THINKING & REASONING, 2005, 11 (4), 349–381

Psychology Press
Taylor & Francis Group

Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples

Wim De Neys, Walter Schaeken, and Géry d'Ydewalle
University of Leuven, Belgium

Two experiments examined the contribution of working memory (WM) to the retrieval and inhibition of background knowledge about counterexamples (alternatives and disablers, Cummins, 1995) during conditional reasoning. Experiment 1 presented a conditional reasoning task with everyday, causal conditionals to a group of people with high and low WM spans. High spans rejected the logically invalid AC and DA inferences to a greater extent than low spans, whereas low spans accepted the logically valid MP and MT inferences less frequently than high spans. In Experiment 2, an executive-attention-demanding secondary task was imposed during the reasoning task. Findings corroborate that WM resources are used for retrieval of stored counterexamples and that people with high WM spans will use WM resources to inhibit the counterexample activation when the type of counterexample conflicts with the logical validity of the reasoning problem.

The ability to think conditional, if-then, thoughts is considered as one of the cornerstones of our mental equipment. As Edgington (1995, p. 235) puts it, "there would not be much point in recognizing that there is a predator in your path unless you also realize that if you don't change direction pretty quickly you will be eaten". Similarly, when someone warns you "If you don't stop bugging me, I'll beat you", and you want to avoid being beaten up by an angry person, you need to draw a conditional inference.

Given the central role that conditional reasoning plays in our causal knowledge system and social interactions, it is not surprising that it has become one of the most intensely studied topics in human reasoning

---

Correspondence should be addressed to Wim De Neys, Lab Experimentele Psychologie, K.U. Leuven, Tiensestraat 102, 3000 Leuven, Belgium.
Email: Wim.Deneys@psy.kuleuven.ac.be

Preparation of the manuscript was supported by grants from the Fund for Scientific Research-Flanders (FWO-Vlaanderen). We would like to thank Walter Schroyens for his comments on an earlier draft of the manuscript.

http://www.tandf.co.uk/journals/pp/13546783.html DOI: 10.1080/13546780442000222

research (Evans, Newstead, & Byrne, 1993). Many reasoning theories appeal to the notion of a limited-capacity working memory in their explanation of reasoning performance (e.g., Braine & O'Brien, 1998; Johnson-Laird & Byrne, 1991; Rips, 1994). Although the proposed reasoning mechanisms differ, the central assumption is that reasoning errors may occur when the capacity of working memory is overburdened.

There is evidence for a general link between working memory capacity and performance in a range of reasoning tasks (e.g., Barrouillet, 1996; Capon, Handley, & Dennis, 2003; Gilhooly, Logie, & Wynn, 1999; Kyllonen & Christal, 1990; Süss, Oberauer, Wittmann, Wilhelm, & Schulze, 2002): People with higher scores on standard working memory tests tend to draw more logically correct conclusions. A few studies have established this link in the specific case of conditional reasoning (e.g., Barrouillet & Lecas, 1999; Markovits, Doyon, & Simoneau, 2002). Some researchers have even moved beyond a merely correlational approach and have shown that burdening working memory with a secondary task gives rise to conditional reasoning errors (e.g., Klauer, Stegmaier, & Meiser, 1997; Meiser, Klauer, & Naumer, 2001; Toms, Morris, & Ward, 1993).

While there is some evidence for the involvement of working memory in conditional reasoning, there is an important caveat in the current studies. These studies have almost exclusively (excepting Markovits et al., 2002) adopted "abstract" conditionals of the form "If square, then circle". We label these conditionals "abstract" because people have no prior knowledge about the relation that the conditional expresses. In everyday life we typically reason with meaningful and content-rich conditionals (e.g., "If you put fertiliser on plants, then they grow well"). Here our long-term memory contains prior knowledge about the conditional (e.g., you might think of the fact that in order to grow well, the plants also need sunlight) and it is long established that this knowledge has a massive impact on the inferences people draw (e.g., Staudenmayer, 1975). It was precisely to sidestep this background knowledge effect that studies on the role of working memory in conditional reasoning have explicitly preferred content-lean conditionals (Barrouillet & Lecas, 1999; Meiser et al., 2001). However, the ultimate goal of a psychological reasoning theory is to account for people's daily life reasoning (e.g., Galotti, 1989; Johnson-Laird, 1983; Johnson-Laird & Byrne, 2002; Oaksford & Chater, 1998). If working memory is assumed to be involved in reasoning, it is crucial to examine its role in reasoning with the typical content-rich conditionals we use in everyday life. The present article starts this examination. We focus on two important functions: The retrieval and inhibition of background knowledge about counterexamples from long-term memory.

Working memory (WM) is often characterised as a hierarchically organised system in which specific storage and maintenance components subserve a central component responsible for the control of information processing (e.g., Baddeley & Hitch, 1974; Cowan, 1995; Engle & Oransky, 1999). The controlling component or "central executive" is conceived as a limited-capacity system that regulates the allocation of attentional resources. Executive functioning is mediated by the prefrontal cortex (Wickelgren, 1997). Performance on standard working memory tests is assumed to reflect primarily central executive capacity (Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Bleckley, Conway, & Engle, 2001). In line with this view, the present investigation of working memory and everyday conditional reasoning focuses on the role of the central executive and not on the storage or slave systems. We also note that in our discussion we treat WM resources as a domain-free attentional capacity. Thus, WM resources are hypothesised to be modality aspecific (e.g., we do not distinguish a separate verbal and visual working memory; see Engle, 2002; Engle & Oransky, 1999).

Investigations of conditional reasoning typically ask people to assess arguments of the following four kinds (in their abstract form):

Modus Ponens (MP) If p then q, p, therefore q
Modus Tollens (MT) If p then q, not q, therefore not p
Denial of the Antecedent (DA) If p then q, not p, therefore not q
Affirmation of the Consequent (AC) If p then q, q, therefore p

In standard logic, MP and MT are considered valid inferences, while AC and DA inferences are considered fallacies. So, when you are told "If fertiliser is put on plants, then they grow well" and you receive the information that fertiliser was indeed put on the plants, then logic tells you to accept the conclusion that the plants grow well (an MP inference). Likewise, if you receive the information that the plants do not grow well, you should infer that the plants were not fertilised (an MT inference). On the other hand, logically speaking, from the information that the plants grow well, you should not infer that the plants were fertilised (an AC inference). Likewise, upon knowing that the plants were not fertilised, you should reject the conclusion that therefore the plants will not grow well (a DA inference).

Research on the impact of background knowledge about the conditional relation has showed that there are at least two important kinds of information stored in long-term memory that affect the inferences people draw: alternative causes and disabling conditions. Both are referred to as "counterexamples" (Byrne, 1989). Consider for example the conditional

If the brake is depressed, then the car slows down.

This conditional expresses a causal relation between a cause, depressing the brake, specified in the first (the antecedent) part of the conditional, and an effect, slowing down, specified in the second (the consequent) part of the conditional. An alternative cause (alternative) is a possible cause that can produce the effect mentioned in the conditional while a disabling condition (disabler) prevents the effect from occurring despite the presence of the cause. For example, possible alternative causes for the conditional are:

> Running out of gas, having a flat tyre, shifting the gear down, …

The alternatives make it clear that it is not necessary to depress the brake in order to slow the car down. Other causes are also possible.

Possible disabling conditions are:

> A broken brake, accelerating at same time, skid due to road conditions, …

If such disablers are present, depressing the brake will not result in the slowing down of the car. The disablers make it clear that depressing the brake is not sufficient for the slowing down of the car. Additional conditions have to be fulfilled.

Pioneering studies have examined the impact of counterexample retrieval on conditional reasoning by the explicit presentation of counterexamples (e.g., Byrne, 1989; Rumain, Connell, & Braine, 1983). For example, Byrne (1989) found that explicitly mentioning a possible disabler like "If the library is open, then Ann studies late in the library" for the conditional "If she has an essay to write, then she studies late in the library" decreased acceptance of MP and MT. Further studies established the importance of the outcome of the search for stored counterexamples. Cummins (1995; see also Cummins, Lubart, Alksnis, & Rist, 1991; Thompson, 1994) manipulated the availability of possible counterexamples. She adopted conditionals for which pilot work indicated that people could retrieve many counterexamples (e.g., a conditional with many possible alternatives: "If you study hard, then you pass the exam") or few counterexamples (e.g., a conditional with few possible alternatives: "If you grasp the glass with your bare hands, then your fingerprints are on it"). For conditionals with many alternatives, where successful retrieval was very likely, AC and DA were less accepted than for conditionals with only few possible alternatives. Likewise, MP and MT were less accepted when a conditional had many disablers than when only a few were available.

