

# Everyday conditional reasoning: A working memory–dependent tradeoff between counterexample and likelihood use

NIKI VERSCHUEREN, WALTER SCHAEKEN, and GERY D'YDEWALLE  
*University of Leuven, Leuven, Belgium*

Considerable evidence has revealed that working memory capacity is an important determinant of conditional reasoning performance. There are two accounts describing the conditional inference process, the probabilistic and the mental models accounts. According to the mental models account, reasoners retrieve and integrate counterexample information to attain a conclusion. According to the probabilistic account, reasoners base their judgments on probabilistic information. It can be assumed that reasoning according to the mental models process would require more working memory resources than would solving the inference on the basis of probabilistic information. By means of a verbal report study, we showed that participants with low working memory capacity more often use probabilistic information, whereas participants with higher working memory capacity are more likely to use counterexample information. Working memory capacity thus not only relates to reasoning performance, it also determines which process reasoners will engage in.

Working memory is conceived of as a cognitive system with a constrained capacity that limits performance in various tasks. In its most general conceptualization, it is seen as controlled attention (see, e.g., Kane, Bleckley, Conway, & Engle, 2001), whereas more functional approaches refer to aspects of moment-to-moment monitoring of processing activities and the maintenance, manipulation, and recombination of information as its main executive tasks (see, e.g., Miyake & Shah, 1999; Oberauer, Süß, Wilhelm, & Wittmann, 2003). Research has shown that working memory capacity is a crucial determinant of performance in a variety of cognitively complex tasks (see, e.g., Kyllonen & Christal, 1990). Studies on abstract syllogistic reasoning problems have yielded strong evidence for a link between working memory capacity and reasoning performance: The higher the working memory capacity of participants, the more these participants give normative answers (Copeland & Radvansky, 2004; Gilhooly, Logie, & Wynn, 1999). The present study focuses on the role of working memory capacity in conditional reasoning.

The involvement of working memory capacity in conditional reasoning has already been demonstrated for abstract or content-lean inference making: Toms, Morris, and Ward (1993) found disruptive effects of secondary tasks, and Barrouillet and Lecas (1999) concluded on the

basis of a developmental study that there is a significant correlation between working memory span (measured by counting span) and successful reasoning performance. However, reasoning on abstract conditionals differs from reasoning with everyday conditionals (see, e.g., Bindra, Clarke, & Shultz, 1980; Venet & Markovits, 2001). Markovits, Doyon, and Simoneau (2002) looked at the relation between individual differences in (visual and verbal) working memory and reasoning performance on both concrete and abstract conditionals. For both types of conditionals, they found some significant correlations between visual and verbal working memory capacity and reasoning performance, but the correlational patterns were different for abstract versus concrete reasoning problems. The authors did not explain this difference but concluded that it is necessary to consider content as an important process-related variable.

One explanation for the differences in performance on abstract and concrete reasoning tasks is that in the case of everyday conditionals, participants spontaneously take their background knowledge into account (for a review, see Politzer & Bourmaud, 2002). Although this contextualization process is characteristic of everyday and common-sense reasoning, the information integration process is omitted when research focuses on abstract conditionals. Consequently, the working memory resources needed to perform abstract reasoning tasks may be different from the resources needed to perform concrete reasoning tasks. The retrieval of stored long-term memory information relevant to the task at hand and the coherent combination of the retrieved elements with the elements that are currently actively attended to is considered to be a key role of working memory (Baddeley & Logie, 1999; Rosen & Engle, 1997).

---

This research was supported by funding from the F.W.O.-Vlaanderen (Belgium). We thank Wim De Neys for providing us with the Gospan scores. We also feel greatly indebted to Joost De Cock for co-rating the verbal report answers. Correspondence relating to this article should be sent to N. Verschueren, University of Leuven, Laboratory of Experimental Psychology, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: niki.verschueren@psy.kuleuven.ac.be).

Given the abundant evidence of the strong and ubiquitous impact of available background knowledge on reasoning performance (see, e.g., Cummins, Lubart, Alksnis, & Rist, 1991; Thompson, 2000), it is crucial to investigate working memory involvement in common-sense reasoning.

The observed relationship between working memory capacity and reasoning performance is generally explained by assuming that reasoning requires an analytic thinking process that taps working memory resources. However, there is evidence that reasoning problems are often solved by heuristic processes that do not rely as heavily on working memory capacity (see, e.g., Evans, 1993; Klaczynski, 2001). Research on syllogistic reasoning has revealed that the relative importance of heuristic and analytic reasoning processes relates to individual differences in working memory capacity (Copeland & Radvansky, 2004; Gilhooly et al., 1999): Highly skilled participants use more demanding reasoning processes, whereas reasoners with a more limited working memory capacity rely on less demanding processes. The present article investigates whether this relationship between working memory capacity and the relative importance of heuristic and analytic reasoning processes can also be found in everyday conditional reasoning.

Most reasoning studies investigating the role of working memory have assumed that conditional inferences are solved by use of a reasoning mechanism that can be described by the mental models account (Johnson-Laird & Byrne, 1991, 2002; Markovits & Barrouillet, 2002). In these studies, reasoning by a mental models mechanism is conceived of as a highly demanding reasoning process (Barrouillet & Lecas, 1999; Klauer, Stegmaier, & Meiser, 1997; Markovits et al., 2002). Some studies of working memory involvement in propositional reasoning have assumed that there are alternative, heuristic processes that blur the role of working memory in reasoning (see, e.g., Klauer et al., 1997; Meiser, Klauer, & Naumer, 2001), but as yet no specific information has been provided on how this heuristic reasoning mechanism functions. Other recent research has shown that conditional reasoning performance is also determined by probabilistic information (George, 1997; Liu, Lo, & Wu, 1996; Oaksford, Chater, & Larkin, 2000). We will argue that a reasoning process relying on probabilistic information can be considered a heuristic, nondemanding mechanism, whereas the mental models theory describes a rather analytic reasoning mechanism. Both reasoning mechanisms—the mental models and probabilistic accounts—have already been contrasted elsewhere (Verschueren, Schaeken, & d'Ydewalle, 2003a, in press; Weidenfeld & Oberauer, 2003). Below, we will briefly explain the two reasoning processes for solving conditional inferences, with a particular focus on their working memory demands.

### Two Reasoning Mechanisms and Their Working Memory Demands

Everyday causal conditional reasoning problems consist of a causal “if-then” rule and a categorical premise that

refers to either the antecedent or the consequent proposition. Two examples of conditional reasoning problems are:

#### *Modus Ponens (MP)*

If you study hard, then you will pass the exam.  
Someone studies hard.  
Will he pass the exam?

#### *Affirmation of the Consequent (AC)*

If you bump your head, then you will get a headache.  
Someone has a headache.  
Did this person bump his head?

In standard logic, MP is considered a valid inference, whereas AC is fallacious. This valid/invalid distinction does not apply to everyday reasoning. We will not score the answers as correct or incorrect but will instead focus on the contextualization process: Which information do participants use for making the inference?

**Mental models account.** According to the mental models theory, reasoners form a representation of the content of all relevant situations and draw a conclusion that is congruent with the models they actively represent (Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992). Markovits and colleagues adapted the mental models theory to reasoning with everyday conditionals. In their account, the automatic retrieval of relevant information assumes a central place (Markovits & Barrouillet, 2002; Markovits, Fleury, Quinn, & Venet, 1998). We will illustrate the reasoning process with the sentence *If you pull a cat's tail, then the cat gets angry*. Reasoners start their reasoning process by representing the content of the causal rule as a possible situation, namely *pull tail* → *angry cat*. On the basis of this single model, reasoners can already provide answers for MP and AC: For both MP (*Someone pulls a cat's tail. Does the cat get angry?*) and AC (*A cat is angry. Did someone pull its tail?*) they will answer “yes.” The active consideration of the problem content will lead to an automatic activation of additional relevant information from long-term memory; elements that are sufficiently activated will be represented and taken into account. For causal reasoning, the relevant information consists of alternative causes and disabling conditions, together referred to as counterexamples. For MP, the categorical premise will trigger the retrieval of disablers. A disabler is a situation that prevents the cause from producing the effect—for example, *the cat is asleep* or *the cat is sedated*. When reasoners retrieve at least one disabler, they will not accept the default MP conclusion. In the case of AC, a search for alternative causes will be triggered. An alternative cause is a cause that can produce the effect regardless of the presence of the given cause—for instance, *take away the cat's food* or *buy a dog*. When reasoners retrieve an alternative cause, the representation of the initial model and the alternative cause informs them that there is more than one possible cause for the effect. As a result, they conclude that the default AC conclusion does not necessarily follow.

In this reasoning process, several of the processing factors rely on working memory capacity:

1. *Representation and maintenance.* Both the problem content and all models of relevant situations that are taken into account have to be represented. The larger the number of models that participants have to represent and maintain, the heavier the load on working memory. This assumption is crucial to the mental models theory and has already been substantiated with experimental results (Barrouillet & Lecas, 1999). Reasoners are aware of the counterexamples they retrieve, and they can verbalize this information without effort (Verschuere, Schaeken, De Neys, & d'Ydewalle, 2004). This evidence indicates that the counterexamples are not only represented in working memory space but are even actively attended to.

2. *Retrieval.* It has been found that retrieval efficiency suffers from dual-task loads (Rosen & Engle, 1997), indicating that working memory is involved in the retrieval of information needed for active processing. The Rosen and Engle study sustains the idea that there are two modes of long-term memory retrieval: (1) A retrieval mode that relies on automatic activation guided by semantic association; with this retrieval mechanism, the central control component is responsible for the monitoring of errors and for the controlled suppression of previously recalled items. (2) A retrieval mechanism that relies on active cue generation; this cue generation functions as part of a controlled, strategic search for relevant knowledge and requires additional working memory resources. It has been found that this controlled search is more prominent for reasoners with high working memory capacity than for those with low working memory capacity (Rosen & Engle, 1997). De Neys (2003) showed that the strategic, working memory-dependent retrieval of counterexample information plays a considerable role in causal conditional reasoning.

3. *Inhibition.* Inhibition is considered as a central executive function of working memory (see, e.g., Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Participants with a high working memory capacity are more efficient in inhibiting the automatic retrieval of counterexamples when this retrieved information conflicts with normative standards (De Neys, 2003; Markovits & Barrouillet, 2002). Some reasoners with a high working memory capacity will allocate additional working memory resources to block or inhibit the retrieval of disablers for the logically—not necessarily the psychologically—valid MP inferences.

4. *Premise integration.* When reasoners have a representation of both the conditional sentence and at least one counterexample, they have to integrate this information in order to see that there are two different conclusions for the same problem. There is wide agreement that information manipulation and integration are crucial tasks of working memory (Miyake & Shah, 1999). The integration of mental models is given as a key example of the coordinative function of working memory (Oberauer, Süss, Schulze, Wilhelm, & Wittmann, 2000).

When summing up these different processing aspects, it is straightforward to conclude that the mental models reasoning process taps heavily on working memory resources.

**Probabilistic account.** Research has revealed that probabilistic properties also affect the inference process (see, e.g., George, 1997; Stevenson & Over, 1995). These findings suggest that everyday causal reasoning cannot be fully understood in terms of the discrete counterexample retrieval and accounting processes set forth by the mental models theory (Evans, Handley, & Over, 2003; Oberauer & Wilhelm, 2003). Conditional inferences can also be solved by taking likelihood<sup>1</sup> information into account. In this case, some have argued that reasoners consult their long-term memory to retrieve the likelihood of the conclusion given the categorical premise (Liu et al., 1996; Oaksford et al., 2000). For solving the MP problem *If you pull a cat's tail, then the cat gets angry. Someone pulls a cat's tail. Does the cat get angry?* reasoners activate all situations in which a cat's tail has been pulled and infer the overall likelihood that the cat got angry. When this likelihood is high, reasoners will accept the default MP conclusion. The endorsement of MP is thus directly proportional to the likelihood of the effect following from the cause, or  $L(\text{effect}|\text{cause})$ . The reasoning process for AC is conceived analogously with MP. In the case of AC, reasoners search for relevant situations in which the effect occurs and induce  $L(\text{cause}|\text{effect})$ , the likelihood that the cause preceded the occurring effect. In order to solve the AC problem *If you pull a cat's tail, then the cat gets angry. A cat got angry. Did someone pull its tail?*, reasoners consider all situations in which a cat got angry and verify the likelihood that its anger was caused by pulling its tail. The AC inference acceptance rates are directly proportional to the derived likelihood estimation.

We will now discuss the role of working memory for the different aspects of this second reasoning process in analogy with the four processing aspects described above.

1. *Representation and maintenance.* The propositions that are brought together in working memory space are the problem content and the resulting likelihood estimate. Regarding the representation of the background knowledge needed to infer a likelihood estimate, it is assumed that reasoners are not consciously aware of the range of situations they consult. Only the end product—the likelihood estimate itself—is available to consciousness. Most models of working memory assert that there are close connections between the content of working memory, attention, and awareness, although the three entities do not coincide (for a review, see Miyake & Shah, 1999). The observation that the situations used for inferring likelihood estimations are not available to consciousness leads us to assume that these situations are not explicitly represented in working memory but rather are briefly accessed in order to produce an automatic likelihood estimate.

2. *Retrieval.* There is no active controlled search process based on cue generation. However, in its supervising function, the central control component should be involved in engaging and monitoring the automatic retrieval process leading to the likelihood estimation. Note, however, that this monitoring does not have to be as profound as when individual counterexamples are to be retrieved and taken into account.

3. *Inhibition.* The low-level retrieval of situations necessary to infer a likelihood estimate is assumed to occur in a heuristic, automatic manner. Although the central control component can control for errors, there is no room for active blocking or inhibition of automatically retrieved information.

4. *Premise integration.* There is no need for premise integration, because the likelihood estimate is based on all relevant situations at a time. The final conclusion directly mirrors the obtained likelihood estimate.

On the basis of this functional analysis, it can be concluded that the reasoning process relying on likelihood estimates does not tap as heavily on working memory resources as does the reasoning process that takes counterexamples into account.

Verschueren et al. (in press) showed that the process that takes counterexamples into account can be considered a slow, analytical reasoning mechanism, whereas the process that bases conclusions on likelihood estimates can be seen as both fast and heuristic. They suggested that the working memory capacity of the individual reasoner could be an important determinant of the relative importance of the two reasoning processes. We assume that when sufficient working memory resources are available, reasoners prefer to reason analytically, because conclusions based on counterexample information are considered to be more informative and accurate than those based on likelihood estimates (see Moskowitz, Skurnik, & Galinsky, 1999, for a related account).

### **The Thinking-Aloud Setup in Research on Conditional Reasoning**

Conditional reasoning performance is classically investigated by looking at inference acceptance rates: Participants are asked either to indicate on an 11-point acceptance scale whether they accept the inference (see, e.g., Cummins et al., 1991) or to tell whether the stated conclusion necessarily follows (see, e.g., Markovits et al., 2002). These acceptance rates can be linked to likelihood estimations as well as to counterexample information, but the discrete answers do not inform us about which information the participants actually used for deciding whether to accept the conclusion. In order to monitor which information participants use for making inferences, we opted for a thinking-aloud setup.

