Chapter 11

Counterexample retrieval and inhibition during conditional reasoning: Direct evidence from memory probing

Wim De Neys

In one of Bill Waterson's hilarious 'Calvin and Hobbes' cartoons, Calvin is standing in the garden with a big can of water. The plants in the garden are on the verge of wilting and with an evil grin on his face, Calvin proudly exclaims that it is up to him to decide if they get water or not. He keeps on shouting that their very lives are in *his* hands and that *he* controls their fate. Calvin's aspirations of absolute power are short-lived, however. In the next frame, we can see a very disappointed Calvin standing in the middle of a sudden rain shower that is providing the victorious plants with all the water they could ever dream of.

This cartoon is decorating my office door. It is a superb illustration of a classic reasoning fallacy. Calvin knows that if you water plants, they stay healthy. Based on this conditional (e.g., 'if, then') knowledge he infers that if he doesn't water the plants, it follows that they will die. In logic, this inference is known as the Denial of the Antecedent fallacy. A conditional links a specific precondition or cause (e.g., you water the plants) with a specific consequence or effect (e.g., the plants stay healthy). Logically speaking, a conditional utterance implies that if the precondition is met, the effect should always follow. However, this does not mean that if the precondition is not met, the effect cannot occur. There might be other conditions that can result in the occurrence of the effect (e.g., 'it might rain'). Most adults will spontaneously think of such alternative causes and this will help them to avoid Calvin's fallacious reasoning (and utter disappointment).

Over the last two decades it has been argued that the search for such alternatives (or 'counterexamples') lies at the core of the conditional reasoning process. It is assumed that when faced with a conditional, people will spontaneously search their semantic memory for stored alternative causes. Retrieval of such alternatives helps people to reject invalid inferences such as the above illustrated DA or related Affirmation of the Consequent (AC) fallacy.

Numerous studies showed that the easier it is for people to think of possible alternatives, the less they accept the invalid AC and DA inferences. In one of the classic studies Cummins, Lubart, Alksnis, and Rist (1991) measured in a pretest how many alternative causes people could generate for a set of conditionals. When Cummins ~~et al.,~~ used these conditionals in a reasoning task with different participants, she found that the invalid inferences were less frequently accepted when a conditional had many vs. few possible alternatives. Likewise, Markovits and Quinn (2002) measured individual differences in people's memory retrieval capacity (i.e., how easily they could come up with alternatives for a set of conditionals) in a pretest. A subsequent reasoning test showed that participants with a better retrieval capacity were less likely to commit the AC or DA fallacy.

Retrieving stored alternatives from memory makes people better logical reasoners: It helps us to reject *invalid* inferences. Unfortunately, people's tendency to take stored background knowledge about the conditional relation into account can also bias their reasoning. During reasoning people will not only spontaneously think of alternative causes of the conditional but also of so-called disabling conditions. These are possible conditions that prevent the effect from occurring even though the proper cause was present. When people take such disabling conditions into account they will also start rejecting *valid* inferences. For example, standard logic tells us that whenever the antecedent of a conditional occurs, we should conclude that the consequent will follow. This is the famous Modus Ponens (MP) inference, one of the most basic rules in classic logic. A dramatic illustration of people's failure to draw a simple MP inference is the disaster at the nuclear power plant in Chernobyl. The well-trained Russian operators clearly knew that if a specific safety test turned out positive, this implied that the reactor was overheated and should be shut down. Nevertheless, when that ill-fated day in 1986 the crucial safety test indeed turned out positive the operators did not draw the simple MP conclusion. What happened was that, just as many people in standard reasoning studies, they spontaneously thought of possible disablers such as 'the test is wrong' or 'maybe it's just an exercise'. Taking these disabling conditions into account resulted in a failure to draw the valid conclusion (and one of the biggest disasters in modern history).