In addition, reasoning performance has been related to individual differences in the efficiency of the counterexample retrieval process (e.g., Janveau-Brennan & Markovits, 1999; De Neys, Schaeken, & d'Ydewalle, 2002b). In these studies participants were first presented with a generation task where they were asked to generate, in a limited time, as many counterexamples as possible for a set of conditionals. The same participants then received a conditional reasoning task with different conditionals. Janveau-Brennan and Markovits found that the more alternatives one could generate in the generation task, the more AC and DA were rejected in the reasoning task. Likewise, De Neys et al. observed that better disabler generation capacity resulted in lower MP and MT acceptance ratings.

We hypothesised that working memory capacity would be important for everyday conditional reasoning because an efficient counterexample retrieval would require WM resources. Although reasoning theories generally conceive the process where background knowledge is accessed as an undemanding, automatic mechanism (e.g., Cummins, 1995; Evans, 2002; Evans & Over, 1996; Newell & Simon, 1972; Stanovich & West, 2000), De Neys, Schaeken, and d'Ydewalle (2005) recently presented evidence against the popular automaticity claim in a memory retrieval study with familiar conditionals. De Neys et al. asked participants to generate as many counterexamples as possible in a limited time for a set of familiar, causal conditionals (e.g., If the match is struck, then it lights). Results indicated that participants higher in WM capacity were better at retrieving counterexamples and that burdening the executive resources with a secondary task also decreased the counterexample retrieval efficiency. Consistent with recent memory studies (e.g., Moscovitch, 1995; Rosen & Engle, 1997), the findings implied that in addition to an automatic counterexample search process based on a passive spreading of activation, people also allocate WM resources to a more active and efficient search process.

Following Markovits and Barrouillet (2002), De Neys et al. (2005) proposed an elementary sketch of the counterexample retrieval process. It was suggested that when drawing conditional inferences, reasoners construct and maintain a mental representation of the premises in working memory. Next, activation automatically starts to spread from the mental representation in WM towards associated counterexamples in long-term memory (Anderson, 1993; Cowan, 2001). The stored counterexamples are thereby conceived as nodes in a semantic network (Anderson, 1983). A counterexample will be retrieved when a node's activation level crosses a critical threshold. More strongly associated (i.e., more "salient") counterexamples have lower retrieval thresholds and will be retrieved more easily (De Neys, Schaeken, & d'Ydewalle, 2003a; Quinn & Markovits, 1998). The spreading of activation would require little in the way of executive attention and can suffice to activate the most strongly associated counterexamples. Available WM resources would be used next for an active, strategic search to access new counterexamples: The larger the WM-resource pool is, the more resources can be allocated to the search, and the more efficient the search will be.

If De Neys et al. (2005) are right that efficient counterexample retrieval for everyday, causal conditionals requires WM resources, then inter-individual differences in WM capacity should also affect reasoning performance with these conditionals. Remember that the extent to which a reasoner accepts, for example, the Denial of the Antecedent (DA) and Affirmation of the Consequent (AC) inferences depends on the number of alternatives one can retrieve. De Neys, Schaeken, and d'Ydewalle (2003b) observed, for example, that inference acceptance linearly decreased with every retrieved counterexample. Thus, the more alternatives that can be retrieved, the less DA and AC are accepted. Consequently, a less efficient alternative retrieval should result in a higher DA and AC acceptance. In Experiment 1 we therefore compare the everyday conditional reasoning performance of individuals with high ("high spans") and low ("low spans") scores on a measure of working memory capacity. Since a high WM span should allow more efficient counterexample retrieval, we expect that the high spans will be less inclined to accept AC and DA compared to low spans.

Since disabler retrieval results in lower MP and MT acceptance ratings, one could also expect that, because of the more efficient disabler search, high spans will more frequently reject the MP and MT inferences. Remember, however, that whereas AC and DA are logical fallacies, MP and MT are logically valid. Rejecting AC and DA is in line with standard logic, whereas rejecting MP and MT is not. For people who approach the logical standards, there might be a dissonance during reasoning between the tendency to reject MP and MT and the valid status of these inferences.

In the reasoning literature there is some debate about whether people are able to adhere to normative standards such as standard logic in reasoning (e.g., Evans & Over, 1996; Stanovich & West, 2000). Individual difference studies indicate that at least people of high cognitive capacity appear to have a logical "decontextualisation" tendency: A basic ability to put background knowledge aside when it conflicts with the logical standards (e.g., Stanovich & West, 2000). In these studies cognitive ability is typically operationalised in terms of scores on general intelligence tests that show a strong connection to working memory test performance (see Engle et al., 1999). If high spans have an elementary notion of logical validity, this should conflict with the automatic disabler retrieval component. In this case one might hypothesise that high spans will use their working memory resources for an active blocking or inhibition of the disabler retrieval (e.g., Gilinsky & Judd, 1994; Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Stanovich & West, 1998, 2000).

The inhibition of responses deemed inappropriate is considered as one of the key executive functions (e.g., Baddeley, 1996; Dempster & Corkill, 1999; Engle et al., 1999; Miyake & Shah, 1999; Shallice & Burgess, 1993). It has also been demonstrated that these inhibition mechanisms can be targeted at memory traces to control retrieval (e.g., Anderson & Bell, 2001; Conway & Engle, 1994; Radvansky, 1999; for a review see Levy & Anderson, 2002).

Markovits and Barrouillet (2002) have already put forward the possibility of a disabler inhibition process in conditional reasoning. In a study with young children, Simoneau and Markovits (2003) also showed that a task designed to measure the efficiency of inhibitory processing predicted whether or not MP was accepted (see Handley, Capon, Beveridge, Dennis, & Evans, 2004, for related findings on syllogistic reasoning). More general evidence for an inhibitory mechanism in reasoning comes from a neuroimaging study with highly educated participants (mostly graduate students) on syllogistic reasoning (Goel, Buchel, Frith, & Dolan, 2000). In conditions where the logical status of the conclusion conflicted with background knowledge (e.g., a valid but unbelievable conclusion like "Some of the communists are golfers. All of the golfers are capitalists. Therefore, some of the communists are capitalists.") two regions of the right prefrontal cortex (Brodmann areas 8 and 46/45) were specifically activated. Goel et al. argued that this activation reflects an inhibitory mechanism that is blocking the impact of background knowledge (see also Goel & Dolan, 2003).

If high spans rely on a disabler inhibition mechanism during everyday conditional reasoning, we should see higher MP and MT acceptance ratings for the high spans, as compared to the low spans. Indeed, given that De Neys et al.'s (in press) memory study showed that high spans have a superior disabler retrieval capacity, a higher MP and MT acceptance during reasoning would indicate that high spans refrain from taking disablers into account in the reasoning task. This would be consistent with the hypothesis of a general WM-dependent inhibition or decontextualisation mechanism (Stanovich & West, 2000). The prediction was tested in Experiment 1. Experiment 2 presents additional, more direct evidence for the role of working memory in the retrieval and inhibition of counterexamples by examining the effects of a secondary WM load on reasoning performance.

EXPERIMENT 1

In Experiment 1 we compared the performance of people with low and high WM span in an everyday conditional reasoning task. Retrieving alternatives is known to decrease acceptance of the logically fallacious AC and DA inferences (e.g., Byrne, 1989; Cummins, 1995; De Neys et al., 2003b; Janveau-Brennan & Markovits, 1999; Quinn & Markovits, 1998). De Neys et al. (2005) showed that high spans are better at retrieving alternatives. Therefore, we expect that high spans will be less inclined to accept AC and DA compared to low spans.

Retrieving disablers is known to decrease acceptance of the valid MP and MT inferences (e.g., Byrne, 1989; Cummins, 1995; De Neys et al., 2003b).

The inhibition hypothesis states that high spans will use their WM resources to inhibit spontaneous disabler access. Remember we stated that the counterexample retrieval process starts with an automatic spreading of activation, after which WM resources will be recruited for a more active search. The inhibition hypothesis entails that where retrieving disablers is concerned, high spans will not use WM resources for an active search but rather for an inhibition of automatically activated disablers. If the hypothesis is correct, we should observe that high spans accept MP and MT more than the low spans, who allocate their WM resources primarily to retrieval. Thus, the central test is an interaction between WM capacity and inference type: Whereas high spans should tend to accept MP and MT to a greater extent, low spans should show higher AC and DA acceptance ratings.

Following Cummins (1995), the number of available counterexamples for the conditionals in the present reasoning task varied systematically (i.e., half of the conditionals had many/few possible alternatives/disablers). This manipulation is important because, under the assumption that high spans have some kind of basic logic notion, it might be suggested that they will use this notion to reject the fallacious AC and DA inferences. Thus one might argue that a lower AC and DA acceptance rating for high spans does not result from counterexample retrieval but rather from a purely abstract, content-free reasoning ability. If high spans' inference acceptance were solely determined by their logical knowledge, the availability of alternatives should not affect the conclusions. We expect that the alternative search process is crucial for both high and low spans (e.g., De Neys et al., 2005). Retrieval will be more successful (i.e., more counterexamples will be retrieved) when many counterexamples are stored. Thus, if retrieval of alternatives determines the inferences that people draw, inference acceptance should be affected by the number of available alternatives. Contrary to the abstract reasoning ability hypothesis, we therefore predict that both low and high spans will be affected by the number-of-alternatives factor.

