring	The delivery man rung the doorbell