Fortunately, the situation is not as dramatic as the Chernobyl example would suggest. Although the pervasive impact of disabler retrieval on people's willingness to draw the MP or related Modus Tollens (MT) inference is well-documented (e.g., Byrne, 1989; De Neys, Schaeken, & d'Ydewalle, 2002, 2003a) there is evidence that indicates that people manage to selectively block the impact of disablers. For example, stressing the logical nature of a reasoning task typically results in better performance on the valid problems, and this performance boost has been attributed to a selective inhibition of disablers (Vadeboncoeur & Markovits, 1999). A number of recent studies also argued that people's inhibitory capacity (i.e., the ability to resist interference from inappropriate memory activations) contributes to sound reasoning on the valid MP problems (e.g., De Neys, Schaeken, & d'Ydewalle, 2005; Handley, Capon, Beveridge, Dennis, & Evans, 2004; Markovits & Doyon, 2004; Simoneau & Markovits, 2003). Markovits and Doyon, for example, measured people's susceptibility to interference with a task based on a test where people had to refrain from the automatic tendency to complete a sentence with a strongly associated word. Such tests have been used as a measure of inhibition in patients with cerebral lesions (Burgess & Shallice, 1996). Participants also solved a conditional reasoning task with problems that had very salient disablers. Markovits and Doyon observed that people who performed well on the interference resistance measure did manage to reason correctly with the valid problems. As the authors noted, these peoples' inhibitory capacity apparently also helped them to discard the disabling information in the reasoning task.

As a result of the above findings, it has become very popular in the reasoning literature to characterize conditional reasoning as an interplay of a counterexample retrieval and inhibition process (e.g., De Neys et al., 2005; Markovits & Barrouillet, 2002; Quinn & Markovits, 2002; Simoneau & Markovits, 2003). Despite the popularity of this characterization, however, it is clear that the framework has moved a lot of the explanatory burden to memory mechanisms (e.g., search for alternatives and inhibition of disablers). If a memory researcher were to look at the studies, she would notice that the evidence for the postulated memory processes is typically quite indirect. Indeed, most reasoning studies focus on the output of the reasoning task. We tend to infer characteristics of the memory mechanism based on peoples' reasoning performance. For example, we typically measure the counterexample availability in a pilot study and test how people reason with the conditionals afterwards. When people reject invalid inferences for which the pretest showed that there were alternatives available, we attribute the rejection to successful retrieval of an alternative. Of course, the fact that the pilot work shows that people can easily think of an alternative or that the participant in question is very good at memory retrieval does not in itself imply that the alternative was also accessed and used while the participant was solving the reasoning task. Likewise, the fact that someone who tends to accept valid inferences also tends to score 'very good' on an inhibition test does not prove that disablers were actively discarded during the reasoning process. These assumptions are not unreasonable, but the problem is that they make strong postulations about memory activations during the reasoning process that are not unequivocally validated.

Validating the popular assumptions about the background knowledge retrieval and blocking calls for a more direct memory probing. The present study introduces a classic procedure from the memory literature to accomplish this goal. In the study participants solved a standard conditional reasoning task where they had to evaluate a set of valid and invalid arguments. After each reasoning problem participants were presented with a lexical decision task. In this task participants have to determine whether a string of presented letters is a word or not. Since the work of Meyer and Schvaneveldt (1971), the task has become one of the most popular methods to probe semantic memory activations. Half of the strings that were presented were non-words (e.g., 'golfrixnt'). When participants had just been presented with an invalid conditional argument (e.g., 'If you water the plants, they stay healthy. You do not water the plants. Therefore, they do not stay healthy') half of the words that were presented in the lexical decision task were possible alternatives (i.e., target words, e.g., 'rain') whereas the other half were completely unrelated words (e.g., 'letter'). If people think of possible alternatives while solving the reasoning problem, this should result in a classic facilitation effect in the lexical decision task. The remaining activation following the retrieval of the alternatives during reasoning should results in faster lexical decision times (i.e., the time people need to decide whether the letter string is a word) for the target than for the unrelated words.

When participants had just solved a valid conditional argument (e.g., 'If the test is positive, then the reactor is overheated. The test is positive. Therefore, the reactor is overheated) the target words in the subsequent lexical decision task were words that were closely associated with possible disablers (e.g., 'exercise') whereas the other half of the words were completely unrelated to the disablers. It is well established in memory studies that when people have to temporarily neglect information or avoid using it, recall of this information will be distorted (e.g., MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003; Neil, 1997; Tipper, 1985). Hence, temporarily putting your knowledge about possible disablers aside during reasoning should also hinder subsequent recall of these disablers and associated knowledge. Hence, if people really attempt to discard possible disablers that pop-up in mind during reasoning one would expect to see longer lexical decision times for target words than for unrelated words after people solved valid arguments.

In sum, the crucial prediction based on the counterexample retrieval and inhibition framework is that the lexical decision times should show an interaction between the type of reasoning problem and word type. If people retrieve alternatives while solving invalid problems, they should be faster to recognize the target vs. unrelated words. If people inhibit disablers while solving valid problems, this pattern should reverse and target words should take longer to recognize than unrelated words.