The inhibition hypothesis entails that both span groups should be affected by the number of disablers of the conditionals. Having an intuitive notion of some basic logical principles does not guarantee that one will always draw correct, logical inferences (Jacobs & Klaczynski, 2002; Klaczynski, 2001a). It is explicitly assumed that inhibiting the retrieval process is difficult and WM-resource demanding (Stanovich & West, 1998, 2000). Although high spans might have some logical competence, the actual performance will depend on the demands of the inhibition process. De Neys et al. (2005) proposed that, except in specific cases, the automatic retrieval process would not be very successful for causal conditionals. The specific cases will be counterexamples that are very strongly associated with the conditional (i.e., stored counterexamples with a low activation threshold, see

De Neys et al., 2005; Markovits & Quinn, 2002). The strength of association between a counterexample and a conditional has been shown to affect successful retrieval (De Neys et al., 2003a; Quinn & Markovits, 1998). Conditionals with many counterexamples typically also have more strongly associated counterexamples (De Neys et al., 2002b, Experiment 1). Thus, more instances will have to be inhibited for the conditionals with many possible disablers. Under the assumption that high spans' disabler inhibition is an attention-demanding process, one should expect that an increase in the inhibition demands will result in a less successful inhibition. Therefore, the inhibition claim entails that although high spans should show higher MP and MT acceptance (vs low spans) overall, even the high span group should show an impact of the number of disablers.

Method

Screening for working memory capacity. We screened participants for working memory capacity using a version of the Ospan task (La Pointe & Engle, 1990) adapted for group testing (Gospan, for details see De Neys, d'Ydewalle, Schaeken, & Vos, 2002a). The main adaptation was that we first presented the operation from an operation–word pair on screen (e.g., "IS $(4/2) - 1 = 5$?"). Participants read the operation silently and pressed a key to indicate whether the answer was correct or not. Responses and response latencies were recorded. After the participant had typed down the response, the corresponding word (e.g., "BALL") from the operation–word string was presented for 800 ms. As in the standard Ospan, three sets of each length (from two to six operation–word pairs) were tested and set size varied in the same randomly chosen order for each participant. The Ospan score was the sum of the recalled words for all sets recalled completely and in correct order.

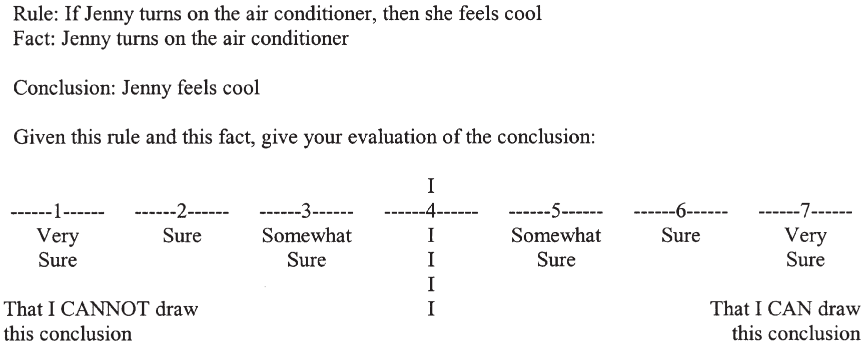
Participants were tested in groups of 38 to 48 at the same time in a large computer room with an individual booth for every participant. Data for participants who made more than 15% maths errors or whose mean operation response latencies deviated by more than 2.5 standard deviations of the sample mean were discarded. The internal reliability coefficient alpha for the Gospan was .74 and the corrected correlation between standard Ospan and Gospan score reached .70 (see De Neys et al., 2002a).

Participants. A total of 52 first-year psychology students from the University of Leuven, Belgium, participated in return for psychology course credit or 5 euro. These participants were identified from a larger pool of 426 first-year psychology students who had participated in the Gospan task: 26

participants were selected from the top quartile of the distribution ("high spans") and 26 were selected from the bottom quartile ("low spans"). Between 45 and 90 days intervened between a given individual's participation in the Gospan task and the reasoning task. None of the participants had received any training in formal logic.

Materials. Eight causal conditionals from the generation study of De Neys et al. (2002b) were selected for the reasoning task (see Appendix Table A1). The conditionals were selected so that the number of available counterexamples constituted a 2 (few/many) × 2 (alternatives/disablers) design with two items per cell. The eight conditionals were embedded in the four inference types (MP, DA, MT, and DA), producing a total of 32 inferences for each participant to evaluate.

The experiment was run on computer. Each argument was presented on screen together with a 7-point rating scale and accompanying statements. This resulted in the following format:



Rule: If Jenny turns on the air conditioner, then she feels cool
Fact: Jenny turns on the air conditioner

Conclusion: Jenny feels cool

Given this rule and this fact, give your evaluation of the conclusion:

			I			
------1------	------2------	------3------	------4------	------5------	------6------	------7------
Very Sure	Sure	Somewhat Sure	I I I I	Somewhat Sure	Sure	Very Sure

That I CANNOT draw
this conclusion

That I CAN draw
this conclusion

Each of the 32 arguments was presented in this way. The premises and conclusion were presented in yellow. The remaining text appeared in white on a black background.

Procedure. All participants were tested individually for the reasoning task. Reasoning task instructions were presented verbally and on screen. They showed an example item that explained the specific task format. Participants were told that the task was to decide whether or not they could accept the conclusions. Care was taken to make sure participants understood the precise nature of the rating scale. Instructions stated that there were no time limits. The experimental session was preceded by one practice trial.

Participants said out loud the number reflecting their decision. The experimenter typed this number down on a keyboard connected to the computer running the reasoning task. Following the key press, participants saw the text "NEXT ITEM" (grey letters on black background) for 800 ms after which the next item was presented.

We constructed four sets of eight inferences each. All eight inferences in a set were based on different conditionals. There were two sets with four MP and four AC inferences in each set, and another two sets with four MT and four DA inferences in each set. The order of presentation of the inferences within a set was random. The conditionals for the four inferences of the same type in a set were taken from the four different cells within the 2 (few/many) × 2 (alternatives/disablers) design that the conditionals constituted. Half of the participants received the sets in the order MP/AC, MT/DA, MP/AC, and MT/DA. For the other half the sets were presented in the reversed order MT/DA, MP/AC, MT/DA, and MP/AC. After two sets (i.e., 16 inferences) were evaluated, item presentation was paused until participants decided to continue.

Participants were told that they could evaluate the conclusions by the criteria they personally judged relevant (see Cummins, 1995). Thus, no explicit attempts were made to instruct people to reason logically. Although participants were still situated in a laboratory setting, this should allow people to reason as they would in everyday life (Cummins, 1995; see also Galotti, 1989).

Results

Each participant evaluated inferences based on two different conditionals within each cell of the 2 (number of alternatives) × 2 (number of disablers) × 4 (inference type) cell of the design. The mean of these two observations was calculated. These means were subjected to a 2 (span group) × 2 (number of alternatives) × 2 (number of disablers) × 4 (inference type) mixed model ANOVA with span group as between-subjects factor and number of alternatives, number of disablers, and inference type as within-subject factor. The within-subject part of the design is a replication of Cummins (1995) and De Neys et al. (2002b). The primary foci of the present study are the interactions with the span group factor.

Effects involving repeated measures with more than two levels are reported as significant if they pass the Greenhouse and Geisser (1959) correction criterion.

As Figure 1 shows, there was a significant interaction between inference type and span group, $F(3, 150) = 4.60$, $MSE = 2.02$, $p < .005$. Consistent with the hypotheses, high spans accepted the MP and MT inferences more than low spans, whereas the low spans accepted AC and DA more than high spans, $F(1, 50) = 7.7$, $MSE = 3.47$, $p < .009$.

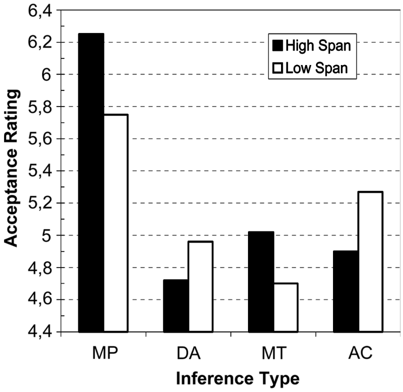


Figure 1. High and low spans' mean acceptance rating of the four inference types. The rating scale ranged from 1 (very sure cannot draw this conclusion) to 7 (very sure can draw this conclusion).

The results for the number of alternatives, number of disablers, and inference type factors completely replicated the standard findings of Cummins (1995) and De Neys et al. (2002b): There was an effect of inference type, $F(3,\ 150) = 30.88$, $MSE = 2.02$, $p < .0001$, and this effect interacted with number of disablers, $F(3,\ 150) = 24.50$, $MSE = 0.76$, $p < .0001$, and number of alternatives, $F(3,\ 150) = 28.76$, $MSE = 0.75$, $p < .0001$. The number of disablers primarily affected MP and MT acceptance ratings; for conditionals with many disablers MP, $F(1,\ 50) = 26.9$, $MSE = 0.4$, $p < .001$, and MT, $F(1,\ 50) = 10.43$, $MSE = 1.2$, $p < .003$, were accepted less than for conditionals with few disablers. The number of alternatives primarily affected AC and DA acceptance ratings; AC and DA were accepted more when there were only few possible alternatives than when there were many of them, $F(1,\ 50) = 125.29$, $MSE = 1.1$, $p < .001$, and $F(1,\ 50) = 94.38$, $MSE = 1.16$, $p < .001$. As did De Neys et al., we also observed an impact of alternatives on MP, $F(1,\ 50) = 24.65$, $MSE = 0.36$, $p < .001$, and MT acceptance, $F(1,\ 50) = 11.81$, $MSE = 0.9$, $p < .005$, and of disablers on AC, $F(1, 50) = 30.78$, $MSE = 0.53$, $p < .001$, and DA, $F(1, 50) = 18.05$, $MSE = 0.96$, $p < .001$, acceptance.

More important is the question of whether the effects of number of alternatives and disablers interacted with span group. Figure 2 illustrates the main findings (see Appendix Table A2 for a complete overview).

High and low spans were equally affected by the number of alternatives; the Span × Number of Alternatives, $F(1,\ 50) = 2.35$, $p > .13$, and the Span × Number of Alternatives × Inference Type interactions were not significant, $F(3, 150) < 1$. More specifically, neither on AC, $F(1, 50) < 1$, nor on DA, $F(1,\ 50) < 1$, did span group affect the effect of number of alternatives.

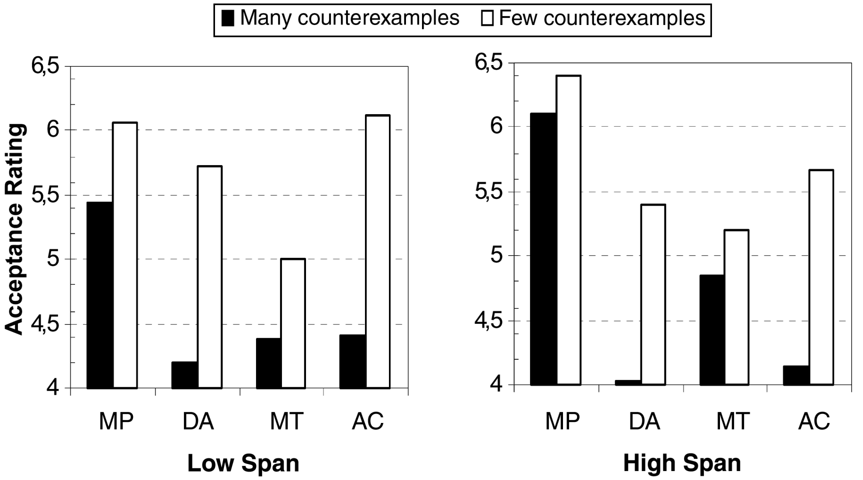


Figure 2. High and low spans' mean inference acceptance ratings as a function of the number of available disablers (MP and MT) and alternatives (AC and DA). The rating scale ranged from 1 (very sure cannot draw this conclusion) to 7 (very sure can draw this conclusion).

The impact of the number of disablers on MP and MT seemed somewhat smaller for the high spans, but neither the Span × Number of Disablers, $F(1,\ 50) > 1$, nor the Span × Number of Disablers × Inference Type interaction, $F(3,\ 150) = 1.62$, $MSE = 0.76$, $p > .18$, reached significance. More specifically, neither on MP, $F(1, 50) < 1$, nor on MT, $F(1, 50) < 1$, did span group affect the effect of number of disablers.

It could be argued that the effect of the number of disablers on high spans' MP and MT acceptance in the present analysis resulted from an aggregation confound. That is, the overall number-of-disablers effect might be caused by a subset of high spans, whereas the vast majority would not be affected. However, an examination of the individual acceptance patterns indicated that there were only 2 out of 26 high spans (1 out of 26 low spans, $p_1 = .08$, $n_1 = 26$, vs $p_2 = .04$, $n_2 = 26$, $p > .25$) whose mean MP and MT ratings were not affected by the number-of-disablers factor. Hence it is rather unlikely that the reported findings could be attributed to an aggregation confound.

Discussion

Results of Experiment 1 are in line with the hypothesised role of working memory in conditional reasoning: The lower AC and DA acceptance ratings for high spans support the hypothesis that WM capacity mediates the retrieval of counterexamples during reasoning (e.g., De Neys et al., 2005).

The higher WM capacity, the more efficient the search process, and the less AC and DA will be accepted. The higher MP and MT acceptance ratings for high spans, despite their intrinsic superior disabler-retrieval capacity, indicate that high spans actively try to refrain from taking disablers into account. This finding is consistent with the inhibition hypothesis.

Both high and low spans' inference acceptance was affected by the number of possible alternatives. These findings establish that even for high spans the retrieval of alternatives is a crucial factor in the inference evaluation.

The number of disablers of a conditional also affected both span groups. This indicates that high spans' disabler inhibition is not complete. Remember we proposed that the counterexample retrieval starts with an automatic spreading of activation. Except in specific cases this automatic retrieval process would not be very successful for causal conditionals. The specific cases will be counterexamples that are very strongly associated with the conditional (De Neys et al., 2005; Quinn & Markovits, 2002). The strength of association between a counterexample and a conditional has been shown to affect successful retrieval (De Neys et al., 2003a; Quinn & Markovits, 1998). Conditionals with many counterexamples typically also have more strongly associated counterexamples (De Neys et al., 2002b, Experiment 1). Thus, more instances will have to be inhibited for the conditionals with many possible disablers. Therefore, the inhibition should be somewhat less successful; occasionally a disabler will "slip through". Consequently, although MP and MT acceptance will be higher for high spans overall, even high spans' MP and MT acceptance will be affected by the number of available disablers.[1]

The present results are supported by an experiment of Markovits et al. (2002), conducted independently from our study. Whereas Markovits et al. did not examine the crucial effect of the number of counterexamples, they did observe that the higher a participant's score on a WM task, the less frequently AC and DA inferences were accepted and the more frequently MP inferences were accepted (no relation was observed for MT). Although the results may have been affected by the fact that Markovits et al. explicitly instructed participants to reason logically, the data pattern does point to the generality of the present findings.

---

[1] One might note that our hypothesis merely stated that both span groups would be affected by the number of counterexamples factor. No assumptions were made about the size of the effect. The problem is that possible differences in effect size will depend on the exact number of activated counterexamples for the many and few conditionals in both span groups. Currently we cannot determine this figure unequivocally. Post hoc, the absence of a Counterexample × Span interaction at least suggests that the difference between the number of activated counterexamples for the few and many conditionals is constant in both span groups.

We mentioned in the introduction that De Neys et al. (2002b, Experiment 3) found that a higher efficiency of the disabler retrieval process is associated with lower MP and MT acceptance. Given that high spans are better at retrieving disablers, these findings might seem to contradict the presently observed stronger tendency for high spans to accept MP and MT. However, De Neys et al.'s sample consisted of 40 randomly selected participants. The inhibition hypothesis only claims that the inhibition occurs for people with the highest cognitive abilities. Therefore, the present study specifically selected participants from the top quartile of first-year psychology students' WM capacity distribution. Thus, when the top WM levels are excluded (or are small in number, as was probably the case in De Neys et al.) we would indeed expect that higher WM capacity (and thus better retrieval) would result in lower MP and MT acceptance. In this respect it would be interesting to complement the present findings in future studies with a more continuous approach, where reasoners from a broader capacity range (i.e., not just the top and bottom levels) are tested.

Although the results of Experiment 1 support the hypothesised role of working memory in the retrieval and inhibition of counterexamples, the findings remain, of course, purely correlational. Additional, more decisive evidence is required. Experiment 2 presents a more direct test of the hypotheses.

EXPERIMENT 2

If working memory resources are used for retrieval and inhibition of counterexamples during reasoning, putting a secondary load on the executive WM resources should interfere with the proper functioning of these processes. We hypothesised that low spans primarily allocate their WM resources to retrieval. Under load conditions retrieving counterexamples will be less efficient and thus successful retrieval of alternatives and disablers will be less likely. Since successful alternative retrieval will decrease AC and DA acceptance, and disabler retrieval will decrease acceptance of MP and MT, we predict that under load conditions (due to the less efficient retrieval process) low spans' acceptance of the four different inferences will increase. Since high spans also have difficulties in retrieving alternatives under load conditions (De Neys et al., 2005) we expect that the secondary task will also increase high spans' level of AC and DA acceptance.

The inhibition hypothesis states that high spans are using WM resources to prevent the automatic activation of disablers. Since the inhibition process is explicitly assumed to be WM resource demanding (e.g., Kokis et al., 2002; Stanovich & West, 2000), the inhibition should be less successful under load. Therefore, in contrast to AC and DA acceptance, a WM load should tend to decrease high spans' MP and MT acceptance. Thus, for high spans the load effects on AC/DA and MP/MT should interact.

We decided to adopt the complex, executive tapping task from De Neys et al. (2005) as secondary task in the present study. The task was based on the work of Kane and Engle (2000) and Moscovitch (1994). These studies showed that tapping a complex novel tapping sequence (e.g., index finger – ring finger – middle finger – pinkie) put a premium on efficient executive WM functioning. De Neys et al. established that the complex tapping task disrupted counterexample retrieval.

Method

Design. We selected two new groups of high and low spans for the present experiment. Participants were presented with the reasoning task of Experiment 1 while working memory was burdened with an attention-demanding secondary task. The performance of the high and low spans in Experiment 1 served as a baseline for the effect of introducing the WM load.

Participants. A total of 47 first-year psychology students from the University of Leuven, Belgium, participated in the experiment in return for course credit or 5 euro. These participants were identified from the same pool of screened students as the participants in Experiment 1. None of the participants in Experiment 2 had participated in Experiment 1. A total of 23 participants were selected from the top quartile of the Gospan distribution ("high spans"), and 24 were selected from the bottom quartile ("low spans"). Between 99 and 133 days intervened between a given individual's participation in the Gospan screening task and participation in Experiment 2. None of the participants had received any training in formal logic.

Materials. Participants were presented with the same reasoning task with the same procedure as in Experiment 1. A program executed by a second computer collected the finger-tapping data. Participants tapped on the "V", "B", "N", and "M" on the (querty) keyboard of the second computer

Procedure. Each participant was tested individually. The experiment started with a tapping practice phase. The tapping procedure was based on Kane and Engle (2000) and De Neys et al. (2005). Participants were asked to continuously tap the complex index finger – ring finger – middle finger – pinkie sequence with their non-dominant hand. The experimenter demonstrated the tapping sequence and instructed participants to tap it at a "comfortable and consistent" rate.

The choice for the executive tapping task was inspired by the work of De Neys et al. (2005) which suggested that this secondary task had an appropriate difficulty level in a conditional reasoning context. One should note that before reasoners can start retrieving or inhibiting counterexamples they have to read and mentally represent the premises of the inference problem first. Such reading or comprehension processes may also demand WM capacity (e.g., Just, Carpenter, & Keller, 1996). One thus needs a secondary task that interferes with the retrieval process but leaves the more elementary representational processes unaffected. Indeed, if the secondary task were to be so demanding that participants could not read and mentally represent the premises, the findings would clearly not be very informative. De Neys et al. (2005) observed that the complex tapping task decreased the retrieval performance in a counterexample generation task. But participants were still able to generate some counterexamples under load. This would not be possible if complex tapping prevented participants from processing the conditional and factual information of a generation task item. Since the information in De Neys et al.'s generation task items closely resembles the information presented in a conditional inference problem, the secondary task should be well suited for the present experiment.

Participants began with five 30-second practice trials of tapping. Participants always received on-line accuracy feedback: Whenever a wrong finger (key) was tapped the computer emitted an "error" tone (300 ms, low pitch). For the last three practice trials participants also received response time feedback (a 600 ms, high-pitch "speed" tone). Participants received examples of the "error" and "speed" tones, and their different meanings were explained. The computer determined the feedback cut-off times for each participant individually: During the second 30-s practice trial, the computer calculated the mean inter-tap interval and added 150 ms to it. This became the feedback cut-off for the next practice trial. Thus, if any one inter-tap interval was more than 150 ms slower than the established mean from the prior practice trial, the computer immediately emitted a "speed" tone (600 ms, high pitch).

During all tapping practice trials participants were instructed to focus on a fixation cross presented at the centre of the computer screen in front of them. Thus participants could not watch their fingers while tapping.

After the final 30-s practice trial participants received the instructions for the reasoning task. The experimenter explained that the practice tapping speed had to be maintained in the upcoming task. During the reasoning task participants always tapped with accuracy and response time feedback, but the "speed" tapping tone was only given for the final finger (i.e., pinkie). Thus, no tones were emitted if the first three fingers were tapped too slowly.

The reasoning task began with a "BEGIN TAPPING" instruction screen. This "baseline tapping" signal remained onscreen for 15 s, during which participants tapped with response time and accuracy feedback.

Following the 15-s baseline tapping a signal ("NEXT ITEM", presented for 1 s on a blue background) announced the beginning of the reasoning task. The experimenter typed down participants' oral responses on a keyboard connected to the computer running the reasoning task. The computer program also kept track of the mean time elapsed between the presentation of an item and the experimenter's key press. In both Experiment 1 and Experiment 2 we used this time record as a raw measure of participants' mean inference latency. We only recorded the timing to have a general indication of the secondary task impact on the time participants needed for an inference. For this goal the less reliable nature of the present procedure is not problematic. For a specific study on inference latencies in a similar reasoning task we refer to De Neys et al. (2002, Experiment 2).

Following the key press, participants saw the text "NEXT ITEM" (grey letters on black background) for 800 ms after which the next item was presented. After each set of eight inferences the reasoning task was paused. As in Experiment 1, the set order was reversed for half of the participants in each span group (MP/AC, MT/DA, MP/AC, and MT/DA vs MT/DA, MP/AC, MT/DA, and MP/AC) and the order of presentation of the inferences within a set was random.

Participants continuously tapped the sequence until the reasoning task was paused. As in Experiment 1, participants could take as much time for every inference as they wanted. After the break participants started with 15 s baseline tapping, after which the warning signal announced the presentation of the next set of inferences.

Results

Reasoning task. We tested the effects of WM load on inference performance by comparing inference acceptance of the high and low spans in Experiment 1 (no-load) and 2 (load). We calculated participants' mean inference acceptance for every inference type. This resulted in a 2 (load) × 2 (span group) × 4 (inference type) design with load and span group as between-subjects factors and inference type as within-subject factor. The Span group × Inference Type part of the design had already been tested in Experiment 1. Here we focus on the effect of the Load factor.

Figure 3 shows the mean acceptance ratings of the four different inference types for high and low spans under no-load and load conditions. We tested the central hypotheses with planned contrasts. For low spans there was a main effect of WM load. As Figure 3 shows, low spans tended to accept all

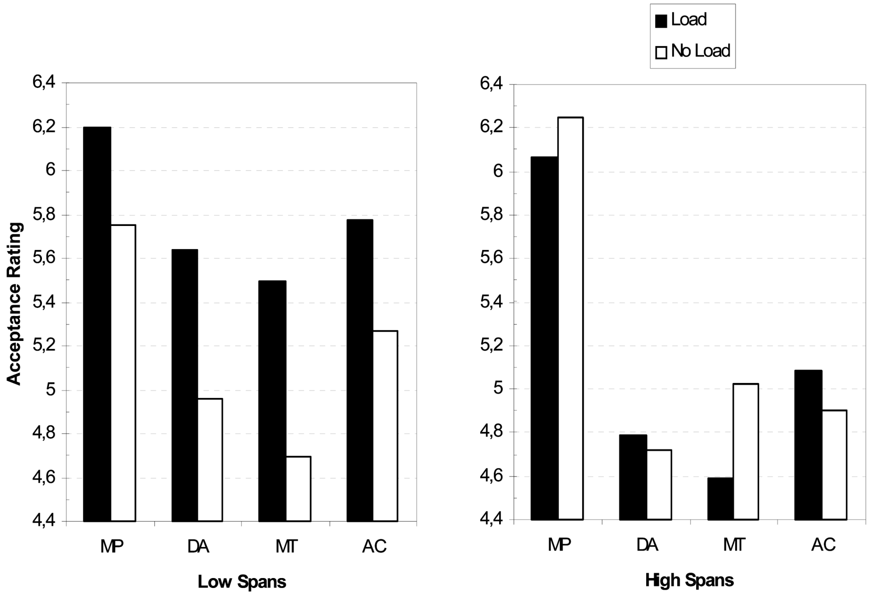


Figure 3. Low and high spans' mean acceptance rating of the four inference types while concurrently tapping the complex finger pattern ("Load") and when there was no secondary task imposed ("No Load"). The rating scale ranged from 1 (very sure cannot draw this conclusion) to 7 (very sure can draw this conclusion).

inferences more when WM was burdened with a secondary task, $F(1, 95) = 7.56$, $MSE = 9.76$, $p < .008$. For high spans this was not the case. Indeed, the effect of the WM load interacted with span group, $F(1, 95) = 4.97$, $MSE = 9.76$, $p < .03$. Consistent with the predictions, for high spans, the load effects on AC and DA on the one hand and MP and MT on the other hand differed. As Figure 3 indicates, the WM load tended to marginally increase AC and DA acceptance whereas it decreased MP and MT acceptance, $F(1, 95) = 3.31$, $MSE = 2.80$, $p < .075$.

The increase of AC, $F(1, 95) < 1$, and DA, $F(1, 95) = 1.77$, $MSE = 5.22$, $p > .18$, acceptance ratings under load seemed to be larger for the low than for the high spans but the effect did not reach significance. Consistent with the inhibition hypothesis, the load impact on MP, $F(1, 95) = 4.84$, $MSE = 2.04$, $p < .035$, and MT, $F(1, 95) = 8.71$, $MSE = 4.32$, $p < .004$, differed for low and high spans. Whereas the low spans accepted MP and MT more under load, burdening WM tended to result in lower MP and MT acceptance for the high spans.

There were no specific expectations about the impact of the WM load on the effect of the number of alternatives and disablers. For completeness we entered the number of alternatives and disablers as within-subject factors in the design. This resulted in a 2 (load) × 2 (span group) × 2 (number of alternatives) × 2 (number of disablers) × 4 (inference type) mixed model factorial. An ANOVA showed that besides a significant Load × Number of Alternatives, $F(1, 95) = 3.96$, $MSE = 9.76$, $p <$ 05, and Load × Span × Number of Alternatives interaction, $F(1, 95) = 6.80$, $MSE = 1.25$, $p < .015$, the load factor did not affect any other factor or interaction of factors in the design (see Appendix Table A2 for the raw data). The significant interactions seemed to be caused by the fact that whereas high spans' inference acceptance under load increased both for conditionals with many and few alternatives, low spans' increase was more pronounced for the many alternative conditionals. The finding makes sense. Under no-load conditions successful retrieval for conditionals with few alternatives will already be rather unlikely for the low spans. Thus, an additional WM load will not make much difference here. Consequently, the load effect for low spans will be stronger for the many alternatives conditionals.

Inference latencies. The time that elapsed between the presentation of an item and the experimenter's key press after participants had evaluated the inference was used as a measure of participants' inference latency. The mean inference latencies were subjected to a 2 (load) × 2 (span group) between-subjects ANOVA. Results showed that both high and low spans needed about 3 s longer to evaluate an inference when concurrently tapping the finger pattern; mean inference latency was 8.42 s ($SD = 1.79$) under no load and 11.44 s ($SD = 3.72$) under load, $F(1, 95) = 27.25$, $MSE = 8.34$, $p < .0001$, with no effect of Span, $F(1, 95) < 1$ or Span × Load interaction, $F(1, 95) < 1$.

Tapping task. For the tapping task we analysed the mean number of correct taps per second across two relevant tapping periods: The "baseline" period presents the average tapping performance during the four 15-s periods that preceded the presentation of each inference set. The "reasoning" period refers to the average tapping performance during the periods between presentation of an inference and the participant's evaluation response. Table 1 shows the results.

A 2 (period, within-subjects) × 2 (span group, between-subjects) ANOVA indicated that tapping performance decreased when participants were reasoning, $F(1, 45) = 4.14$, $MSE = 0.18$, $p < .05$. The decrease in tapping performance was similar for high and low spans, Period × Span interaction, $F(1, 45) < 1$. High spans tended to perform somewhat better than low spans but the effect did not reach significance, $F(1, 45) < 1$. The data thus establish that high and low spans were not differentially trading-off reasoning and tapping performance.

TABLE 1
Mean number and SD of correct taps per second during baseline tapping and during concurrent conditional reasoning

	Period			
	Baseline		*Reasoning*	
Span group	*M*	*SD*	*M*	*SD*
High spans (n = 23)	2.29	0.85	2.09	0.57
Low spans (n = 24)	2.10	0.70	1.96	0.57

Discussion

As predicted, low spans' acceptance of all four inference types increased when working memory was burdened by the tapping task. De Neys et al. (2005) already showed that counterexample retrieval was less efficient when the tapping task demanded WM resources. The present experiment established the link between the decreased search efficiency and inference acceptance. This supports the hypothesis that WM capacity is important for the retrieval of counterexamples in everyday reasoning.

For high spans the load effect interacted with the type of inference. The working memory load tended to increase AC and DA acceptance, as with low spans, but in contrast to the low spans, MP and MT acceptance tended to decrease under load. This pattern corroborates the hypothesis that high spans are using their working memory to inhibit activated disablers. The inhibition process explains the higher MP and MT acceptance for high spans in the absence of a WM load. Since the inhibition is resource demanding, it will be less efficient under load. Therefore, automatically activated disablers that are otherwise inhibited will decrease MP and MT acceptance. Low spans on the other hand allocate their working memory resources at retrieval. When this retrieval becomes less likely under load, MP and MT will be more accepted.

Interestingly, even low spans showed a higher DA and MT acceptance under load. In contrast to AC and MP, the DA and MT inferences involve negations. Therefore, DA and MT are typically labelled more complex inferences than AC and MP (Braine & O'Brien, 1998; Johnson-Laird & Byrne, 1991; Oaksford, Chater, & Larkin, 2000; Schroyens, Schaeken, & d'Ydewalle, 2001). Markovits and Barrouillet (2002), for example, have argued that accepting MT and DA requires that people retrieve an instance of a "complementary class" from memory. Such a complementary class refers to cases in which both the relationship and the events concerned are complementary to those specified in the original conditional (i.e., cases where events different from p are related to not-q). For example, the complementary class for the conditional "If it rains, then the streets get wet" would be composed of related events such as "If the sun shines, the streets are dry" or "If it is only cloudy, the streets are dry".

The fact that low spans' MT and DA acceptance increased under load indicates that they could still retrieve a complementary class example. One could suggest this finding shows that retrieval of a complementary class example is rather automatic, so that it is not affected by the WM load. However, this would conflict with the general contention that processing the negation in the categorical premise of the DA and MT problem requires additional processing capacity (e.g., De Neys et al., 2002b; Johnson-Laird & Byrne, 1991; Oaksford & Chater, 2001; Schroyens et al., 2001; Schroyens, Schaeken, Fias, & d'Ydewalle, 2000). However, Markovits and Barrouillet (2002) suggest an explanation in terms of priority of retrieval instead of demands of retrieval. Markovits and Barrouillet state that when reasoners are confronted with an MT or DA problem they will first search for instances of the complementary class before alternatives or disablers are retrieved (see also Schroyens et al., 2001, for a parameterised model that captures this processing assumption). Consistent with this claim, the increased acceptance might point to the priority of the complementary class search. Because of the priority, the few WM resources that are still available under load would be primarily allocated to retrieval of the complementary example. Therefore, retrieval of a complementary instance could still be successful, but additional disabler or alternative retrieval would become rather unlikely. Consequently, DA and MT acceptance will also increase under load.

To our knowledge the present experiment is one of the first to use a dual task methodology to study reasoning with meaningful, realistic conditionals. The findings present an interesting extension of previous dual task studies with abstract material (e.g., Meiser et al., 2001; Toms et al., 1993). These studies typically found that burdening working memory gave rise to reasoning errors (e.g., a higher rejection of the valid MT under load in Toms et al., 1993). This supported the general contention of reasoning theories like mental logic and mental models that capacity limitations in working memory are a major source of fallacious reasoning. Traditional reasoning theories have focused on the role of WM in the manipulation and storage of the basic mental representations (be they mental rules or mental models) of a reasoning problem. We deliberately selected a secondary task that would not interfere with these basic representational processes. This allowed a more subtle examination of the WM contribution to everyday conditional reasoning. We observed for example that low spans' MP and MT acceptance increased under load. Remember that MP and MT are both logically valid. Thus, working memory limitations actually resulted in a better logical performance here. Despite their larger pool of WM resources, high spans did not show the same effect. To explain these findings, reasoning theories will need to take count of the role of working memory in the retrieval and inhibition of background knowledge.

GENERAL DISCUSSION

The present study examined the role of working memory (WM) capacity in everyday reasoning with causal conditionals. We reported two experiments that are among the first ones in the field to introduce a dual task methodology.[2] Over these studies a stable pattern emerged that established the central role of WM capacity. We focused on two crucial functions: Retrieval and inhibition of counterexamples stored in long-term memory. Experiment 1 compared the performance of a group of people classified as high and low spans in an everyday conditional reasoning task. Successful retrieval of an alternative during conditional reasoning decreases acceptance of the AC and DA inferences, whereas disabler retrieval decreases MP and MT acceptance. Consistent with the predictions, low spans' less efficient alternative retrieval resulted in higher acceptance ratings for the AC and DA inferences. Experiment 2 showed that making counterexample retrieval less likely by burdening WM led to a higher acceptance of every inference type for the low spans. Thereby, the availability of WM resources for counterexample retrieval determines the kind of inferences people draw.

In contrast with AC and DA, MP and MT are logically valid inferences. Based on Stanovich and West's (2000) work on individual differences in reasoning we hypothesised that people of high cognitive ability (for example people in the upper regions of the WM capacity distribution) might have a basic notion of logical validity. Since disablers lead to rejection of MP and MT, this logic notion should conflict with the disabler retrieval. The inhibition hypothesis suggested that high spans use WM resources to inhibit automatically activated disablers. Consistent with this claim, high spans in Experiment 1 showed higher MP and MT acceptance ratings than low spans. If high spans are indeed using WM resources to inhibit disabler retrieval in a conditional reasoning task, the efficiency of the inhibition should be affected by an attention-demanding secondary task. The results of Experiment 2 corroborated the inhibition hypothesis.

The study contributes to recent research that aims to characterise and model the background knowledge search process during (everyday)

---

[2] In a different context, Oaksford, Morris, Grainger, and Williams (1996) have already used dual task methodology to assess the impact of mood states on performance in a hypothesis-testing task with a realistic conditional rule.

conditional reasoning (e.g., De Neys et al., 2002b, 2003a, 2003b, 2005; Markovits & Barrouillet, 2002; Markovits & Quinn, 2002; Quinn & Markovits, 2002; Simoneau & Markovits, 2003). Below we incorporate the present findings in a sketch of the elementary components of the counterexample retrieval process (e.g., De Neys et al., 2005; Markovits & Barrouillet, 2002). Finally, we point to the broader implications of the present findings.

Components of counterexample retrieval

It is assumed that the retrieval process starts with an automatic spreading of activation from an implicit retrieval cue. In a conditional reasoning task this retrieval cue will be the conditional and the categorical premise (e.g., Markovits & Barrouillet, 2002). More precisely, the cue would be the mental representation of these premises stored in working memory. As suggested by many authors, it is assumed that activation will automatically start to spread from the information stored in WM (or "the focus of attention" see Cowan, 1995) to related long-term memory elements (Anderson, 1993; Cowan, 1995; see also Markovits & Barrouillet, 2002). The spreading of activation requires little in the way of executive attention and this component is important for both high and low spans. Subsequently, both span groups will use their WM resources to monitor the automatic retrieval to prevent errors and re-access of previously retrieved counterexamples. Finally, available WM resources will be used for an active generation of cues to access new instances.

The active cue generation will allow a much more efficient retrieval than the passive spreading of activation. Educated adult reasoners (e.g., undergraduate students) will typically use WM resources for an active cue generation. The more resources that are available, the more successful the retrieval will be.

This mechanism can be used for the retrieval of both alternatives and disablers. When it concerns retrieving disablers, however, people from the top of the WM capacity distribution will not use their WM resources for an active cue generation but for an inhibition of automatically activated disablers.

Note that the present model characterises the inhibition process as targeted at preventing the retrieval of disablers. However, this does not imply that the disabler inhibition cannot occur at a later stage in the reasoning process. The inhibition might also contribute to the discarding of a disabler after successful retrieval rather than to the prevention of the retrieval *per se*. We make no strong claims about the exact locus of the inhibition phenomenon. The crucial point is that the inhibition will prevent the disablers having their full impact on the reasoning process.

Currently, reasoning theories mainly focus on a specification of how retrieved counterexamples affect the reasoning process (e.g., Byrne, Espino, & Santamaria, 1999; Johnson-Laird & Byrne, 2002; Oaksford & Chater, 1998; Politzer, in press; Thompson, 2000). Explaining how a reasoning problem is represented and how additional information alters these representations is of course a fundamental component of a reasoning theory. However, the equally important question of how the information is retrieved has not yet been dealt with. The characteristics of the search process itself remain largely unknown (Johnson-Laird & Byrne, 1994; Oaksford & Chater, 2001). The present WM-based specification of the retrieval process will contribute to filling this "knowledge" gap.

Working memory and inhibitory processing

Over the last few years, cognitive research on the nature of inhibitory processing has been subject to increasing interest (e.g., Anderson & Spellman, 1995; Dempster & Brainerd, 1995; Friedman & Miyake, 2004). Although inhibition is generally considered as a key executive function, some studies have observed that executive WM and inhibition are dissociable constructs (e.g., Dempster & Corkill, 1999; Handley et al., 2004). Here it is important to note that Friedman and Miyake (2004) showed that inhibition in itself is not a unitary construct but that different inhibitory functions, with a different relation to executive functioning, can be distinguished. The present study conceived inhibition at the most general level. More detailed future work may further characterise the precise nature of the inhibitory process during reasoning. Nevertheless, the study indicated that whatever the precise nature of the inhibitory processing that is important for everyday conditional reasoning may be, it requires executive WM resources for proper operation.

Dual process theory

The present study has some important implications for dual process theories of reasoning. Remember that the dual process theories distinguish two types of reasoning systems (e.g., Evans & Over, 1996; Sloman, 1996; Stanovich & West, 2000). In general, the first system (System 1) is characterised by automatic, heuristic processing, whereas the second system is characterised by analytic, WM-dependent processing. System – 1 processes would tend towards an automatic contextualisation of a problem with prior knowledge. System – 2 processes on the other hand would decontextualise a problem and allow reasoning according to normative standards. Thereby, System 2 would serve as an override system for the output provided by System 1. Stanovich and West (e.g., 1998, 2000) suggested that System – 2 thinking would be characteristic of those higher in cognitive ability.

The evidence for a disabler inhibition mechanism in everyday conditional reasoning supports Stanovich and West's (e.g., 1998, 2000) notion of a general WM-resource-dependent decontextualisation system: Whenever background knowledge conflicts with normative standards during reasoning, high spans will allocate executive WM resources to block or override the background knowledge impact.

However, the present findings also point to a problematic feature of the framework. Dual process theorists typically conceive the process where background knowledge is accessed as an undemanding, automatic mechanism (e.g., Evans & Over, 1996; Newell & Simon, 1972; Stanovich & West, 2000). Consistent with the critique of De Neys et al. (2005) our dual task experiment showed that retrieval of counterexamples during reasoning (thus a "contextualisation" or System – 1 process) is not purely automatic but depends on WM resources or controlled processing. This questions the general characterisation of System – 1 processing as automatic and effortless. Dual process theories will need to differentiate among different types of background knowledge retrieval processes (see De Neys et al., 2005).

Logic in everyday life

The evidence for high spans' disabler inhibition indicates that people with high cognitive abilities have some basic notion of logical validity or its subsidiary assumptions. High spans apparently adhere to the basic principle in first-order conditional logic that the truth of the antecedent implies the truth of the consequent. This basic logical notion would be the direct cause of the disabler inhibition. This does not mean that high spans intuitively master the propositional calculus or that they would reason by applying formal, logical rules. The present results clearly showed that even high spans' inference acceptance ratings were affected by the number of available alternatives and disablers. If high spans were using an abstract logical "database" to reason, such content mediation would not be expected.

What the data point to is a minimal notion of the logical fact that a conditional rule excludes the possibility that the consequent does not occur when the antecedent occurs. Typically, previous demonstrations of people's ability to adhere to the normative standard of first-order logic have explicitly instructed people to adhere to this norm (Evans & Over, 1996; Markovits et al., 2002; Stanovich & West, 1998; but see Klaczynski, 2001b). Although this shows that some people are able to reason in accordance with the normative standard when they are properly instructed, it does not show that this ability has something to do with reasoning in everyday life (when the norm is not provided explicitly). This critique is not applicable to the present study. Participants were not instructed to reason logically and none of them had received any training whatsoever in formal logic. The present findings thus question the claim of a number of authors that standard logic has no bearing on everyday human reasoning (e.g., Harman, 1986; Oaksford & Chater, 1998). The disabler inhibition phenomenon indicates that it has an impact for some people, albeit in a minimal form.

As in most reasoning studies, we referred to first-order, "textbook" logic as the logical norm (Evans, 2002). Note, however, that despite its widespread use in psychological reasoning studies the status of standard logic as the correct normative system for conditional reasoning is debated (e.g., Edgington, 1995; Evans, 2002; Oaksford & Chater, 1998). Logicians have constructed alternative logical systems with different validity characteristics. When we claim that participants higher in WM capacity manage to inhibit the disabler retrieval, no claims are made about the quality of the reasoning process. It is not claimed that high spans are "better" reasoners. One could argue that low spans adhere to a different normative system where there is simply no need for a disabler inhibition. However, the fact that high spans do tend to adhere to a standard logical norm must at least give pause for thought before discarding the notion of a standard logic-based normative rationality completely (Evans, 2002).

Differences in task interpretation?

One might wonder whether high and low spans' different reasoning performance can simply be attributed to a different task interpretation. High spans interpret the reasoning task as a logical task, whereas low spans perceive it as a task where as much background information as possible needs to be taken into account. We concur that at a meta-level high spans' disabler inhibition may be described as the result of a different, "logical" task interpretation. However, one should be aware that simply proposing a difference in task interpretation as the actual explanation for the present findings will beg the question. Indeed, one would still need to specify why the task interpretation of both groups differs. In the end one would need to assume here that high spans base their interpretation on some elementary logic notion that low spans lack or deem irrelevant. Furthermore, one would need to explain how a different interpretation results in a different performance. Here, one will probably also need to take some kind of inhibition phenomenon into account. Thus, we object to any mere reference to a difference in task interpretation as explanation for the present findings, because it does not explain anything at the psychological processing level.

One could also argue that since the task instructions did not instruct participants to reason logically, high spans are actually erring in that they mistakenly treat an everyday reasoning task as a "logical" one. We believe that such an argument is fallacious and detracts attention from the crucial message. As argued above, it is not clear at present what the correct norm for everyday reasoning should be. It might even be the case that different reasoners adhere to different norms. Therefore we stressed that the fact that high spans seem to adhere to some basic, standard logical principle does not allow one to draw any conclusions about the "quality" or "correctness" of the reasoning process. Along the same lines it makes no sense to label high spans' performance as erroneous by changing the norm. The crucial finding in the present study is that in a situation where people have to evaluate everyday conditional inferences, high spans spontaneously take count of a standard logical principle. The findings indicate that this notion affects performance by means of a WM-dependent inhibition process. Whatever the normative status of standard logic may be, the message is that, to some extent, it is mediating high spans' everyday reasoning. One may wonder why high spans do adhere to this particular norm. Although such considerations fall outside the scope of the present study, they are intriguing and deserve further research. Our point is that one should refrain from wondering why high spans adhere to a *faulty* norm. Couching the discussion in these evaluative terms is not warranted and may prevent further progress.

CONCLUSION

The present study indicates that WM capacity plays a crucial role when people reason with familiar, causal conditionals for which they have access to relevant background knowledge. By the mediation of counterexample retrieval and inhibition, WM capacity has a profound impact on the inferences people draw in daily life reasoning.

Manuscript received 25 June 2004
Revised manuscript received 19 November 2004

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of prepositional knowledge. *Journal of Experimental Psychology: General*, *130*, 544–570.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *102*, 68–100.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49A*, 5–28.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Barrouillet, P. (1996). Transitive inferences from set-inclusion relations and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1408–1422.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning*, *5*, 289–302.

Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

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

Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, *40*, 347–373.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking and Reasoning*, *9*, 203–244.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, *123*, 354–373.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

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.

Dempster, F. N., & Brainerd, C. J. (1995). *Interference and inhibition in cognition*. San Diego, CA: Academic press.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica*, *101*, 395–416.

De Neys, W., d'Ydewalle, G., Schaeken, W., & Vos, G. (2002a). A Dutch, computerized, and group administrable adaptation of the operation span test. *Psychologica Belgica*, *42*, 177–190.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002b). 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 counterexample retrieval for causal conditionals. *Thinking and Reasoning*, *11*, 123–150.

Edgington, D. (1995). On conditionals. *Mind*, *104*, 235–329.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

Engle, R. W., & Oransky, N. (1999). Multi-store versus dynamic models of temporary storage in memory. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 515–555). Cambridge, MA: MIT Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*, 978–996.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, *105*, 331–351.

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology*, *11*, 473–498.

Gilinsky, A., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging*, *9*, 356–371.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage*, *12*, 504–514.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, *87*, B11–B22.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

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.

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

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of judgment and decision making during childhood and adolescence. *Current Directions in Psychological Science*, *11*, 145–149.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology*, *35*, 904–911.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Cambridge University Press.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1994). Models, necessity, and the search for counterexamples. *Behavioral and Brain Sciences*, *17*, 775–778.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, *103*, 773–780.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, *130*, 169–183.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 336–358.

Klaczynski, P. A. (2001a). Analytic and heuristic processing influences on adolescent reasoning and decision making. *Child Development*, *72*, 844–861.

Klaczynski, P. A. (2001b). Framing effects on adolescent task representation, analytic and heuristic processing, and decision making: Implications for the normative/descriptive gap. *Applied Developmental Psychology*, *22*, 289–309.

Klauer, K. C., Stegmaier, R., Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking and Reasoning*, *3*, 9–47.

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity. *Intelligence*, *14*, 389–433.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1118–1133.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trend in Cognitive Sciences*, *6*, 299–305.

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

Markovits, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking and Reasoning*, *8*, 97–107.

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

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica*, *106*, 303–327.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, MA: Cambridge University Press.

Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and medial temporal cortex. *Neuropsychology*, *8*, 524–534.

Moscovitch, M. (1995). Models of consciousness and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1341–1356). Cambridge, MA: MIT Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, UK: Psychology Press.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*, 349–357.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 883–899.

Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 476–492.

Politzer, G. (in press). Premise interpretation in conditional reasoning. In D. Hardman & L. Macchi (Eds.), *Reasoning and decision making*. New York: John Wiley.

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition*, *68*, B93–B101.

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

Radvansky, G. A. (1999). Memory retrieval and suppression: The inhibition of situation models. *Journal of Experimental Psychology: General*, *128*, 563–579.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT press.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*, 211–227.

Rumain, B., Connell, J., & Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults. *Developmental Psychology*, *19*, 471–481.

Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2001). The processing of negations in conditional reasoning: A meta-analytic study in mental model and/or mental logic theory. *Thinking and Reasoning*, *7*, 121–172.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle, G. (2000). Heuristic and analytic processes in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1713–1734.

Shallice, T., & Burgess, P. W. (1993). Supervisory control of thought and action. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 171–187). Oxford: Oxford University Press.

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.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*, 161–188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–726.

Staudenmayer, H. (1975). Understanding conditional reasoning with meaningful propositions. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. (pp.55–79). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Süss, H., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability and a little bit more. *Intelligence*, *30*, 261–288.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory and Cognition*, *22*, 742–758.

Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition*, *76*, 209–268.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *Quarterly Journal of Experimental Psychology*, *46A*, 679–699.

Wickelgren, I. (1997). Getting a grasp on working memory. *Science*, *275*, 1580–1582.

APPENDIX

TABLE A1

The conditionals for the reasoning task of Experiment 1 and 2

If John studies hard, then he does well on the test.
If Bart's food goes down the wrong way, then he has to cough.
If the trigger is pulled, then the gun fires.
If the intensity of light increases, then the pupils of the eyes grow smaller.
If Jenny turns on the air conditioner, then she feels cool.
If water is poured on the campfire, then the fire goes out.
If the ignition key is turned, then the car starts.
If Tom grasps the glass with his bare hands, then his fingerprints are on it.

TABLE A2
Acceptance ratings

	Low span								*High span*							
	MP		*DA*		*MT*		*AC*		*MP*		*DA*		*MT*		*AC*	
Counterexample group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
No Load (Experiment 1)																
Few alternatives	6.03	.71	5.72	.78	5.01	1.13	6.13	.58	6.38	.64	5.40	1.46	5.16	1.21	5.67	1.25
Many alternatives	5.47	.97	4.19	1.11	4.38	.96	4.41	1.07	6.12	.78	4.03	1.58	4.88	1.17	4.13	1.64
Few disablers	6.06	.78	4.66	1.00	5.01	1.01	4.90	.90	6.39	.65	4.43	1.46	5.20	1.21	4.71	1.29
Many disablers	5.44	.91	5.25	.93	4.38	1.23	5.63	.75	6.11	.78	5.00	1.50	4.85	1.17	5.10	1.45
Load (Experiment 2)																
Few alternatives	6.29	.55	6.07	.83	5.68	1.00	6.24	.86	6.28	.74	5.40	1.39	4.83	1.36	5.88	1.38
Many alternatives	6.10	.67	5.21	1.35	5.31	1.07	5.30	1.14	5.85	.95	4.17	1.26	4.35	1.35	4.28	1.52
Few disablers	6.34	.49	5.47	1.36	5.70	1.04	5.64	1.03	6.23	.87	4.44	1.33	4.90	1.23	4.82	1.31
Many disablers	6.05	.63	5.81	.81	5.29	1.03	5.91	1.00	5.85	.73	4.96	1.30	4.32	1.33	5.04	1.40

High and low spans' mean acceptance rating of the four inference types as a function of the number of alternatives and disablers when no secondary task was imposed ("No Load") and when concurrently tapping the complex finger pattern ("Load").