Memory & Cognition
2002, 30 (6), 908-920

Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework

WIM DE NEYS, WALTER SCHAEKEN, and GÉRY D'YDEWALLE
University of Leuven, Leuven, Belgium

This study tested and refined a framework that proposes a mechanism for retrieving alternative causes and disabling conditions (Cummins, 1995) during reasoning. Experiment 1 examined the relation between different factors affecting retrieval. The test revealed high correlations between the number of possible alternative causes or disabling conditions and their strength of association and plausibility. Experiment 2 explored the hypothesis that due to a more extended search process, conditional inferences would last longer when many alternative causes or disabling conditions were available. Affirmation of the consequent (AC) and modus ponens (MP) latencies showed the hypothesized pattern. Denial of the antecedent (DA) and modus tollens (MT) inferences did not show latency effects. The experiment also identified an effect of the number of disabling conditions on AC and DA acceptance. Experiment 3 measured efficiency of disabler retrieval by a limited time, disabler generation task. As predicted, better disabler retrieval was related to lower acceptance of the MP and MT inferences.

Conditional reasoning involves making inferences on the basis of an "if–then" relation and is considered one of the cornerstones of human reasoning. Research on conditional reasoning has been trying to identify the factors and processes that affect the performance on these "if–then" inference problems. In a standard conditional inference task, people are asked to assess arguments of the following four kinds:

Modus ponens (MP)	If p then q, p therefore q
Modus tollens (MT)	If p then q, not q therefore not p
Denial of the antecedent (DA)	If p then q, not p therefore not q
Affirmation of the consequent (AC)	If p then q, q therefore p

In standard propositional logic, MP and MT are considered valid inferences and DA and AC are regarded as fallacies.

A growing body of evidence is showing that people's knowledge about the relation between the p (antecedent) and q (consequent) part of the conditional has a considerable effect on the reasoning process. In particular, the role of knowledge of alternative causes and disabling conditions has attracted interest (see Politzer, in press, for a review).

An alternative cause (alternative) is a possible cause that can produce the effect mentioned in the conditional, whereas a disabling condition (disabler) prevents the effect from occurring despite the presence of the cause. Consider the following conditional:

If the air conditioner is turned on, then you feel cool

Possible alternative causes for this conditional are as follows:

Taking off some clothes, the weather cools, swimming . . .

The alternatives make it clear that it is not necessary to turn on the air conditioner in order to feel cool. Other causes are also possible.

Possible disabling conditions are as follows:

Air conditioner is broken, having fever, window open . . .

If such disablers are present, turning on the air conditioner will not result in feeling cool. The disablers make it clear that it is not sufficient to turn on the air conditioner in order to feel cool. Additional conditions must be fulfilled.

Rumain, Connell, and Braine (1983) showed that when a possible alternative was explicitly presented to participants, the AC and DA inferences were less endorsed. Byrne (1989) found a similar effect on MP and MT when a possible disabling condition was mentioned. In addition, using familiar relations (e.g., "If an animal has feathers, it is a bird") for which people have ready access to alternatives, Markovits (1986) showed that even without explicit

Preparation of the manuscript was supported by grants from the Fund for Scientific Research-Flanders (FWO) and the Belgian Program on Interuniversity Poles of Attraction, Convention No. P4/19. The paper was greatly improved by comments from Evan Heit, Denise Cummins, Mike Oaksford, and Ruth Byrne. Correspondence should be addressed to W. De Neys, Department of Psychology, Tienestraat 102, B-3000, Leuven, Belgium (e-mail: wim.deneys@psy.kuleuven.ac.be).

presentation, awareness of a possible alternative decreased the number of AC and DA inferences.

Cummins and colleagues (Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991) have conducted further seminal work on the impact of alternatives and disablers in conditional reasoning. Cummins (1995) and Cummins et al. (1991) directly addressed the role of stored knowledge of alternatives and disablers by examining the effect of the number of available alternatives and disablers. In a pretest they identified causal conditionals (i.e., conditionals that express a causal relation) for which participants generated many or few alternatives and disablers. These conditionals were then adopted for a conditional reasoning task. Results showed that people's acceptance of DA and AC inferences decreased for conditionals with many possible alternatives. In addition, the number of disabling conditions affected the acceptance of the MP and MT inferences. If there were many conditions that could disable the relation between antecedent and consequent, people tended also to reject these valid inferences. Alternatives and disablers were not explicitly presented, showing that a crucial factor in causal conditional reasoning is the number of alternative causes and disabling conditions people can think of.