With respect to the validation, it is important to note that despite the popularity of the retrieval and inhibition framework there are possible alternative accounts. At the start of the psychological reasoning research era in the 1950s, for example, the reasoning community was convinced that the human reasoning engine did not care about the content of a problem and focused exclusively on the structure of a problem. According to this traditional, logistic view, reasoning is considered to be nothing more than applying a set of stored logical rules. Hence, people would accept or reject inferences based on whether the structure of the problem would (mis)match with their logical knowledge without any need to postulate an additional counterexample retrieval and blocking mechanism. The fact that people who score better on more general cognitive ability measures, such as retrieval efficiency or interference resistance, also reason better could be explained by assuming that these more gifted people also have a better, more precise logical database. Such a traditional logistic view faces it own problems but it underscores the point that the postulated activation of stored alternatives and blocking of disablers needs some direct validation.

AQ: Please provide the updated date.

In addition, recent studies have explicitly argued against the role of counterexample retrieval (e.g., Geiger & Oberauer, in press; Verschueren, Schaeken, & d'Ydewalle, 2005). Geiger and Oberauer argued that people's willingness to draw conditional conclusions would be estimated from the frequency of exceptions regardless of what causes them. In this view, people would not be retrieving specific stored counterexamples from memory but would rather go through a quick and less demanding probabilistic estimation process to compute the frequency with which exceptions to the rule occur. For example, when deciding to conclude whether or not the light inside the fridge will go on if they open the fridge, people would roughly estimate the number of times that the light inside the fridge did not go on in the past without searching for specific 'reasons' or counterexamples (e.g., the light bulb might be broken). The present study will allow us to test these alternative accounts. If the frequency or logistic views are right and people do not retrieve counterexamples during reasoning, then we should not observe any facilitation or distortion effects in the subsequent lexical decision task.

Experiment

Method

Participants

Twenty undergraduates studying at York University (Toronto, Canada) participated voluntarily. All participants were native English speakers or had lived more than 10 years in Canada or the United States.

Material

Reasoning task Participants were presented with a standard conditional reasoning task where they were asked to evaluate the validity of eight conditional inferences (four valid and four invalid problems). The four valid problems had a logically valid argument structure (either a Modus Ponens or the related Modus Tollens inference). We specifically selected conditionals for the valid problems for which previous pilot generation work (e.g., Cummins, 1995; De Neys et al., 2002, 2003b) had showed that people readily thought of many salient disabling conditions. It is assumed that taking these disablers into account normally results in the rejection of the valid inference (e.g., the introductory Chernobyl example) and will need to be avoided. The four invalid problems had a logically invalid argument structure (either an Affirmation of the Consequent or the related Denial of the Antecedent inference). For these invalid problems we specifically selected conditionals for which the pilot work showed that people could generate many possible alternative causes. It is assumed that taking these alternatives into account during reasoning will help rejecting the invalid inferences. Bellow is an example of the format of the reasoning task:

If Bart's food goes down the wrong way, then he has to cough.
Bart has to cough

Therefore, Bart's food went down the wrong way.
1. The conclusion is valid
2. The conclusion is invalid

A complete overview of all eight problems can be found in the Appendix.

Note that we selected conditionals with many disablers for the valid problems to make sure that disablers would be activated during the reasoning task. If people cannot think of a possible disabler there would not be a conflict between semantic knowledge and logic and no need for an inhibition process. Likewise, for the invalid problems we selected conditionals with many possible alternatives.

In theory, it is possible that when a conditional has very salient disablers, people will also think of these when solving an invalid problem. Although it has been shown that this has only a limited impact on people's judgment, we wanted to exclude any possible bias on the recall measure. Therefore, we made sure that the selected conditionals for the invalid problems had only a small number of disablers so that successful disabler retrieval was unlikely. Likewise, the selected conditionals for the valid problems had only a small number of possible alternatives.

To make sure that the reasoning task was not too repetitive we presented two subtypes of valid problems (MP and MT) and two subtypes of invalid problems (AC or DA). Since lexical decision data were similar for these respective subtypes, they were collapsed in the valid and invalid category.

Lexical decision task After each problem a total of 20 letter strings was presented. Participants indicated whether the string was a word or not by pressing one of two response keys. Half of the letter strings were non-words, the other half were English words. Five of the presented words were target words that were closely related to the possible counterexamples (i.e., disabling conditions for valid problems and alternative causes for invalid problems) of the conditional in the reasoning task. The other five words were unrelated to these counterexamples.

All target words were selected from the pilot generation material of De Neys et al. (2002; 2003b) where participants generated possible alternatives and disablers for the set of conditionals. Targets for valid problems were single words associated with frequently generated disabling conditions (e.g., 'exercise' or 'error' in the Chernobyl example) whereas target words for invalid problems were associated with frequently generated alternative causes (e.g., 'rain' or 'shower' in the Calvin example). Unrelated words were selected with the help of the online version of the Edinburgh Word Association Thesaurus (Kiss, Armstrong, Milroy, and Piper, 1973). After we had constructed an initial list of target and unrelated words, two raters were asked to validate the classifications. In the few cases that judgments diverged the specific word was replaced with an alternative that all parties could agree on. A complete overview of the selected words can be found in the Appendix.

The length and word frequency of presented targets and unrelated words was matched on each problem. The word strings were presented in random order with the restriction that targets could not be presented on consecutive trials.

Procedure

All participants were tested individually. Participants were first familiarized with the task format. They were shown an example of a reasoning problem and practiced the lexical decision task. It was clarified that in the actual experiment both tasks would always alternate. Instructions for the reasoning task were taken from Vadeboncoeur and Markovits (1999) and explicitly stressed the logical nature of the reasoning task. The eight reasoning problems were presented in random order. Each reasoning trial began with the appearance of a fixation point for 0.5 s, which was replaced by a problem to solve. The problem remained on the screen for a maximum of 15 s or until participants responded 'valid' or 'invalid' by pressing one of two response keys (previous work showed that people needed no longer than about 12 s to solve similar problems, e.g., De Neys et al., 2002). The lexical decision trials (with the 20 words specifically selected for that problem) started immediately after the key press. Words were presented in the centre of the screen and participants were instructed to respond as quickly as possible, while avoiding errors.

After the lexical decision trials, the experiment was briefly paused until the participant was ready to continue with the next reasoning problem.

Results and discussion

Reasoning task

The reasoning task was properly solved. Overall, 85% of the valid problems and 69% of the invalid problems were correctly evaluated with only small inter-individual performance variation (SD valid problems = .22, SD invalid problems = .26). These numbers are close to what Vadeboncoeur and Markovits (1999) obtained with a similar reasoning task and indicate that the lexical decision trials did not bias reasoning performance.

Lexical decision task

The central question concerns participants' performance on the lexical decision task. The dependent measure was the mean time (ms) taken by subjects to characterize each letter string as a word or non-word. As in the classic lexical decision studies, incorrect classifications of the letter strings were infrequent (less than 6% error rate across all trials) and where they did occur they were excluded from the analysis. Lexical decision times were submitted to a 2 (reasoning problem: valid or invalid) x 2 (word type: target or unrelated) repeated measures ANOVA. Figure 11.1 shows the results.

There was a main effect of problem type, $F(1, 19) = 8.72$, $p < .01$, $\eta^2_p = .31$, whereas the effect of word type was not significant, $F(1, 19) < 1$. As predicted, the two factors also interacted, $F(1, 19) = 5.42$, $p < .05$, $\eta^2_p = .22$. As Figure 11.1 shows, after solving an invalid problem lexical decision times for target words were faster than for unrelated words. When participants had just solved a valid problem, lexical decision times showed the opposite pattern with slower responses for target words than for unrelated words. Planned contrasts showed that lexical decision times on the target words were significantly longer after solving a valid problem than after solving an invalid problem $F(1, 19) = 14$, $p < .005$, $\eta^2_p = .42$. Lexical decision times did not differ for the unrelated words, $F(1, 19) < 1$.

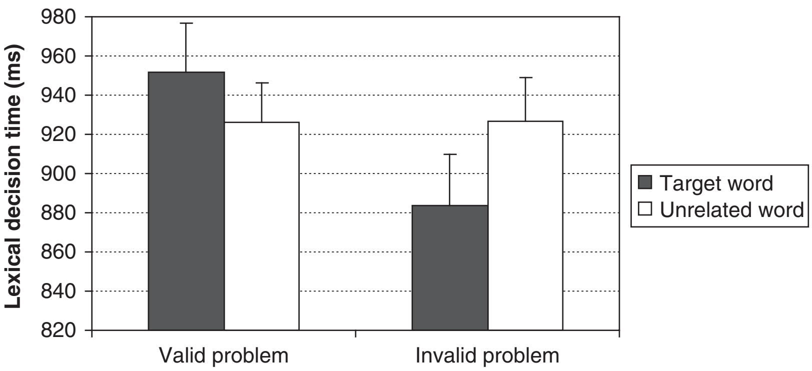

Fig. 11.1 Lexical decision times (ms) for target and unrelated words after participants solved valid and invalid conditional inference problems. Error bars are standard errors.

Conclusion

The present results nicely fit with the counterexample retrieval and inhibition view. When participants had solved an invalid problem, ~~lexical decision times for~~ target words that were related to possible alternatives were recognized faster than unrelated words. This facilitation effects provides direct evidence for the claim that the alternatives had already been accessed during the reasoning task. On the other hand, when participants had solved valid problems the target words that were related to possible disablers were recognized slower than unrelated words. The finding that the memory access for disabling information was temporarily impaired is prima facie evidence for the claim that this information was inhibited during the reasoning process. Note that the observed facilitation and distortion was specifically tied to the target words. Unrelated words that were not associated with possible counterexamples were not affected. Hence, it is not the case that conditional reasoning generally facilitated or impaired memory access. Results indicate that only semantic knowledge that was specifically associated with possible counterexamples had been activated. These findings are hard to reconcile with any framework that would deny the role of counterexample retrieval and inhibition in conditional reasoning. The memory probing approach indicates that the search for counterexamples lies at the very heart of the conditional reasoning process.

References

Burgess, P. W. & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, **34**, 263–273.

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

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

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

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory and Cognition*, **30**, 908–920.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology*, **42**, 177–190.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory and Cognition*, **31**, 581–595.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, **11**, 349–381.

Geiger, S. & Oberauer, K. (in press). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory and Cognition*.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. St. B. T. (2004). Working memory, inhibitory control, and the development of children's reasoning. *Thinking and Reasoning*, **10**, 175–195.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitkin, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies.* Edinburgh, UK: University Press.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation*, *Vol. 43* (pp. 163–206). San Diego, CA: Academic Press.

Markovits, H. & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, **22**, 5–36.

Markovits, H. & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory and Cognition*, **32**, 592–601.

AQ: Please provide the updated date.

Markovits, H. & Quinn, S. (2002). Efficiency of retrieval correlates with 'logical' reasoning from causal conditional premises. *Memory and Cognition*, **30**, 696–706.

Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence upon retrieval operations. *Journal of Experimental Psychology*, **90**, 227–234.

Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **23**, 1291–1305.

Quinn, S. & Markovits, H. (2002). Conditional reasoning with causal premises: Evidence for a retrieval model. *Thinking and Reasoning*, **8**, 179–191.

Simoneau, M. & Markovits, H. (2003). Reasoning with premises that are not empirically true : Evidence for the role of inhibition and retrieval. *Developmental Psychology*, **39**, 964–975.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, **37A**, 571–590.

Vadeboncoeur, I. & Markovits, H. (1999). The effect of instruction and information retrieval on accepting the premises in a conditional reasoning task. *Thinking and Reasoning*, **5**, 97–113.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). Everyday conditional reasoning: A working memory–dependent tradeoff between counterexample and likelihood use. *Memory and Cognition*, **33**, 107–119.

ED:Should the Appendix come before references?

Appendix

Table A1 List of conditionals for the reasoning task

Valid problems:
1. If the trigger is pulled, then the gun fires. (MP)
2. If the correct switch is flipped, then the porch light goes on. (MT)
3. If the ignition key is turned, then the car starts. (MP)
4. If the match is struck, then it lights. (MT)
Invalid problems:
5. If Bart's food goes down the wrong way, the he has to cough. (DA)
6. If Mary jumps in the swimming pool, the she gets wet. (AC)
7. If the apples are ripe, then they fall from the tree. (DA)
8. If the water is poured on the campfire, then the fire goes out.(AC)

Table A2 List of selected target and unrelated words in the lexical decision task for each of the eight reasoning problems

	Target	**Unrelated**
Conditional 1	unloaded	waitress
	blank	onion
	broken	forest
	safety	author
	jammed	monkey
Conditional 2	failure	manager
	blackout	overcoat
	old	amp
	burnedout	arrowhead
	removed	leagues
Conditional 3	damaged	invites
	busted	floral
	empty	guest
	wrong	shore
	refill	beaver

Conditional 4	wet	tar
	damp	lion
	worn	cady
	used	face
	softly	spring
Conditional 5	cold	data
	sick	wave
	dry	paw
	asthma	lentil
	smoke	smirk
Conditional 6	rain	rice
	shower	writer
	hosed	zebra
	bath	fare
	splashed	gangster
Conditional 7	storm	ivory
	wind	wall
	birds	paint
	shaken	stream
	dropped	surgeon
Conditional 8	died	mark
	smother	pigment
	sand	card
	extinguish	politician
	blanket	article