Solving a conditional reasoning task in a thinking-aloud setup differs from solving inferences in a standard deduction task. Whereas classic deduction tasks do not provide information on the background knowledge used for everyday reasoning, the present thinking-aloud study

will enable us to monitor the information reasoners use. It can be assumed that the information retrieval observed in thinking-aloud experiments also governs performance in standard deduction tasks, but as these tasks do not provide information about which information is used, this assumption is as yet mere speculation, and generalizations of thinking-aloud results to performance in these tasks should be made with caution.

Conditional reasoning entails a conscious consideration of premises and answers and lends itself well to explicit verbalization. Especially for tasks involving higher level cognitive processes operating on verbal types of information—such as propositional reasoning—verbal reports can tap the conscious and easily verbalizable processes (Payne, 1994). Concurrent verbalization does not affect the basic performance or the gross structure of thought processes when the verbal reports entail verbalizations of covert but actively attended-to verbal codings (Ericsson & Simon, 1984). Two sets of experimental observations have shown that verbalization does not disrupt the normal course of the reasoning process. Verschueren, Schaeken, and d'Ydewalle (2003b) ran a detailed analysis of a similar thinking-aloud study and found that the robust effects of counterexample availability and inference types reported in literature on conditional inferences were corroborated. Dieussaert, Schaeken, Schroyens, and d'Ydewalle (1999) used both a thinking-aloud setup and a discrete answer evaluation format to investigate strategies on inference suppression by additional premise presentation. They concluded that “the specific procedure of thinking-aloud has not altered the reasoning process of the participants” (p. 228). Other studies using verbal reports have already revealed important representational and semantic information about different reasoning tasks (see Ford, 1995, on syllogistic reasoning and Stenning & Van Lambalgen, 2001, on the selection task). The present study continues on this line of research and investigates working memory involvement on a level that is different from that of previous approaches to conditional reasoning: It combines the maximal informativeness of a thinking-aloud setup with an experimental selection of different working memory span groups.

### **The Present Experiment**

In this experiment, participants with low, medium, and high working memory capacity were selected and asked to think aloud when solving causal inference problems. Working memory capacity was measured by use of an operation span test (La Pointe & Engle, 1990; Turner & Engle, 1989). This test is often used in working memory research for investigating the relation between working memory capacity and performance on complex cognitive tasks (see, e.g., Engle, Tuholski, Laughlin, & Conway, 1999); it is taken to measure general executive functioning and control capacity (storage and processing). The reasoning task was performed in a thinking-aloud setting. The free-production answers were examined for the use of likelihood and counterexample information,

and in cases in which both information types were combined, we checked the order in which the information was used.

Our main prediction was that when working memory capacity increases, reasoners would more often use counterexample information. The underlying ideas are that reasoning by use of counterexample information is an analytical resource-demanding reasoning process, and that only reasoners with sufficient working memory capacity have the resources to reason according to this mechanism. Reasoners with lower working memory capacity will be more likely to base their answers on likelihood information.

## METHOD

### Participants

Fifty-four students from the University of Leuven participated in the experiment. These participants were recruited from a larger pool of 292 first-year psychology students whose working memory span was tested by use of the Gospan test ( $M = 32.46$ ,  $SD = 10.90$ ). In order to investigate the differential effects of working memory span, we selected 18 participants from three different span groups. The low-span group had a mean Gospan score of 18.5 ( $SD = 4.38$ ,  $Min = 9$ ,  $Max = 24$ ), the medium group had a mean of 32 ( $SD = 1.46$ ,  $Min = 29$ ,  $Max = 34$ ), and the high-span group had a mean score of 47.9 ( $SD = 4.22$ ,  $Min = 42$ ,  $Max = 60$ ). The participants received course credit or were paid for their participation (8 euro/hour). All participants were native Dutch speakers.

### Procedure

**Operation span test.** We screened the participants for working memory capacity by use of the Gospan test (for details, see De Neys, d'Ydewalle, Schaeken, & Vos, 2002). This is a Dutch computerized version of the classic operation span test (La Pointe & Engle, 1990) adapted for group administration. In this operation span test, a mathematical operation was first presented on screen (e.g., "Does  $(4/2) - 1 = 5$ ?"). The participants read this problem silently and had to decide whether the equation was correct. They pressed a key to answer "yes" (equation holds) or "no" (equation does not hold). Responses and response latencies were recorded. Following this answer selection, a word was presented for 800 msec (e.g., *carrot*). After a series of two to six operation-word pairs, the participants were asked to recall the list of two to six unrelated words. Each participant was presented with three sets of each length. Set size varied in the same random order for all participants. The operation span score was the sum of the recalled words for all sets recalled correctly (completely and in the order of presentation). Span scores ranged from 0 to 60. Participants who made more than 15% math errors or whose mean latencies deviated by more than 2.5 standard deviations were excluded.

**Reasoning task.** The participants in the reasoning task were tested individually. The inference problems were presented on paper, with every reasoning problem on a separate sheet. The participants progressed through the sheets in a self-paced manner. The reasoning problems were as follows (AC):

If you water a plant well, then the plant will stay green.  
A plant stays green.  
Was this plant watered well?

The participants were asked to read the premises aloud. They were not forced to accept or reject the inference, and they could provide whatever information they considered relevant to answer the question in an everyday setting. The participants were told that their an-

swers would be recorded and that they were instructed to think aloud. The experimenter interfered only when the participants answered "yes" without further information, by asking "Can you motivate your answer?" This was done to stimulate participants to think aloud. The subsequent motivation that the participant provided on the specific trial was not included in the analyses.

### Design

All participants solved 32 problems; the 16 sentences were presented in AC as well as in MP argument form. Participants first solved 16 problems in one argument form, and after a short break, they solved the 16 problems in the second argument form. The order of argument forms was counterbalanced: In each span group, 9 participants received MP and then AC, the other 9 received AC first and then MP. Overall, it took participants from 15 to 30 min to complete the task.

### Materials

A total of 16 causal sentences were selected based on previous research. They were chosen to reflect a representative sample of causal rules.

1. If water is heated to the boiling point, then the water will boil.
2. If you phone someone, then his telephone will ring.
3. If a predator is hungry, then it will search for prey.
4. If you have a Y-chromosome, then you will be masculine.
5. If you depress the brake, then the car will slow down.
6. If a dog has fleas, then it will scratch itself.
7. If you eat lots of salt, then you will get thirsty.
8. If you sit in the cold draft, then you will catch a cold.
9. If a girl has sexual intercourse, then she will be pregnant.
10. If you water a plant well, then it will stay green.
11. If the light switch is flipped, then the light will go on.
12. If there are advertisements for a product, then the sales records will rise.
13. If you drink lots of Coke, then you will gain weight.
14. If the roads are slippery, then there will be car accidents.
15. If someone is nearsighted, then he will wear glasses.
16. If you go on a diet, then you will lose weight.

The number of available counterexamples for all sentences was measured in a previous study (Verschuere et al., 2004) by use of the standard generation task (used by, e.g., Cummins et al., 1991). There were 4 sentences with few disablers and few alternatives (1–4), 4 sentences with few disablers and many alternatives (5–8), 4 sentences with many disablers and few alternatives (9–12), and 4 sentences with many disablers and many alternatives (13–16). The 16 sentences occurred in the same random order for all participants. The four types of sentences were well distributed.

### Rating

The answers of the participants varied greatly in length and detail.<sup>2</sup> Two independent raters were asked to rate the report answers. For the rating task, each trial got a unique numerical code that made it impossible to see to which condition the individual answers belonged. In a first phase, both raters rated all 1,728 answers on four dimensions: (1) inference acceptance [0/1]; (2) reference to information about probability/frequency [0/1]; (3) reference to counterexample information [0/1]; and (4) if participants referred to counterexample information, they scored the number of specific counterexamples used. The scoring instructions can be found in Appendix A. Next, the two sets of ratings were brought together to verify possible disagreements. Answers for which the ratings disagreed on at least one of the first three dimensions were tagged, and both raters were asked to rerate those answer trials. No information was retained about the dimension(s) where discrepancies were lo-

calized. For AC trials, there were 102 answers (11.81%) rescored; for MP trials, 196 answers were rescored (22.69%). In many instances, simply asking the raters to rescore the selected answers punctually reduced the discordance. After rerating, discrepancies remained for 13 AC answers (1.5%) and for 21 MP answers (2.43%). These trials were excluded from the analysis. For the rating of the number of counterexamples, the interrater reliability was checked by calculating the linear correlation between the numbers scored by the two raters. For MP, the correlation was .935 ( $N = 864, p < .001$ ), and for AC, the correlation was .962 ( $N = 864, p < .001$ ). For further analyses, we used the mean of the two counts.

## RESULTS

The verbal reports were transcribed from audiotape to text files. Each answer was scored for inference acceptance and the use of counterexamples and likelihood information. Appendix B gives an overview of the different types of answers. Overall, participants referred to likelihood information in 19.8% of the trials, they referred to counterexamples in 62.6% of the trials, and the rate of inference acceptance was 25.8%.

With inference acceptance, there was no information available about the background knowledge that participants might have used to reach their conclusions. It is possible that reasoners did not rely on background knowledge and just satisfied the question. However, it is also possible that they reached conclusions after consulting relevant background knowledge, and in that case infer-

ence acceptance could relate either to a likelihood estimation that is high enough to accept the conclusion or to a lack of available counterexamples. The two latter types of inference acceptance can be recognized when we look at the subsequent motivation participants gave to "yes" trials (see Examples 1 and 2 of Appendix B). Example 3 illustrates that some participants may not have taken disablers or alternatives into account because they believed that the given conclusion held in normal and default situations. This is because answers that were given after a request for motivation may be considered as retrospective rationalizations instead of concurrent answer registrations. We will analyze inference acceptance rates without speculating on the underlying reasoning process.

We ran our first analysis on trials where participants either accepted the inference, used a likelihood estimation, or referred to a counterexample. A second analysis examined the data of combination trials. Finally, we looked at the number of counterexamples that participants used.

### Effects of Operation Span

Examples 4 to 9 of Appendix B illustrate answers in which participants referred to either likelihood or counterexample information. We verified whether the different span conditions related to differences in the information used, by running an analysis of variance (ANOVA) with a 3 (span group, between subjects)  $\times$  2 (inference type, within subjects)  $\times$  3 (answer type, within subjects)

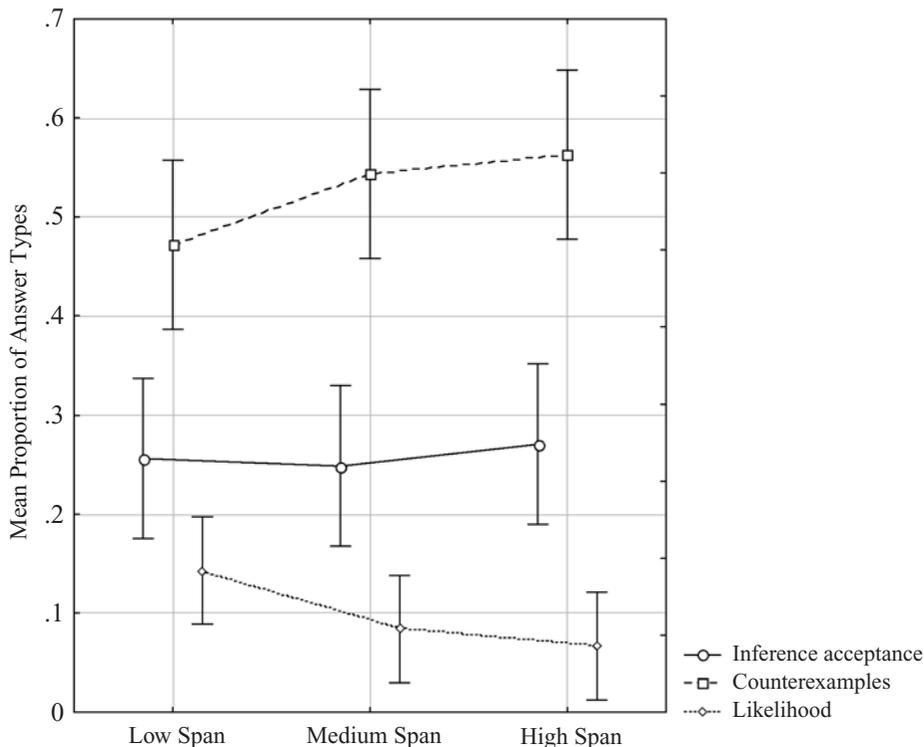


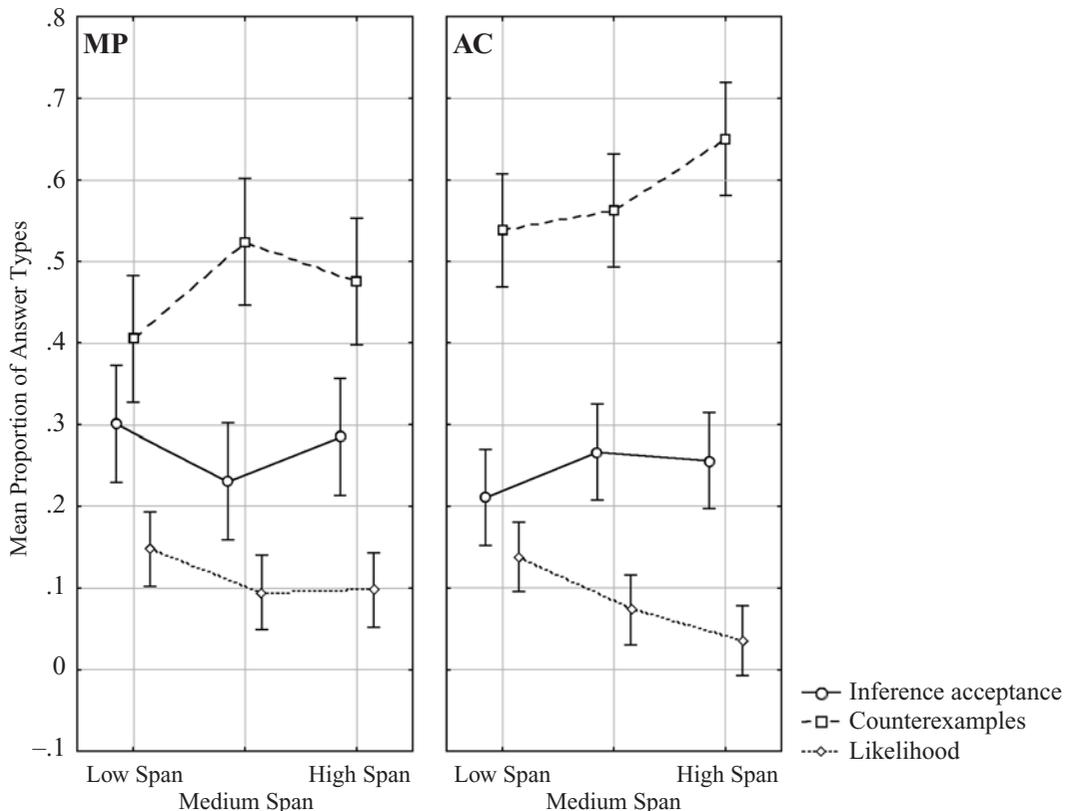
Figure 1. Interaction between the operation span groups (low, medium, high) and the mean proportion of the observed answer types.

design. All effects involving the three-leveled dependent variable of answer type were then analyzed with multivariate statistics.

There was a main effect of answer type [ $F(2,50) = 167.33, p < .001$ ; Wilks's  $\lambda = .130$ ]: There were more counterexample trials (52.6%) than inference acceptance (25.8%) or likelihood trials (9.8%). Two factors influenced the relative occurrence of the different answer types. First, we observed a marginally significant interaction between answer type and operation span group [ $F(4,100) = 2.399, p = .055$ ; Wilks's  $\lambda = .657$ ]. Figure 1 illustrates this interaction (vertical bars denote 95% confidence intervals). Planned comparisons revealed that there was a significant positive linear relation between working memory capacity and the proportion of trials in which participants referred to counterexamples [ $F(1,51) = 4.598, MS_e = 0.150, p < .05$ ;  $M_{low} = .472, M_{med} = .543, M_{high} = .563$ ]. These results corroborate the assumption that participants with a higher working memory capacity are more inclined to rely on counterexample information than are reasoners with a lower working memory capacity. Participants with lower working memory capacity more often used likelihood information [ $F(1,51) = 8.011, MS_e = 0.1054, p < .05$ ;  $M_{low} = .143, M_{med} = .084, M_{high} = .067$ ]. There was no effect of working memory capacity on the inference acceptance rates.

Second, the relative occurrence of the different answer types interacted with the reasoning form [ $F(2,50) = 13.067, p < .01$ ; Wilks's  $\lambda = .657$ ]. Planned comparisons showed no significant difference in inference acceptance rates. There were more likelihood trials for MP (11.3%) than for AC (8.2%) [ $F(1,51) = 8.011, MS_e = 0.105, p < .01$ ]. There were more counterexamples mentioned for AC (58.3%) than for MP (46.8%) [ $F(1,51) = 22.799, MS_e = 0.359, p < .001$ ]. The difference in counterexample use between AC and MP corroborates the findings of Verschueren et al. (2003b). There are two possible explanations for this effect. First, it is possible that finding disablers is harder than finding alternatives (Markovits & Barrouillet, 2002; see Byrne, Espino, & Santamaria, 1999, for a related account). Second, some participants may be more inclined to retrieve alternatives than disablers. The following interaction provides support for this second possibility.

There was a significant interaction of span group, reasoning form, and answer type [ $F(4,100) = 2.874, p < .05$ ; Wilks's  $\lambda = .805$ ]. Figure 2 illustrates this interaction. Following Markovits and Barrouillet (2002), we can assume that some highly skilled reasoners tend to selectively inhibit MP because they recognize that the logical validity of MP renders disabling information irrelevant. In this case, we should find that the relation between counter-



**Figure 2.** Mean proportion of inference acceptance, counterexample, and likelihood trials for the different span groups, separately for MP and AC.

example use and working memory capacity is different for MP than for AC: For AC, we should observe a positive linear relation, and for MP, we should observe an inverse, *U*-shaped relation between working memory capacity and counterexample use. For AC, we indeed found a significant positive linear trend [ $F(1,51) = 5.291, MS_e = 0.113, p < .05$ ], whereas for MP, we found a marginally significant inverse *U*-shaped relation between working memory capacity and the proportion of counterexample trials [ $F(1,51) = 3.145, MS_e = 0.085, p = .082$ ]. For AC, there was a significant negative linear trend between likelihood use and working memory capacity [ $F(1,51) = 11.638, MS_e = 0.095, p < .01$ ]. For MP, there was no significant relation between working memory capacity and likelihood use. We did not find any significant trends between working memory capacity and inference acceptance.

The differential effect of span group on the type of information participants used delivers support for the findings of De Neys (2003), who observed a positive linear relation between AC acceptance measured on a standard deduction task and working memory capacity, whereas the relation between working memory span and MP followed an inverse *U* shape. This result was explained by assuming that an increase in working memory capacity leads to a more efficient counterexample retrieval process but that for MP, some highly skilled participants acknowledge the logical irrelevance of disablers, leading to a selective inhibition of this type of background knowledge. The presently observed trends for counterexample use are in line with this assumption, but the reports of the high-span group did not reveal specific information to substantiate the inhibition hypothesis.

Overall, the present results support our thesis that reasoning by use of counterexamples requires more working memory capacity than reasoning based on likelihood information. The fact that the effects of working memory capacity on the use of counterexamples and likelihood information are complementary indicates that a limited working memory capacity does not simply lead to an imperfect reasoning performance based on counterexample retrieval, but that there is a computationally motivated tradeoff. When working memory resources are limited, participants rely on a less demanding, heuristic reasoning process.

### Answers Combining Counterexample and Likelihood Information

The mere occurrence of combination answers reveals that both reasoning processes do not occur in isolation from each other. Analyzing the combination answers can reveal how both processes are related. The centerpiece of the heuristic/analytic interaction mechanism is the override principle (see, e.g., Evans & Over, 1996; Stanovich & West, 2000): It is generally assumed that the heuristic process starts automatically and by default and that the analytical process can subsequently override or amend the heuristic output. If this heuristic/analytic conceptualization holds, we should find that the likelihood-counterexample

combination is more prevalent than the counterexample-likelihood combination.

To verify how the two processes are combined, the combination answers were first exhaustively divided into three groups: (1) answers in which a likelihood estimate was followed by counterexample information; (2) answers in which the counterexample information was followed by a likelihood estimate; and (3) answers in which the reasoners first mentioned a likelihood estimate, then provided counterexample information, and concluded with another likelihood estimate. Examples can be found in Appendix B (10–18).

For each participant, we calculated the proportion of each of the three types of combination answers. We ran an ANOVA with a 3 (span group, between subjects)  $\times$  2 (reasoning form, within subjects)  $\times$  3 (type of combination answer, within subjects) design. There was no main effect of operation span, nor were there any significant interaction effects involving operation span (all *p* levels were  $> .7$ ).

There was a main effect of combination type [ $F(2,50) = 23.457, p < .001; M_{\text{Like-CE}} = .075, M_{\text{CE-Like}} = .015, M_{\text{Like-CE-Like}} = .011; \text{Wilks's } \lambda = .516$ ]: There were more answers in which a likelihood estimate was followed by counterexample information than vice versa [ $F(1,51) = 40.986, MS_e = 0.196, p < .001$ ], a result that supports the general claim that the slower analytical process relying on counterexample information can complement or override the heuristic process relying on likelihood information (see Verschueren et al., in press).

Second, there was a main effect of inference form: Significantly more combination answers were given for MP than for AC [ $F(1,51) = 12.436, p < .001; M_{\text{MP}} = .044, M_{\text{AC}} = .023$ ]. For MP, there were significantly more combination answers of the first kind—likelihood followed by counterexample information—than there were for AC [ $F(1,51) = 8.407, MS_e = 11.34, p < .001; M_{\text{MP}} = .096, M_{\text{AC}} = .054$ ]. This observation can be explained as follows: If retrieving disablers is indeed more difficult than retrieving alternatives (Markovits & Barrouillet, 2002), reasoners would need more time to process the analytically obtained information, thereby leaving time for the heuristic conclusion to be explicitly produced. Similar comparisons for the other two types of combination answers did not yield significant results.

In sum, the analysis of the combination answers revealed two things: First, we found support for the idea that the heuristic likelihood conclusion is amended or overridden by a conclusion based on counterexample information. Second, the other two types of combination answers showed that in a minority of trials, the counterexample information can also be followed by a likelihood estimate. Inspection of the verbal reports revealed that this likelihood estimate did not always concern L(effect|cause) or L(cause|effect). Reasoners also referred to other types of probabilistic information. We briefly list below some alternative uses of likelihood information.

First, Examples 14 to 16 illustrate that likelihood information can also be used as a sort of summary of the stated counterexamples. This finding supports the idea that reasoners summarize multiple specific counterexamples with a general model (proposed by Markovits & Barrouillet, 2002). Second, likelihood information can be used to indicate that there is still a fair possibility that the default conclusion holds, even though there are counterexamples. The likelihood information is then used to nuance the counterclaim. Examples 19 and 20 illustrate this kind of likelihood use. Third, the likelihood information can be added to indicate the probability of the counterexample mentioned (Example 21) or a probability that the conclusion holds given that certain preconditions (positively formulated counterexamples) are absent (Example 22). This use of likelihood estimations can be linked to the idea that mental models are annotated with a relative weight (Johnson-Laird, 1994). Example 21 can be seen as a counterexample model annotated with likelihood information, whereas Example 22 reveals that the initial model can be annotated with a certain likelihood even when counterexamples have been retrieved (see Example 13 for a variation). A fourth case is when participants refer to a likelihood estimation to indicate the probability that the conclusion holds when counterexamples are present. In Example 23, the participant mentions an alternative cause for which the probability of the effect is higher than it is for the given cause.

These four types of likelihood use indicate that more complex probabilistic processes are involved in everyday reasoning than simply drawing a conclusion in proportion to the heuristically derived likelihood that the conclusion follows given the premises. The present data, however, are too sparse to allow a more thorough analysis of these more complex answers.

### Number of Counterexamples

To check whether there is a positive relationship between the number of counterexamples used and the reasoners' working memory capacity, we ran an ANOVA with a 2 (inference type, within subjects)  $\times$  3 (span group, between subjects) design. The effects parallel those found for the counterexample trials. We found a marginally significant effect of working memory span on the mean number of counterexamples used [ $F(2,51) = 2.680, p = .078; M_{\text{low}} = .728, M_{\text{med}} = .939, M_{\text{high}} = .910$ ]. There was also a marginally significant interaction between span group and inference type [ $F(2,51) = 2.57, MS_e = 14.11, p = .086$ ]: For MP, there was a significant inverse *U*-shaped relation between counterexample use and operation span [ $F(1,51) = 4.364, MS_e = 0.573, p < .05$ ], whereas for AC we observed a positive linear trend between span group and the number of counterexamples [ $F(1,51) = 3.860, MS_e = 0.361, p = .055$ ]. These effects can be attributed to the observed differences in the proportions of counterexample trials. When the analysis was confined to only those trials in which participants

referred to counterexamples, there were no significant effects of working memory capacity on the number of counterexamples.

## DISCUSSION

The present results are a clear indication that for reasoning with everyday conditionals, participants spontaneously take relevant background knowledge into account. The mental models account and recent probabilistic theories can both account for a substantial part of the present observations, and both models remain valuable because they explain a significant proportion of the thinking-aloud results. At the same time, each account by itself delivers an incomplete description of the contextualization process by focusing mainly on one type of background information. The statements that participants reason by taking counterexamples into account or by making inferences probabilistically are both true, but both cover only part of the data. The present results show that reasoners use (and even combine) both types of background information. Thus, in terms of the recent polarization between mental models and probabilistic accounts (see, e.g., Evans et al., 2003; Oberauer & Wilhelm, 2003), we support an intermediate position in which both accounts complement each other (for another alternative account, see Schroyens & Schaeken, 2004). The dual-process idea provides a fruitful way of combining the two accounts in a single framework for contextualized reasoning, while at the same time retaining the key aspects of both accounts. Because we found combination answers where likelihood estimations followed counterexample information, we also conclude that the use of probabilistic information is not confined to the simple heuristic inferences described by the probabilistic reasoning process.

The observed effect of span group on reasoning performance corroborates the idea that working memory is involved in conditional reasoning with everyday causal sentences. Working memory capacity not only relates to the obtained reasoning results but also to the actual reasoning process used for attaining the conclusion. We found that highly skilled reasoners take counterexample information into account and reason by a process that is best described by the mental models account for everyday causal reasoning. On the other hand, reasoners with a lower working memory span rely more often on a simple probabilistic reasoning mechanism. The present results support the idea that the two reasoning processes can be framed in a dual-process perspective (see also Verschueren et al., 2003a, in press). The probabilistic account can be considered to be a heuristic process, a type that is generally assumed to impose only a minor load on working memory; analytical processes—such as the one described by the mental models account—are generally believed to draw more heavily on working memory resources. We found that the reasoning process relying on counterexample information is more often used by highly

skilled participants, which suggests that the analytical reasoning mechanism based on counterexample information can rightly be considered a high-demand reasoning process. The presently observed working memory related differences in the types of information used provide only indirect information about the working memory demands of the processes involved. Specific dual-task studies may reveal whether there is a difference in the actual working memory demands of the two processes.

The present observations suggest that the relation between individual differences and the reasoning performance observed in standard deduction tasks is most likely mediated by different underlying reasoning processes. The most straightforward explanation for the observed relation between span group and reasoning process is that participants choose the reasoning process that maximizes their reasoning performance.

Our findings converge with those of Copeland and Radvansky (2004) on syllogistic reasoning. They found that the degree to which participants relied on an analytic reasoning strategy was related to the operation span size of the participants. The larger their operation span, the more they relied on mental models to solve inferences. Our results can also be linked to the observations of Markovits et al. (2002), who found a significant correlation between verbal working memory and concrete reasoning for participants with high performance, whereas the correlation for the low performers was only marginally significant. They explained this result as follows: The high reasoning performance was based on a high-demand reasoning process (e.g., a mental models account) that drew heavily on working memory, whereas the low performance was attained by applying low-demand, heuristic processes (e.g., a probabilistic account) that put a minor load on working memory resources. The present verbal report results provide the necessary information to substantiate this claim: Highly skilled reasoners mainly refer to counterexample information, whereas reasoners with lower working memory performance are more likely to reason based on probabilistic information.

Instructing participants to think aloud enabled us to monitor concurrently the information they used for deriving conclusions and to check whether participants revised or nuanced their answers. By using spontaneous thinking-aloud instructions, we could investigate the reasoning process in a natural setting, where possible leading effects of instruction and answer options were excluded. From this point of view, verbal report analysis can be considered to be far more informative than the obtained inference acceptance rates from a conditional answer selection task. The downside of verbal report investigation is twofold: First, collecting, transcribing, and rating the answers are very time-consuming activities. Second, it is not yet clear to what extent the thinking-aloud results can be generalized to the results of standard deduction tasks. Further research can better define this generalizability. We suggest two possible methods for obtaining converging evidence: First, a research method like the one used by, for instance, Campbell and Gunter

(2002) and Copeland and Radvansky (2004) may provide additional information about the underlying processes of reasoning. In these studies, after participants solved a particular problem (respectively, a mathematical or a syllogistic reasoning problem) they were immediately asked to indicate how they came to their conclusion. A similar method could be used for investigating the underlying processes of conditional reasoning and could provide converging evidence for the present results. A second—more indirect—way of investigating the relative importance of both processes would be by use of a multiple regression analysis (see Verschueren et al., 2003a, in press).

## CONCLUSION

By use of a thinking-aloud study, we found evidence for two theories describing common-sense reasoning with causal conditional inferences. It was assumed that reasoning with counterexample information draws heavily on working memory resources, whereas the use of simple likelihood estimates imposes a far lesser demand on working memory. The present results show that participants with a relatively small operation span more often use likelihood estimates to attain a conclusion. When working memory capacity increases, participants preferentially use counterexample information to solve an inference. We thus conclude that working memory capacity is not only crucial for reasoning itself but also determines the process that reasoners use for reaching a conclusion. The present investigation underscores the idea that by distinguishing heuristic and analytic reasoning processes, we can obtain a more accurate and differentiated picture of working memory involvement in conditional reasoning.

## REFERENCES

- BADDELEY, A. D., & LOGIE, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.
- BARROUILLET, P., & LECAS, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, *5*, 289-302.
- BINDRA, D., CLARKE, K. A., & SHULTZ, T. R. (1980). Understanding predictive relations of necessity and sufficiency in formally equivalent "causal" and "logical" problems. *Journal of Experimental Psychology: General*, *109*, 422-443.
- BYRNE, R. M. J., ESPINO, O., & SANTAMARIA, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory & Language*, *40*, 347-373.
- CAMPBELL, J. I. D., & GUNTER, R. (2002). Calculations, culture and the repeated operant effect. *Cognition*, *86*, 71-96.
- COPELAND, D. E., & RADVANSKY, G. A. (2004). Working memory and syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *57A*, 1437-1457.
- CUMMINS, D. D., LUBART, T., ALKSNIS, O., & RIST, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274-282.
- DE NEYS, W. (2003). *In search of counterexamples: A specification of the memory search process for stored counterexamples during conditional reasoning*. Unpublished doctoral dissertation, University of Leuven.
- 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.

- DIEUSSAERT, K., SCHAEKEN, W., SCHROYENS, W., & D'YDEWALLE, G. (1999). Strategies for dealing with complex deductive problems: Combining and dividing. *Psychologica Belgica*, **39**, 215-234.
- 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.
- ERICSSON, K. A., & SIMON, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- EVANS, J. ST. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, **48**, 1-20.
- EVANS, J. ST. B. T., HANDLEY, S. J., & OVER, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 321-335.
- EVANS, J. ST. B. T., & OVER, D. E. (1996). *Rationality and reasoning*. Hove, U.K.: Psychology Press.
- FORD, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, **54**, 1-71.
- GEORGE, C. (1997). Reasoning with uncertain premises. *Thinking & Reasoning*, **3**, 161-189.
- GILHOOLY, K. J., LOGIE, R. H., & WYNN, V. (1999). Syllogistic reasoning tasks, working memory, and skill. *European Journal of Cognitive Psychology*, **11**, 473-498.
- JOHNSON-LAIRD, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, **50**, 189-209.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics and inference. *Psychological Review*, **109**, 646-678.
- JOHNSON-LAIRD, P. N., BYRNE, R. M. J., & SCHAEKEN, W. (1992). Propositional reasoning by model. *Psychological Review*, **99**, 418-439.
- 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.
- KLACZYNSKI, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision making. *Child Development*, **72**, 844-861.
- KLAUER, K. C., STEGMAIER, R., & MEISER, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking & Reasoning*, **3**, 9-47.
- 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, & Cognition*, **16**, 1118-1133.
- LIU, I., LO, K., & WU, J. (1996). A probabilistic interpretation of "if-then." *Quarterly Journal of Experimental Psychology*, **49A**, 828-844.
- 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 & Reasoning*, **8**, 97-107.
- MARKOVITS, H., FLEURY, M. L., QUINN, S., & VENET, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development*, **69**, 742-755.
- MEISER, T., KLAUER, K. C., & NAÜMER, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica*, **106**, 303-327.
- MIYAKE, A., FRIEDMAN, N. P., EMERSON, M. J., WITZKI, A. H., & HOWERTER, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, **41**, 49-100.
- MIYAKE, A., & SHAH, P. (EDS.) (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.
- MOSKOWITZ, G. B., SKURNIK, I., & GALINSKY, A. D. (1999). The history of dual-process notions and the future of preconscious control. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 12-41). New York: Guilford Press.
- OAKSFORD, M., CHATER, N., & LARKIN, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 883-899.
- OBERAUER, K., SÜSS, H.-M., SCHULZE, R., WILHELM, O., & WITTMANN, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality & Individual Differences*, **29**, 1017-1045.
- OBERAUER, K., SÜSS, H.-M., WILHELM, O., & WITTMANN, W. W. (2003). The multiple facets of working memory: Storage, processing, supervision, and coordination. *Intelligence*, **31**, 167-193.
- OBERAUER, K., & WILHELM, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 680-693.
- PAYNE, J. W. (1994). Thinking-aloud: Insights into information processing. *Psychological Science*, **5**, 245-248.
- POLITZER, G., & BOURMAUD, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology*, **93**, 345-381.
- ROSEN, V. M., & ENGLE, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, **126**, 211-227.
- SCHROYENS, W., & SCHAEKEN, W. (2004). Guilt by association: On iffy propositions and the proper treatment of mental-models theory. *Current Psychology Letters*, **12** (1). Available at <http://cpl.revues.org/document411.html>.
- STANOVICH, K. E., & WEST, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural & Brain Sciences*, **23**, 645-726.
- STENNING, K., & VAN LAMBALGEN, M. (2001). Semantics as a foundation for psychology: A case study of Wason's selection task. *Journal of Logic, Language, & Information*, **10**, 273-317.
- STEVENSON, R. J., & OVER, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology*, **48A**, 613-643.
- THOMPSON, V. (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.
- TURNER, M. L., & ENGLE, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, **28**, 127-154.
- VENET, M., & MARKOVITS, H. (2001). Understanding uncertainty with abstract conditional premises. *Merrill-Palmer Quarterly*, **47**, 74-99.
- VERSCHUEREN, N., SCHAEKEN, W., DE NEYS, W., & D'YDEWALLE, G. (2004). The difference between generating counterexamples and using them during reasoning. *Quarterly Journal of Experimental Psychology*, **57A**, 1285-1308.
- VERSCHUEREN, N., SCHAEKEN, W., & D'YDEWALLE, G. (2003a). Two reasoning mechanisms for solving the conditional fallacies. In A. Markman & L. Barsalou (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 1176-1181). Mahwah, NJ: Erlbaum.
- VERSCHUEREN, N., SCHAEKEN, W., & D'YDEWALLE, G. (2003b). *A verbal protocol study on causal conditional reasoning* (Tech. Rep. No. 301). Leuven: University of Leuven, Laboratory of Experimental Psychology.
- VERSCHUEREN, N., SCHAEKEN, W., & D'YDEWALLE, G. (in press). A dual-process theory on causal conditional reasoning. *Thinking & Reasoning*.
- WEIDENFELD, A., & OBERAUER, K. (2003). Reasoning from causal and noncausal conditionals: Testing an integrated framework. In A. Markman & L. Barsalou (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 1212-1217). Mahwah, NJ: Erlbaum.

## NOTES

1. Because the estimations participants give do not follow the rules of probability—for example, their estimates of exhaustive and mutually exclusive situations do not add up to 1—we prefer to use the term *likelihood* instead of *probability*.

2. An excerpt from the report data can be found at [www.psy.kuleuven.ac.be/reason/niki/data](http://www.psy.kuleuven.ac.be/reason/niki/data), under the filename `protocol_1.pdf`.

## APPENDIX A

### Scoring Instructions

An example of an AC inference:

If you go on a diet, then you will lose weight.  
Someone lost weight.  
Did this person go on a diet?

#### Inference acceptance (1/0)

Participants accept the conclusion without providing extra information. There is no mention of likelihood adverbs, probability indicators or counterexamples. Examples:

*yes*  
*yes, he went on a diet*

If an answer is scored as inference acceptance, it can *not* be scored on probability or counterexample.

#### Counterexample (1/0)

The participants refer to possible situations that contradict the suggested conclusion. For AC these possible situations refer to alternative causes. Examples:

*unless he got sick*  
*he might (also) have gotten sick*  
*a disease is also possible*

*Special case of counterexamples:* When people do not mention a determinate conclusion but rather say that the conclusion “depends” on certain variables or that the conclusion only occurs when certain “preconditions” are fulfilled. Examples:

*it depends on whether this person got sick*  
*only if this person did not get sick/stayed healthy*

This answer can occur in combination with a likelihood answer, but not necessarily.

(Similar examples and instructions were  
provided for MP.)

When an answer is scored for counterexample, you indicate *how many* counterexamples are mentioned.

#### Number of counterexample (1–10)

##### Probability/likelihood (1/0)

There is mention of a likelihood, probability, chance, frequency, percentage. . . . Examples:

*it is most likely/probable/unlikely/ . . . that he went on a diet*  
*this person never/seldom/sometimes/often/always/ . . . went on a diet*  
*there is a little/fat/ . . . chance that he went on a diet*

This answer can occur in combination with a counterexample, but it is not necessary.

For each counterexample trial, you count the number of counterexamples that the participant mentions.

##### Note

—The scoring does not entail a judgment about the quality of the answer (good, poor). For scoring the answers you only rely on the formal characteristics of the answer.

—In case of motivation request: Ignore the subsequent argumentation of the participant.

## APPENDIX B

### Illustrative Text Examples of the Thinking-Aloud Results (Literal Translation From Dutch)

#### Inference acceptance

- If you depress the brake, then the car will slow down.

The brake of a car is depressed.  
Will this car slow down?

**Participant 4:** Yes. (request for motivation) In *most cases* it is, at least I think so.

- If you have a Y-chromosome, then you will be masculine. Someone has a Y-chromosome.  
Is this person masculine?

**Participant 11:** Yes. (request for motivation) It is a scientific fact, there are almost *no other possibilities*.

- If the light switch is flipped, then the light will go on.  
The light switch is flipped.

Does the light go on?

**Participant 20:** Yes. (request for motivation) It is *normal*, because you know that it is connected with each other and that always when you have flipped the light switch the light goes on. So *in normal cases* the light goes on, that is, when the electricity did not go out or so.

#### Likelihood information

- If a predator is hungry, then it will search for prey.  
A predator goes searching for prey.  
Is this predator hungry?

**Participant 2:** Yes, the *probability is quite high* that this predator is hungry.

- If you go on a diet, then you will lose weight.

Someone goes on a diet.

Does this person lose weight?

**Participant 5:** Uhm . . . *mostly*, this person will lose a few kilos, but whether it also stays that way is still the question.

- If you phone someone, then his telephone will ring.

Someone's phone rings.

Did someone phone this person?

**Participant 51:** Then *it is most likely* [the case] that this person was phoned.

#### Counterexample information

- If the light switch is flipped, then the light will go on.

The light switch is flipped.

Does the light go on?

**Participant 5:** Yes, *unless the lamp is broken*, then the light does not come on.

- If a predator is hungry, then it will search for prey.

A predator goes searching for prey.

Is this predator hungry?

**Participant 10:** Yes, because otherwise you don't go searching for food *unless you want to lay in a winter-store*.

## APPENDIX B (Continued)

9. If you depress the brake, then the car will slow down.  
A car slows down.

Is the brake of this car depressed?

**Participant 56:** I think that the brake must have been depressed, *unless you are driving uphill*, then you also suddenly slow down.

**Likelihood followed by counterexample information**

10. If you go on a diet, then you will lose weight.

A person has lost weight.

Did this person go on a diet?

**Participant 35:** *Probably* he did, but it can also be that *the person has been ill* and ate less because of that [his illness] or so.

11. If there are advertisements for a product, then the sales records will rise.

There are advertisements for a product.

Will the sales records rise?

**Participant 36:** *Most of the time* the sales records will then rise, but if *a product is bad* then the story goes the round under people that it is bad and then it will not sell well.

12. If someone is nearsighted, then he will wear glasses.

Someone is nearsighted.

Does this person wear glasses?

**Participant 55:** Then he *probably* wears glasses but he can also *wear contacts*.

**Counterexample followed by likelihood information**

13. If you sit in the cold draft, then you will catch a cold.

A person sits in the draft.

Does this person catch a cold?

**Participant 5:** That is a possibility but it is also possible that he has still *a strong resistance* or that he is *not tired* and then he will *probably not* catch a cold.

14. If you drink lots of Coke, then you will gain weight.

A person has gained weight.

Did this person drink lots of Coke?

**Participant 9:** It can also have other reasons, for instance, *candy* or *chips* or *greasy food*, so *not always*.

15. If you water a plant well, then it will stay green

A plant is well watered.

Does this plant stay green?

**Participant 22:** Yes, it is actually a precondition to stay green, but a plant can also *catch a disease* or have *too much water* or *too little sun* or so, he will therefore *not always* stay green.

**Likelihood—counterexample—likelihood**

16. If you eat lots of salt, then you will get thirsty.

Someone is thirsty.

Did this person eat lots of salt?

**Participant 24:** I think that the *chances are rather thin* that that caused him to get thirsty, I think that you get thirsty much faster [more likely] because it has been a *long time* that you have got in a little water, and since people are in need of fluids, we are also in need of drinking. Maybe you *just sported* or *it is warm*, or yeah . . . you ate salt, that is also a possibility. The *probability is thus rather low* that that [eating salt] was the cause [of being thirsty]

17. If a predator is hungry, then it will search for prey.

A predator is hungry.

Does this predator go searching for prey?

**Participant 27:** I think *in most cases* it will [search for prey]. I think if it is the oldest animal that just was repudiated from its [leader-] position and that it is lying down somewhere with *heavy injuries* that it cannot do very much any more. Then the poor animal can be hungry but cannot go search for prey. In the end it also depends on the *season*, some seasons are for a lot of animals difficult seasons to find food. Some animals then pass [on searching for prey] and wait for the spring to return when they have plenty of food. But *in most cases* the predator will search for food.

18. If water is heated to 100°C, then the water will boil.

Water is boiling.

Was this water heated to 100°C?

**Participant 27:** Uhm, I think it is *highly likely* [that it was heated], I do not know for sure, but I think that I maybe saw in physics [class] that there can originate a boiling effect at a *lower [temperature]*, I do not know anymore which circumstance [caused it], but I think that it is rather *infrequent*, I think that it is *highly probable* that it was heated to 100°C.

**Extra examples**

19. If you go on a diet, then you will lose weight.

A person has lost weight.

Did this person go on a diet?

**Participant 27:** Yeah . . . going on a diet is still a voluntary action, while if there is no food around than you will not really go on a diet, but you will lose weight. In our western society I think that in most cases, someone [who lost weight] went on a diet. Bwah, you can also have done a *liposuction* or so, but in *most cases* it must have been the case [that he went on a diet].

20. If you phone someone, then his telephone will ring.

Someone's phone rings.

Did someone phone this person?

**Participant 26:** In most cases this person is phoned, at our place the *voicemail* also rings when someone left a message. But it is *highly probable* [that someone phoned].

21. If a girl has sexual intercourse, then she will be pregnant.

A girl is pregnant.

Did this girl have sex?

**Participant 33:** Yes, I think that is right, then she has had sex. Though, you can also be pregnant by *I.V.F.* that can also be [the cause], but yeah . . . that *does not occur that often*.

22. If a girl has sexual intercourse, then she will be pregnant.

A girl has sexual intercourse.

Will this girl be pregnant?

**Participant 37:** Only if she has sex during her *fertile period* she will be pregnant, otherwise she will *most likely not* get pregnant.

23. If you sit in the cold draft, then you will catch a cold.

A person has caught a cold.

Did this person sit in the draft?

**Participant 17:** He does not have to be seated in the cold, it is also possible that he went *playing in the snow*, just wearing a T-shirt, in that case there is a *good chance* that he catches a cold.