Cummins (1995) argued that finding possible alternative causes and disabling conditions affects people's interpretation of the necessity and sufficiency of a cause for bringing about the effect in question. Thompson (1994) showed that the number of disablers and alternatives effect generalized to noncausal conditional relations (e.g., permissions, obligations, and definitions).

The results of Cummins's (1995) experiments imply that during a conditional reasoning task, people search their memory for stored knowledge of alternatives and disablers. Since the outcome of this retrieval process determines which conclusions people are willing to draw, it is of crucial importance for the reasoning community to clarify how the search process is affecting reasoning (Johnson-Laird & Byrne, 1994; Oaksford & Chater, 1998).

In a number of studies, Markovits and collaborators have started to specify this search mechanism, which constitutes the core of their general model of conditional reasoning (see Janveau-Brennan & Markovits, 1999; Markovits, 2000; Markovits, Fleury, Quinn, & Venet, 1998; Markovits & Potvin, 2001; Quinn & Markovits, 1998).

The model states that while making conditional inferences, reasoners will automatically access structures with relevant information in semantic memory.[1] Such a structure contains semantically or propositionally related elements. In causal conditional reasoning, the structures would consist of possible alternative causes and disabling conditions. Alternatives and disablers would be stored in different structures. According to many influential models of long-term memory (e.g., Anderson, 1983; Gillund & Shiffrin, 1984), the probability of retrieving at least one element from such a semantic memory structure will depend on the number of elements within the structure. Thus, the probability of retrieving at least one element from the structure storing alternative causes will be higher for conditionals with many possible alternative causes. In the same way, the probability of retrieving a disabling condition will be higher for conditionals with many possible disablers.

The model thus accounts for the effect of number of alternatives and disablers on the underlying reasoning process. The number of possible alternatives and disablers affects successful retrieval of such an element from the corresponding semantic memory structures. When an alternative cause is retrieved, the original antecedent will no longer be perceived as necessary for bringing about the consequent. As a consequence, the DA and AC inferences will be less accepted. Retrieval of a disabling condition will decrease the perceived sufficiency of the original antecedent for bringing about the consequent. This will result in increased rejection of the MP and MT inferences.

We group these ideas under the heading of "semantic memory framework." Although the framework starts specifying a crucial component of the reasoning process, a number of propertes are not addressed and remain untested. The present study focuses on this issue. We present three experiments where the characteristics of the semantic search process during conditional reasoning are further tested and explored.

First, we address a neglected issue concerning the relation between different retrieval factors. Remember that one of the framework's central claims is that the probability of successful retrieval is higher for conditionals with many alternatives or disablers. Although it is well established that the probability of retrieval is affected by the number of stored elements, other factors are known to affect the retrieval too. The relation between these factors has to be taken into account in order to validate the claim. A prior concern here is the recently identified impact of "strength of association" on conditional reasoning (see De Neys, Schaeken, & d'Ydewalle, in press; Quinn & Markovits, 1998).

Quinn and Markovits (1998) identified the associative strength factor within the semantic memory framework: In the memory structure with possible alternative causes of a conditional, some alternatives will be more strongly associated with the consequent in question than will others. For example, the cause "the dog has fleas" will be more strongly associated with the consequent "a dog scratches constantly" than the cause "the dog has skin disease". Quinn and Markovits showed that in addition to the number of alternative elements in a structure, the relative strength of association facilitated retrieval. In a related study, De Neys et al. (in press) found that the strength of association effect was also present for the disabling conditions.

Participants in Experiment 1 were asked to generate disablers and alternatives for a set of conditionals. As in Cummins's (1995) pretest, we recorded the number of generated alternatives and disablers in order to classify the conditionals in groups with many and few disablers and alternatives. We also recorded how frequently each individual alternative and disabler was generated to measure strength of association (see Quinn & Markovits, 1998). This allowed us to establish the relation between both retrieval factors.

In addition, plausibility ratings for the individual disablers and alternatives were collected. Since more readily available stored elements will tend to be judged more likely (Kahneman, Slovic, & Tversky, 1982), this allowed us to take an additional retrieval factor into account.

The impact of the number of stored disablers and alternatives on inference acceptance is well established. In the second experiment we looked for the first time at the impact on the inference latencies. The additional latency data can contribute to a further characterization of the semantic memory framework: Whether or not one has stored many disablers or alternatives will have clear processing consequences. Of primary interest here is the fact that some studies (e.g., Conway & Engle, 1994) have shown that the time course of a memory search process is affected by the number of elements that are retrieved from a memory structure. It is likely that for conditionals with many disablers or alternatives, more of these elements will be retrieved. Therefore, within the semantic memory framework, one might expect that the time needed for an inference will also be affected by the available number of disablers and alternatives. This was explored in Experiment 2 by recording both acceptance ratings and reaction times for the different inferences in a replication of Cummins (1995). The replication also allowed us to examine the generality of a previously reported (e.g., Liu, Lo, & Wu, 1996; Markovits & Potvin, 2001) additional effect of the number of disablers on AC and DA acceptance.

Finally, in a third experiment we tried to obtain further evidence for the role of the memory search process in conditional reasoning. If the outcome of the search process determines the inferences people draw, we should expect that individual differences in the efficiency of the retrieval process will affect reasoning performance. Janveau-Brennan and Markovits (1999) showed that elementary school children's capacity to generate possible alternative conditions was related to their performance on the AC and DA inferences. Here, we tested whether individual differences in adult reasoners' capacity to retrieve disabling conditions would affect performance on MP and MT inferences. The semantic memory framework states that the increased rejection of MP and MT results from the successful retrieval of a disabling condition. Since the probability of successful retrieval will be higher for people with a more efficient search process, we predict that the better one is at retrieving disablers from semantic memory, the less one will accept MP and MT.

EXPERIMENT 1
Relation Between Retrieval Factors

Whether or not the semantic search process retrieves an alternative or disabler depends on more than the number of stored elements. The probability of successful retrieval will also increase with the associative strength or plausibility of the stored elements. In Experiment 1 we examined the relation between different factors affecting the retrieval of stored alternatives and disablers. Participants generated disablers and alternatives for a set of conditionals. On the basis of the number of generated items, conditionals were classified in groups with many and few alternatives or disablers. Associative strength was measured by recording the frequency of generation of the individual alternatives and disablers. After the generation task, participants rated the plausibility of the generated alternatives or disablers.

Within the semantic memory framework, one would expect to see positive relations between the factors of number, associative strength, and plausibility. That is, if conditionals with more disablers also have more strongly associated and plausible disablers, the claim that the probability of successful retrieval is higher for conditionals with many disablers would be validated.

Method

Participants. Forty students in introductory psychology were enrolled. Half the participants were required to generate possible alternative causes and the other half to generate possible disabling conditions. All participants were native Dutch speakers.

Material. The 16 causal conditionals from Cummins (1995, Experiment 1A) were used for the generation task. Cummins selected these conditionals because they constituted a 2 (few vs. many alternatives) 3 2 (few vs. many disablers) manipulation of the factors of number of alternative causes and number of disabling conditions. Another four causal conditionals that seemed to vary in terms of possible alternatives and disablers were taken from the literature. Item format and instructions were similar to those of the generation task in Cummins and Cummins et al. (1991). Thus, the following format was used:

> Rule: If the air conditioner is turned on, then you feel cool
> Fact: You feel cool, but the air conditioner was not turned on
>
> Please write down as many factors as you can that could make this situation possible.

This is an example of the alternative causes generation task. The format of the disabling conditions generation task was similar except that under the heading fact would appear "The air conditioner was turned on, but you don't feel cool." Formats like these were constructed for each of the 20 conditionals; they were typed one to a page in a booklet. The order in which the conditionals appeared in the different booklets was randomized. Task instruction stressed the importance of producing items that were reasonably realistic and different from each other. Participants were instructed that simple variations of the same, simple idea (for the example above, "taking off shirt," "taking off sweater", "taking off coat") would be scored as a single item and should to be avoided.

Procedure. Participants were run in groups of 2 to 6. The top sheet of the generation task booklet included the written instructions, which were read aloud to the participants. Participants had 1.5 min to write down their answers for each conditional.

After the generation task, participants received written instructions for the plausibility rating task. They were asked to rate the plausibility of the disablers or alternatives they had generated on an 11-point scale, (0 5 very implausible; 10 5 very plausible). Ratings had to be written down in the booklets alongside each generated disabler or alternative. Participants received a practice conditional ("If Bart drinks coffee in the evening, then he doesn't sleep well") with possible disablers or alternatives. Instructions for the disabler rating task made clear that we wanted participants to rate how plausible they judged the generated disablers as "explanations" for the nonoccurrence of the specified effects (e.g., "Bart drank coffee in the evening, but he slept well. How plausible is it that this was due to the fact that the coffee was decaffeinated?"). Participants who had generated alternatives were instructed to rate the plausibility of the generated factors as alternative causes for the specified effects (e.g., "Bart did not sleep well. How plausible is it that this was due to the fact that it was too noisy?").

As in Quinn and Markovits (1998) and De Neys et al. (in press), associative strength was measured by recording how frequently an individual alternative or disabler was generated across participants. For example, if 12 of the 20 participants would generate "taking off sweater" as an alternative for "If the air conditioner is turned on, then you feel cool," "taking off sweater" would receive an associative strength of 60%.

The generation protocols were scored by two independent raters in order to identify unrealistic items and items that were variations of a single idea.

Results and Discussion

Overall, 5.6% of the generated items were disallowed by the raters. Interrater reliability for the alternatives generation task was .95 and for the disablers generation task was .83. These figures are in line with the data from the Cummins (1995) generation task.

In general, the mean number of generated alternatives and disablers for the 16 "Cummins" conditionals was also very similar to what was originally reported. Nevertheless, two original conditionals ("If Alvin reads without his glasses, then he gets a headache" and "If the doorbell is pushed, then it will ring") needed to be replaced because the number of generated disablers differed substantially from those in Cummins (1995). Table A1 in the Appendix presents the 16 selected conditionals. The table contains the respective means together with the most frequently generated alternatives and disablers for every conditional.

We were interested in the differences in generation frequency and plausibility ratings of alternatives and disablers for conditionals classified as having many or few of these conditions. Results are summarized in Table 1. Overall, when a conditional had many alternatives or disablers, the disablers (34% vs. 32%) and alternatives (30% vs. 22%) also tended to have a higher association strength. Spearman rank-order correlations confirmed this positive relation between a conditional's number of possible disablers and the disablers' mean association strength [r_s 5 .43, n 5 16, t(14) 5 1.77, p , .1]. The alternatives showed a similar relation [r_s 5 .65, n 5 16, t(14) 5 3.23, p , .01]. To provide more detail, we also looked at the number of strongly associated alternatives/disablers in both groups. Three frequency levels (disabler or alternative generated by at least 50%, 75%, and 90% of participants) were used as criterion.

As Table 1 indicates, comparing the number of strongly associated elements supported the global analysis: The group of conditionals with a high number of alternatives also has a higher mean number of strongly associated alternatives (at the 50%, 75%, and 90% levels). The disablers show the same trend at the 50% and 75% levels. The high correlations between the number and number of

Table 1
Mean Number and Mean Plausibility of Generated Alternatives/Disablers for Conditionals Classified as Having Many or Few Alternatives/Disablers

		Alternatives		Disablers	
		Many	Few	Many	Few
Mean Number		4.03	1.91	4.03	2.31
Mean Plausibility		6.34	5.08	6.29	5.21
Mean Number AS	50%	3	.88	2.88	1.75
	75%	1.13	0.25	1.5	0.75
	90%	0.63	0.13	0.5	0.5

Note—There are eight conditionals in each group. The last three rows present the mean number of alternatives and disablers generated by at least 50%, 75%, and 90% of the participants. The plausibility rating scale ranged from 0 (*very implausible*) to 10 (*very plausible*). AS, associative strength.

strongly associated (50%) alternatives [r_s 5 .82, n 5 16, t(14) 5 5.82, p , .001] and disablers [r_s 5 .73, n 5 16, t(14) 5 4.01, p , .002] supported these findings.[2]

Thus, conditionals with a higher number of possible disablers or alternatives will also have more strongly associated disablers and alternatives. Since both number and strength of association increase the probability of retrieval, these results validate the claim that successful retrieval is more probable for conditionals with many alternatives or disablers.

Table 1 further shows that the generated disablers and alternatives for conditionals classified as having many disablers/alternatives were also rated as more plausible. For the 16 selected conditionals, Spearman rank-order correlations showed a high positive correlation between the number of generated disablers and the mean plausibility of these disablers [r_s 5 .70, n 5 16, t(14) 5 4.01, p , .002]. The same trend was observed for the alternatives [r_s 5 .62, n 5 16, t(14) 5 2.97, p , .02]. Given the often observed "availability heuristic" (Kahneman et al., 1982), one can assume that the higher plausibility ratings reflect easier retrieval. In this sense, the findings present converging evidence for the strength of association conclusion.

The plausibility rating also gives us an indication of the "quality" of the retrieved disablers and alternatives. Chan and Chua (1994) showed that this "quality" of a presented disabler (i.e., the salience or perceived importance of the disabler in relation to the occurrence of the consequent) affected MP and MT acceptance. This could imply that in addition to the higher retrieval probability, a disabler (or alternative) that is retrieved from a memory structure with many stored elements will have a stronger impact on inference acceptance. Notably, this might be an additional factor that contributes to the effect of the number of alternatives and disablers on conditional reasoning.

EXPERIMENT 2
Inference Study

First, Experiment 2 explored the effect of the number of stored alternatives and disablers on conditional inference latencies. Previous studies have focused solely on the impact on inference acceptance. However, latency data may allow us to further characterize the crucial semantic search process during conditional reasoning.

Experiment 1 showed that for conditionals with many disablers or alternatives, the memory structure storing these elements will typically contain a number of strongly associated elements. Available evidence from memory studies (Conway & Engle, 1994) suggests that (up to four elements) a semantic search process will take longer when the number of elements that are retrieved from a memory structure increases. When one is making conditional inferences, such a longer search process should result in longer inference times. Thus, availability of many alternatives and disablers might result not only in lower AC/DA and MP/MT inference acceptance but, due to an extended search process, also in longer inference latencies. This expectation was explored by recording both acceptance ratings of the conclusions and the time needed to evaluate them in a replication of Cummins (1995, Experiment 1A).

Second, Experiment 2 allowed us to examine the possible impact of the number of disablers on AC and DA inferences. Liu et al. (1996) and Markovits and Potvin (2001, Experiment 3) observed that AC and DA were more accepted when many disablers were available, although Cummins (1995) did not. At present, the semantic memory framework does not incorporate such an effect. In view of the procedural variations in the cited studies, which might restrict comparability with Cummins's work, we used a large-scale (100 participants) replication to assess the generality of the trend.

It is important to note that in the present study we are interested in the general inference latency pattern. Therefore, in contrast to more traditional reasoning studies, we analyzed latencies in terms of the number (few vs. many) of stored alternatives and disablers independent of the specific acceptance ratings. Furthermore, as in Cummins (1995), a graded rating scale was used to measure inference acceptance. Thus, "acceptance/rejection of an inference" should be interpreted relative to the rating scale (i.e., rejection indicates lower acceptance ratings).

In sum, except for the possible impact of disabling conditions on AC and DA, we expected to replicate Cummins's (1995) findings on the acceptance ratings: MP and MT should be rejected more for conditionals with many disablers than for conditionals with few disablers. AC and DA should be rejected more for conditionals with many alternatives than for conditionals with few alternatives. Given the extended search hypothesis, one would expect longer MP and MT latencies for conditionals with many disablers than for conditionals with few disablers. Likewise, we expected longer AC and DA latencies for conditionals with many alternatives than for conditionals with few alternatives.

Method

Participants. One hundred and one undergraduate students from the University of Leuven participated as paid volunteers or for partial fulfillment of a course requirement. All were native Dutch speakers and none had had training in formal logic.

Materials. The 16 conditionals selected in the generation task (see Table A1) were used. The conditionals yielded a 2 (few/many) 3 2 (alternatives/disablers) design with four items per cell. The 16 conditionals were embedded in the four (MP, DA, MT, and DA) inference types, producing a total of 64 inferences for each participant to evaluate.

The experiment was run on computer. The item format was based on Cummins (1995). Each argument was presented on screen together with a 7-point rating scale and accompanying statements. This resulted in the following format:

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

Conclusion: Jenny feels cool

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

```
                                            I
—1————2————3————4————5————6————7—
Very      Sure    Somewhat    I    Somewhat    Sure    Very
Sure                Sure      I      Sure              Sure
                              I
That I cannot draw this       I         That I can draw this
conclusion                                        conclusion
```

Type the number that best reflects your decision about the conclusion:

Each of the 64 arguments was presented in this way. The premises, conclusion, and typed number were always presented in yellow. The remaining text appeared in white on a black background.

Procedure. Participants were run in groups of 2 to 8. Instructions were presented verbally and on screen, showing 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. Participants used the keypad to type the number reflecting their decision and pressed the Enter key when finished. The next item was presented 750 msec after the Enter key had been pressed. The instructions made clear that there were no time limits, but it was stressed that once participants had made their final decision, they had to press the Enter key immediately. The time between the presentation of the item and pressing the Enter key was recorded together with the acceptance rating. After half of the inferences were evaluated, item presentation stopped until participants decided to continue. Instructions stressed that no other breaks were allowed and participants were expected to work through the items one immediately after another. The 64 items were presented in random order. The experimental session was preceded by one practice trial.

It should be pointed out that, as in Cummins's (1995) study, participants were not explicitly instructed to accept the premises as true and to endorse only conclusions that followed necessarily. Instead, participants could evaluate the conclusions by the criteria they personally judged relevant. With Cummins we assume that this encourages people to reason as they would in everyday circumstances. However, we should note that strictly speaking the task is therefore not a deductive inference task (see Evans, 2000). Endorsing the logically valid (MP and MT) and invalid (AC and DA) inferences should therefore not be considered correct or incorrect reasoning. When we refer to the standard nomenclature, we adopt a nominalist stance toward the use of the terms valid inferences and fallacies.

Results and Discussion

Each participant evaluated inferences on the basis of four different conditionals within each 2 (number of alternatives) 3 2 (number of disablers) 3 4 (inference type) cell of the design. The mean of these four observations was calculated. These means were subjected to a 2 3 2 3 4

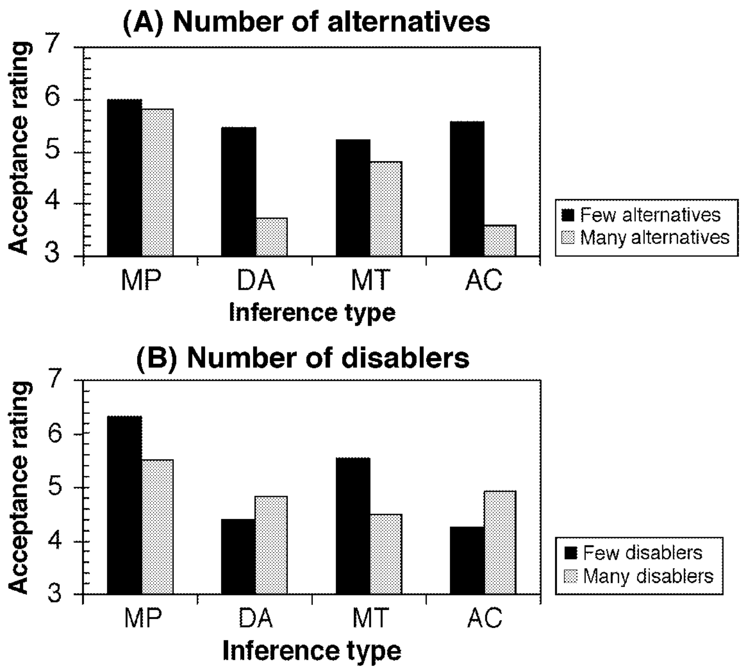

Figure 1. The effect of (A) the number of alternatives and (B) disablers on mean acceptance ratings for the four inference types. The rating scale ranged from 1 (*very sure cannot draw this conclusion*) to 7 (*very sure can draw this conclusion*), with 4 representing *can t tell*.

within-subjects analysis of variance (ANOVA) for acceptance ratings and reaction times.

Effects involving repeated measures with more than two levels were analyzed with multivariate ANOVA tests.

The data from 1 participant were discarded. Mean reaction times differed from the mean reaction times of the sample by more than 3 SD on MP, DA, and MT, and 2.5 SD on AC.

Acceptance ratings. The main effects of inference type, number of disablers, and number of alternatives were all significant [Rao R (3,97) 5 82.19, p , .0001; F(1,99) 5 17.94, MS_e 5 .89, p , .0001; F(1,99) 5 399.39, MS_e 5 1.17, p , .0001]. As in Cummins (1995), the interactions between inference type and number of disablers [Rao R (3,97) 5 106.41, p , .0001] and inference type and number of alternatives [Rao R (3,97) 5 96.34, p , .0001] were also significant. These interactions are depicted in Figure 1.

Planned contrast tests showed that acceptance of DA [F(1,99) 5 353.59, MS_e 5 .85, p , .0001] and AC [F(1,99) 5 383.59, MS_e 5 1.04, p , .0001] was significantly lower for conditionals with many alternatives. Further in line with Cummins (1995), the number of disablers affected acceptance of MP [F(1,99) 5 137.85, MS_e 5 .48, p , .0001] and MT [F(1,99) 5 121.41, MS_e 5 .94, p , .0001], with lower ratings for conditionals with many possible disablers.

These findings replicate Cummins (1995). The present study also identified some additional effects.[3] First, the number of alternatives affected not only AC and DA but also MP [F(1,99) 5 7.88, MS_e 5 .37, p , .01] and MT [F(1,99) 5 29.8, MS_e 5 .6, p , .0001] acceptance. For conditionals with many alternatives, slightly lower MP and MT ratings were obtained. Nevertheless, there was still an interaction between inference type and number of alternatives. As is clear from Figure 1A, the MP and MT effects were smaller than the impact of alternative retrieval on AC and DA.

The MP and MT findings can be explained if one takes into account that in our set of conditionals there was a positive correlation (r_s 5 .37, n.s.) between a conditional's number of possible alternatives and disablers (see Thompson, 2000, Appendix E, for a similar observation). Therefore, conditionals with many alternatives will also tend to have a slightly higher number of disablers than conditionals with few alternatives. Consequently, since retrieving disablers will be somewhat more probable for conditionals with many alternatives, the lower MP and MT acceptance is not surprising.

Second, the findings of Liu et al. (1996) and Markovits and Potvin (2001, Experiment 3) were replicated. In addition to the impact on MP and MT, the number of disablers also affected DA and AC acceptance. Both DA [F(1,99) 5 38.16, MS_e 5 .47, p , .0001] and AC [F(1,99) 5 74,13, MS_e 5 .58, p , .0001] were accepted more when many disablers were available. This effect cannot be explained by the correlation between the number of alternatives and disablers since this would result in a trend in the opposite direction.

A possible explanation points toward an interplay between the disabler and alternatives search processes. Markovits and Potvin (2001) observed that disabler retrieval can occur during a search for alternative causes. This suggests that retrieval of alternatives and disablers is not occurring in complete isolation. Note that at present the specific relation between both search processes is not specified in the semantic memory framework. We can assume that retrieving stored memory elements is resource demanding. A plausible hypothesis would be that retrieving (many) disablers puts a load on the cognitive system. This load would then burden the search for alternative causes. Thus, retrieval of alternatives would be less efficient for conditionals with many disablers, which would result in a lower retrieval probability and thus higher AC and DA acceptance. Consistent with this hypothesis, the effect of the number of disablers was mediated by the available number of alternatives on both AC [F(1,99) 5 6.56, MS_e 5 .34, p , .015] and DA [F(1,99) 5 6.04, MS_e 5 .5, p , .02]: Disabler retrieval affecting AC and DA was less pronounced for the conditionals that had few alternatives. Indeed, when few alternatives are available, successful retrieval is not likely anyway. Thus, an eventual burden of the search process should not be expected to have a major impact here.

Though interesting, the explanation is not unproblematic. For example, given that alternative retrieval did not increase MP and MT acceptance, one needs to explain why retrieval of disablers would burden retrieval of alternatives although the reverse does not occur. Such a pattern suggests that retrieving disablers somehow has priority over retrieving alternatives. It is clear that this issue demands further research. Meanwhile, in the light of recent demonstrations of the interplay between disabler and alternative retrieval processes (Markovits & Potvin, 2001), the overload mechanism should not be discarded.

Reaction times. In order to eliminate biased measures, a trimming procedure was applied to the reaction times before they were subjected to analysis. For every inference type, any latency that was more than 3 SD above a person's mean reaction time for that inference type was discarded. This procedure affected less than 1% of all observations.

The MANOVA revealed a significant main effect of inference type [Rao R (3,97) 5 43.03, p , .0001]. As for the acceptance ratings, we also observed significant interactions between inference type and number of alternatives [Rao R (3,97) 5 4.82, p , .004] and inference type and number of disablers [Rao R (3,97) 5 2.7, p , .05]. These interactions are depicted in Figure 2.

Planned contrasts tests indicated that the AC [F(1,99) 5 5.47, MS_e 5 7.52, p , .025] inference took more time (641 msec) for conditionals with many alternatives (Figure 2A). Likewise, the MP [F(1,99) 5 11.52, MS_e 5 5.27, p , .001] inferences required more time (776 sec) for the conditionals with many disablers (Figure 2B).

These findings show that the higher number of stored alternatives and disablers results not only in lower AC and MP acceptance, but also in increased inference latencies. This corroborates the hypothesized extended semantic search process for conditionals with many alternatives or disablers: When many (strongly associated) alternatives and disablers are available, retrieval of additional alternatives and disablers will result in longer inference latencies.

Although the number of available disablers and alternatives affected MT and DA acceptance, no effects were observed on the latencies. Thus, DA and MT latency data suggest there is no extended search process for these inferences. The findings can be reconciled if one takes into

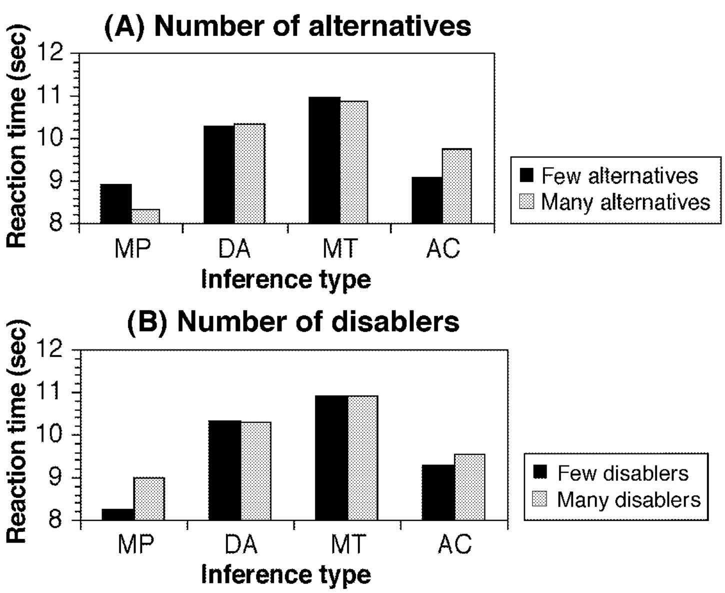

Figure 2. The effect of (A) the number of alternatives and (B) disablers on mean reaction times for the four inference types.

account that the semantic search process during conditional reasoning is possibly affected by inference complexity. DA and MT inferences are more complex than AC and MP inferences. DA and MT, the so-called denial inferences, involve negations, and reasoning theories typically state that these demand more cognitive (working memory) resources (Braine & O'Brien, 1998; Johnson-Laird & Byrne, 1993; Oaksford, Chater, & Larkin, 2000). Rosen and Engle (1997) have shown that semantic memory retrieval is affected by working memory load. Therefore, the increased complexity may affect the semantic search process. Due to the higher load, the search would be less extensive. The additional need to process negations would thus override the extended search for DA and MT inferences.

In general, this implies that for DA and MT based on conditionals with many alternatives/disablers, fewer of these elements will be retrieved than for MP and AC. Therefore, one might expect a less pronounced effect of the number of alternatives and disablers on DA and MT acceptance ratings. However, one has to consider that when the reasoning system is already burdened, retrieval of even a single alternative might cause a cognitive overload. Such an overload can result in increased inference rejection (e.g., Toms, Morris, & Ward, 1993). Therefore, even with a less extensive search process, availability of many alternatives/disablers can still result in considerable decreased MT and DA acceptance. Consistent with this claim, planned contrast tests showed that the number of disablers had a similar impact on MP and MT acceptance ratings. Likewise, AC and DA acceptance was not differentially affected by the number of available alternatives.

As Figure 2A indicates, MP inferences also lasted longer (595 msec) when few alternatives were available [$F(1,99)$ 5 11.9, MS_e 5 2.97, p , .001]. This effect of alternatives on MP reaction times is puzzling. The direction of the effect, with many alternative conditionals having shorter reaction times, was the opposite of the other latency effects. At present, we have no clear-cut explanation for this finding.

EXPERIMENT 3
Individual Differences in Disabler Retrieval Capacity

In Experiment 3 we tried to obtain further evidence for the role of the semantic search process during conditional reasoning. We tested whether individual differences in adult reasoners' capacity to retrieve disabling conditions affected performance on MP and MT inferences. The semantic memory framework states that the increased rejection of MP and MT results from the successful retrieval of a disabling condition. Since the probability of successful retrieval will be higher for people with a more efficient search process, we would expect that the better one is at retrieving disablers from semantic memory the less MP and MT should be accepted.

We do not predict an effect on the inference latencies. Reasoners with a more efficient retrieval process can be expected to be faster at retrieving disablers. Although more efficient retrievers will probably retrieve more disablers, the faster running search process will compensate the extra time needed for the additional retrieval.

Retrieval efficiency was measured by looking at the number of generated disablers in a generation task that was presented to 40 participants after the inference task of Experiment 2. The generation task used four new conditionals and four conditionals already presented in the inference task. The new conditionals were included because generation for the old conditionals may be biased by familiarity. On the other hand, retrieval of disablers for adult participants may be highly specialized and conditional specific. In this latter case, generation of disablers for the new conditionals would not be informative for the retrieval efficiency for conditionals in the inference task. Comparing the results for the old and new conditionals allows us to sidestep this possible complication.

Method

Participants. Forty participants that participated in Experiment 2 took part in the present experiment.

Materials. Disabler retrieval capacity was measured by requesting participants to generate possible disablers for eight conditionals. Half the conditionals had already been presented to participants in the inference study (5 old conditionals), and the other half were different (5 new conditionals). Half the old and new conditionals had been classified in previous work as possessing many disabling conditions, and the other half had only few possible disabling conditions. