Working Memory Span and Everyday Conditional Reasoning: A Trend Analysis

Wim De Neys (Wim.Deneys@psy.kuleuven.ac.be)
Walter Schaeken (Walter.Schaeken@psy.kuleuven.ac.be)
Géry d'Ydewalle (Géry.dYdewalle@psy.kuleuven.ac.be)
Department of Psychology, K.U.Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

Abstract

This study presents evidence for the role of working memory (WM) capacity in the retrieval and inhibition of counterexamples (alternatives and disablers) during everyday conditional reasoning. A total of 292 university students were given a measure of WM-capacity and a reasoning task with everyday, causal conditionals. Results show that the acceptance ratings of the logically valid Modus Ponens and Modus Tollens inferences follow a quadratic, U-shaped trend in function of WM-capacity, while acceptance ratings of the logically invalid Affirmation of the Consequent and Denial of the Antecedent inferences follow a negative linear trend. Findings support the claim that participants highest in WM-capacity spontaneously inhibit the disabler retrieval process during everyday reasoning.

Introduction

Suppose that you are given the following conditional information: "*If you pull the trigger, then the gun fires*". Next, you are also told that "*The gun fired*". Would you draw the conclusion "*The trigger was pulled*" on the basis of this information? Likewise, suppose that in addition to the conditional "*If the match is struck, then it lights*" you are told that "*The match is struck*". Would you then draw the conclusion "*The match will light*"? In the present study we examine how the extent to which people do draw such conditional conclusions is related to their working memory capacity.

Research on conditional reasoning typically focuses on peoples performance on four kinds of arguments: Modus Ponens (MP, e.g., 'If p then q, p therefore q), Affirmation of the Consequent (AC, e.g., 'If p then q, q therefore p'), Modus Tollens (MT, e.g., 'If p then q, not q, therefore not p'), and Denial of the Antecedent (DA, e.g., 'If p then q, not p, therefore not q). The first (p) part of the conditional is called the antecedent and the second (q) part is called the consequent. In standard logic, MP and MT are considered valid inferences, while AC and DA inferences are considered fallacies. Thus, standard logic would tell you to reject the AC conclusion "*The trigger was pulled*" in the first introductory example and to accept the MP conclusion "*The match will light*" in the second example.

When people reason with realistic, content-rich conditionals, the inferences they draw are affected by prior knowledge about the conditional. For example, when one thinks about the fact that the match might be wet, one will be reluctant to infer that the match will light when it is struck. Since in daily-life people typically reason with content-rich conditionals it is crucial for cognitive reasoning theories to address the impact of background knowledge on inference acceptance (Johnson-Laird & Byrne, 1994, 2002; Oaksford & Chater, 1998). In the last few years this issue has become one of the main foci of interest in the literature (e.g., Byrne, 1989; Byrne, Espino, & Santamaria, 1999; Cummins, 1995; De Neys, Schaeken, & d'Ydewalle, 2002; Markovits & Barrouillet, 2002; Thompson, 1994, 2000; Rumain, Connell, & Braine, 1983).

At least two important kinds of information stored in long-term memory have been shown to affect the inference acceptance: Alternative causes and disabling conditions. An alternative cause (alternative) is a condition, besides the original antecedent, that can bring about the consequent (e.g., lighting the match with another fire in the introductory example). A disabling condition (disabler) is a condition that prevents the antecedent from bringing about the consequent (e.g., the match being wet in the introductory example).

It is well established that when reasoners retrieve an alternative during conditional reasoning they will tend to reject the fallacious AC and DA inferences (e.g., Rumain et al., 1983; Cummins, 1995; Janveau-Brennan & Markovits, 1999; Quinn & Markovits, 1998). Retrieval of a disabler results in rejection of the valid MP and MT inferences (e.g., Byrne, 1989; Cummins, 1995; Thompson, 1994; De Neys, Schaeken, & d'Ydewalle, 2002, 2003b, in press). Further on, we adopt Byrne's (1989) terminology and refer to alternatives and disablers as counterexamples.

Recently, peoples conditional reasoning performance with realistic, causal, conditionals has been related to working memory (WM) capacity (De Neys, Schaeken, & d'Ydewalle, 2003a; Markovits, Doyon, & Simoneau, 2002). De Neys et al. showed that WM affects everyday conditional reasoning by mediation of the counterexample retrieval process. In a first experiment, participants were asked to generate as much counterexamples as possible in limited-time for a set of conditionals. Results indicated that participants higher in WM-capacity were better at retrieving alternatives and disablers. 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: The larger the WM-resource pool is, the more resources can be allocated to the search, and the more efficient the search will be. In a further experiment, a group of low and high spans (participants in the bottom and top quartile of first-year psychology students' WM-capacity distribution, respectively) were tested in an everyday conditional reasoning task. Because of the more efficient alternative retrieval De Neys et al. reasoned that high spans (vs. low spans) should be more likely to reject the fallacious AC and DA 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, 2002; Stanovich & West, 2000). It is generally assumed that all people have a basic "contextualisation" tendency to search stored background knowledge (e.g., counterexamples) associated with the reasoning problem. However, recent individual difference studies (e.g., Stanovich & West, 1998, 2000) indicate that at least people of high cognitive (WM) capacity also appear to have a logical, "decontextualisation" tendency: A basic ability to put background knowledge aside when it conflicts with the logical standards. Remember that in standard logic MP and MT are valid inferences. Since disabler retrieval will result in the rejection of MP and MT, a basic validity notion should conflict with the disabler retrieval process. De Neys et al. reasoned that high spans would therefore use their WM-resources for an active inhibition of the automatic disabler search. Despite the better intrinsic retrieval capacities, this inhibition process should result in higher MP and MT acceptance ratings for the high (vs. low) spans. Results of the study supported the predictions.

In an additional dual-task study the hypotheses were further confirmed. The basic assumption stated that lower spans allocate WM-capacity to the disabler retrieval, while high spans allocate WM-capacity primarily to the retrieval inhibition. Consistent with the hypothesis, the dual-task study showed that a less efficient disabler retrieval under WM-load resulted in higher MP acceptance ratings under load (vs. no load) for low spans, while the less efficient inhibition resulted in lower MP ratings under load (vs. no load) for high spans.

Thus, there is evidence for the claim that high spans are inhibiting the disabler retrieval process during conditional reasoning. Inhibition of cognitive processes deemed inappropriate is indeed one of the key executive working memory functions (e.g., Baddeley, 1996; Levy & Anderson, 2002; Miyake & Shah, 1999). The basis of the inhibition during reasoning would be high spans' minimal notion of the standards of first-order logic. High spans would adhere to the logical principle that the truth of the antecedent implies the truth of the consequent. This principle excludes the possibility that the consequent does not occur when the antecedent occurs (i.e., a disabler).

Note that it is assumed that the inhibition would only occur for people highest in WM-capacity. If this assumption is correct, it follows that people with medium WM-capacities should show the lowest MP and MT acceptance. Indeed, on one hand medium spans (vs. high spans) should not inhibit disabler retrieval. On the other hand, medium spans will have a more efficient counterexample retrieval than low spans because they can allocate more resources to the search. Thus, the disabler retrieval during conditional reasoning should be most successful for medium spans. Consequently, it is expected that the MP and MT acceptance ratings follow a U-shaped curve in function of WM-capacity: Due to the limited resources, people lowest in WM-capacity will not be very successful at retrieving disablers and should therefore show rather high levels of MP and MT acceptance. Because of the more efficient disabler retrieval, MP and MT acceptance should decrease for the medium spans. Because of the disabler inhibition, MP and MT acceptance ratings should increase again for reasoners higher in WM-capacity.

Since retrieving alternatives results in the rejection of AC/DA inferences and accepting AC/DA is erroneous in standard logic there is no conflict between a basic logical notion and the retrieval of alternatives. Moreover, De Neys et al. (2003a) already showed that the extent to which high spans accept AC and DA is mediated by the alternative retrieval. Therefore, the higher WM-capacity is, the more efficient the alternative retrieval will be, and the less AC and DA should be accepted. Contrary to MP and MT, AC and DA acceptance ratings should therefore follow a negative linear trend in function of WM-capacity. These predictions were tested in the present experiment.

The present study will complete the previous work of De Neys et al. (2003a) and Markovits et al. (2002). De Neys et al. compared only a group of participants from the bottom and top quartile of the WM-capacity distribution. Thus , the crucial 'medium' span group was missing. Markovits et al. looked only at the linear correlation between WM-capacity and inference acceptance. Given the a priori expectation of a quadratic, U-shaped function for MP and MT, a mere linear correlation analysis is not informative here.

With reference to the everyday nature of our reasoning task we want to stress two crucial distinctions with more formal, deductive reasoning tasks. First, the study adopts meaningful, causal conditionals so that participants have access to relevant background knowledge about alternatives and disablers. In addition, contrary to most conditional reasoning studies participants are not specifically instructed to reason logically (e.g., participants are not instructed to accept the premises as always true or to derive only conclusions that follow necessarily). Although participants are still situated in a laboratory setting this should allow and encourage people to reason as they would in everyday life (Cummins, 1995; see also Galotti, 1989).

Method

Participants

A total of 292 first-year psychology students from the University of Leuven (Belgium) participated in the experiment in return for course credit. None of the students had had any training in formal logic.

Material

Working memory task. Participants' working memory capacity was measured using a version of the Operation span task (Ospan, La Pointe & Engle, 1990) adapted for group testing (Gospan, for details see De Neys, d'Ydewalle, Schaeken, & Vos, 2002). In the Ospan-task participants solve series of simple mathematical operations while attempting to remember a list of unrelated words The main adaptation in the Gospan is that the operation from an operation-word pair is first presented separately on screen (e.g., 'IS (4/2) – 1 = 5 ?'). Participants read the operation silently and press a key to indicate whether the answer is correct or not. Responses and response latencies are recorded. After the participant has typed down the response, the corresponding word (e.g., 'BALL') from the operation-word string is presented for 800 ms. As in the standard Ospan three sets of each length (from two to six operation-word pairs) are tested and set size varies in the same randomly chosen order for each participant. The Gospan-score is the sum of the recalled words for all sets recalled completely and in correct order.

Participants were tested in groups of 21 to 48 at the same time. Participants who made more than 15% math errors or whose mean operation response latencies deviated by more than 2.5 standard deviations of the sample mean were discarded (participants already in the bottom quartile of the Gospan-score distribution were not discarded based on the latency criterion). De Neys, d'Ydewalle et al. (2002) reported an internal reliability coefficient alpha of .74 for the Gospan. The corrected correlation between standard Ospan and Gospan-score reached .70.

Reasoning task. Sixteen causal conditionals from the generation study of De Neys et al. (2002) and Verschueren, De Neys, Schaeken, and d'Ydewalle (2002) were selected for the reasoning task. The number of possible counterexamples of the selected conditionals varied systematically. The number of counterexamples constituted a 2 (few/many) x 2 (alternatives/disablers) design with four conditionals per cell. Two conditionals in each cell were embedded both in the MP and DA inferences, while the other two were embedded both in the AC and MT inferences. This produced a total of 32 inferences for each participant to evaluate. The experiment was run on computer. The following item format was used:

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

Each argument was presented on screen together with a 7-point rating scale ranging from 1 (*Very certain that I cannot draw this conclusion*) to 7 (*Very certain that I can draw this conclusion*) with 4 representing *can't tell*. Participants typed down the number that best reflected their evaluation of the conclusion.

Procedure

All participants started with the Gospan task. After a 5 min break the reasoning task was presented. Reasoning task instructions were presented 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. Participants used the keypad to type the number reflecting their decision. The experimental session was preceded by one practice trial.

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) x 2 (alternatives/disablers) design that the conditionals constituted. Participants received the sets in the order MP/AC, MT/DA, MT/DA, MP/AC. Eight conditionals were used for the first two sets and the remaining eight conditionals for the inferences in the last two sets. Thus, the inferences in the first and last two sets were always based on different conditionals. After two sets (i.e., 16 inferences) were evaluated, item presentation was paused until participants decided to continue.

As we pointed out, the task instructions did not mention to accept the premises as true or to endorse conclusions that follow necessarily. Instead participants were told they could evaluate the conclusions by the criteria they personally judged relevant.

Results

Ten participants were discarded (about 3.5 % of the sample) because they did not meet the operation correctness or latency requirements of the WM-task. The remaining 282 participants were split in five span groups based on the quintile boundaries of the

Gospan-score distribution. Mean Gospan-score for the successive span groups was 18.21 (SD = 4.1), 26.97 (SD = 1.69), 32.26 (SD = .99), 37.56 (SD = 1.35), and 47.32 (SD = 5.07) for the fifth and top quintile group.

Each participant evaluated eight inferences of the same inference type. The mean of these eight observations was calculated. These means were subjected to a 5 (span group) x 4 (inference type) mixed model ANOVA with span group as between-subjects factor and inference type as within subjects factor. The effects of inference type were analyzed with multivariate ANOVA tests.

There was a main effect of span group, $F(1, 277) = 3.31$, $MSE = 2.35$, $p < .015$, and inference type, Rao $R(3, 275) = 180.16$, $p < .0001$, and the two factors also interacted, Rao $R(12, 727) = 1.83$, $p < .045$. More specifically, the impact of span group on AC and DA differed from the impact of span group on MP and MT, $F(1, 277) = 2.72$, $MSE = .33$, $p < .03$. Figure 1 shows the results.

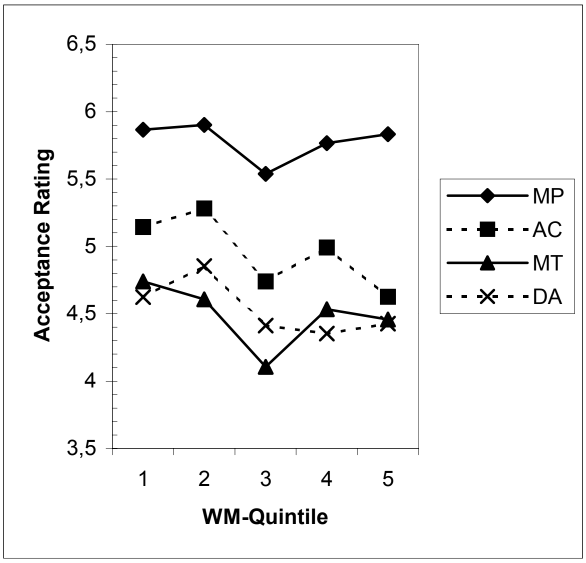

Figure 1. Mean acceptance rating of the four inference types for participants in the successive span groups. WM-Quintile 1 stand for the bottom quintile. The rating scale ranged from 1 (very sure cannot draw this conclusion) to 7 (very sure can draw this conclusion).

As figure 1 indicates the MP and MT acceptance ratings indeed followed a U-shaped curve with people in the middle quintile ('medium' spans) showing the lowest acceptance levels. The trend analysis established the quadratic nature of this trend for MP, $F(1, 277) = 3.13$, $MSE = .55$, $p < .08$, and MT, $F(1, 277) = 3.3$, $MSE = 1.33$, $p < .075$. Neither on MP, $F(1, 277) < 1$, nor on MT, $F(1, 277) = 1.88$, $MSE = 1.33$, $p > .17$, the quadratic trend was mediated by a linear trend.

As expected, AC and DA acceptance showed a different pattern with higher span groups tending towards lower acceptance ratings. The trend analysis established the linear trend in the AC, $F(1, 277) = 11.97$, $MSE = .9$, $p < .001$, and DA ratings, $F(1, 277) = 3.95$, $MSE = 1.25$, $p < .05$, while a quadratic trend was absent, both $F(1, 277) < 1$.

In order to make sure there was indeed a linear trend without quadratic trend mediation on AC/DA on the one hand, and a quadratic trend on MP/MT on the other hand, we ran an additional analysis on the combined MP/MT and DA/AC acceptance ratings. This more powerful analysis confirmed the findings [linear trend, $F(1, 277) = 8.87$, $MSE = 1.71$, $p < .005$, without quadratic trend mediation, $F(1, 277) < 1$, on AC/DA, and a quadratic trend, $F(1, 277) = 4.46$, $MSE = 1.30$, $p < .04$, without linear trend, $F(1, 277) = 1.67$, $MSE = 1.30$, $p > .2$, on the combined MP/MT ratings]. The analysis also confirmed that there were significant AC/DA by MP/MT interactions for the quadratic, $F(1, 277) = 5.16$, $MSE = .66$, $p < .025$, and linear trend, $F(1, 277) = 4.42$, $MSE = .66$, $p < .04$.

Discussion

The present results establish that MP and MT acceptance ratings follow a quadratic, U-shaped trend in function of WM-capacity, while the AC and DA acceptance rating patterns follow a negative linear trend. MP and MT acceptance ratings were lowest for the 'medium spans' in the middle quintile of the WM-capacity distribution. This pattern is precisely what one would expect if the disabler inhibition during conditional reasoning occurs only for people highest in WM-capacity. In the absence of a disabler inhibition process, medium spans can allocate more WM-resources to the disabler retrieval than lower spans. Because high spans will inhibit the disabler retrieval, the search will be most successful for the medium spans. Consequently, the middle group shows the lowest MP and MT acceptance ratings.

It is assumed that the basis of high spans' disabler inhibition is a minimal notion of the logical principle that the truth of the antecedent implies the truth of the consequent. While this notion conflicts with the possibility that the consequent does not occur when the antecedent occurs (i.e., a disabler), it does not conflict with the possibility that the consequent occurs in the absence of the antecedent (i.e., an alternative). Thus, the process where alternatives are retrieved from long-term memory should not be inhibited. Successful retrieval of alternatives results in the rejection of the AC and DA inferences. Since higher WM-capacity allows a more efficient retrieval, higher WM-resources will lead to lower AC and DA acceptance ratings in the reasoning task. Consistent with this framework, contrary to MP and MT, AC and DA acceptance linearly decreased for the successive span groups. These findings demonstrate the crucial role of WM in the mediation of the counterexample retrieval during everyday conditional reasoning.

In an earlier study, De Neys et al. (2002, Experiment 3) measured the efficiency of the disabler retrieval process in a generation task and linked this with performance in a conditional reasoning task. It was observed that a more efficient disabler retrieval process resulted in lower MP and MT acceptance ratings. Later, De Neys et al. (2003a) found that although high spans were better at retrieving disablers, they nevertheless accepted MP and MT more than low spans. These findings seem to contradict each other. However, De Neys et al. (2002) only measured retrieval efficiency for 40 randomly selected participants (WM-capacity was not assessed). De Neys et al. (2003a) specifically selected participants from the top and bottom quartile of first-year psychology students' WM-capacity distribution. As De Neys et al. (2003a) argued, the top WM-spans were probably small in number in the sample of 40 participants. The present data show that in that case higher WM-capacity (and thus better retrieval) should indeed result in lower MP and MT acceptance. Hence, the trend analysis allows us to reconcile both studies.

The quadratic trend in the MP and MT acceptance ratings implies that, in general, people with the highest and lowest WM-capacity will accept MP and MT to the same extent. Of course, when the performance of a selected group of high and low spans is contrasted (e.g., see De Neys et al., 2003a), ratings may differ depending on the relative position of the participants in both groups on the WM-capacity distribution. Bluntly put, if the low group tends somewhat more towards the middle than the high group, the high group will also show somewhat higher ratings and vica verca. The trend analysis provides the bigger picture here. This underlines the importance of examining the inference acceptance patterns over the whole WM-capacity distribution.

Markovits et al. (2002) calculated linear correlations between reasoning performance and WM-capacity in a conditional reasoning task with the kind of causal conditionals adopted in the present study. Consistent with the present findings Markovits et al. found significant correlations for AC and DA: Higher WM-capacity resulted in a more frequent rejection of the AC ($r = .21$) and DA ($r = .20$) inferences. However, they also reported a smaller but significant linear correlation for MP. In the present study there was clearly no sign of a linear trend on MP. We suspect that the difference with the present findings lies in the task instructions. Markovits et al. were interested in the study of deductive reasoning with realistic conditionals. Therefore participants were explicitly instructed to reason logically (i.e., participants had to assume that the premises were always true). We are interested in the reasoning process people use in everyday life whatever the nature of this process may be (e.g., deductive or probabilistic). Therefore, in our studies people can evaluate the conclusions by the criteria they personally judge relevant. Our data show that reasoners in the top levels of the WM-distribution are able to inhibit the disabler retrieval. The findings thereby indicate that high spans spontaneously (without being instructed) adhere to a standard logical norm in their reasoning. Markovits' data might suggest that when the norm is explicitly presented even medium span will tend to adhere to it. There is some evidence (George, 1995; Vadeboncoeur & Markovits, 1999) suggesting that stressing the logical nature of the task reduces the MP rejection. Thus, when properly instructed even medium spans might to some extent block the disabler retrieval and consequently show a boost in MP acceptance. Because of the larger WM-capacity pool the resource demanding inhibition will still be more successful for the high spans. Since disabler retrieval is already unlikely for participants lowest in WM-capacity, the instructions should only have minimal impact on low spans' acceptance ratings. Therefore, one could expect a more (positive) linear trend on MP and MT acceptance with standard instructions. The higher WM-capacity is, the more successful the inhibition will be, and the more MP and MT will tend to be accepted.

An important final remark concerns the status of standard, first-order logic as a normative reasoning system. As in most reasoning studies, we always refer to first-order, 'textbook' logic as the logical norm (Evans, 2002). However, note that despite its widespread use in psychological reasoning studies the status of standard logic as the correct normative system for conditional reasoning is heavily debated (e.g., Edgington, 1995; Evans, 2002; Oaksford & Chater, 1998). Logicians have constructed alternative logical systems with different validity characteristics. Van Lambalgen and Stenning (2002), for example, worked out a nonmonotonic logic where rejecting MP and MT in the light of possible disablers is considered valid. 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 medium and low spans adhere to a different normative system where there is simply no need for a disabler inhibition. However, the disabler inhibition phenomenon does suggest that cognitively skilled reasoners have a basic notion of the standard logical principle that a conditional utterance excludes the possibility that the consequent does not occur when the antecedent occurs. Therefore, the findings do question the opposite claim that standard logic would have no bearing on peoples everyday life reasoning (e.g., Oaksford & Chater, 1998). For some people, to some extent, it has.

Acknowledgements

We thank Kristien Dieussaert and Sara Verbrugghe for their assistance in conducting the experiment. Preparation of the manuscript was supported by grants from the Fund for Scientific Research-Flanders (FWO).

References

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

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.

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

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

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). *Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples*. (Tech. Rep. No. 295). Leuven, Belgium: University of Leuven, Laboratory of Experimental Psychology.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology, 15,* 161-176.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (in press). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*.

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

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

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

George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86*, 93-111.

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

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.

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. *Trends 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.

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

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

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.

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*.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology, 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.

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

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

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

Van Lambalgen, M., & Stenning, K. (2002). *'Suppression effect' and nonmonotonicity*. Unpublished manuscript.

Verschueren, N., De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Working memory capacity and the nature of generated counterexamples. *Proceedings of the Annual Conference of the Cognitive Science Society, 24,* 914-919.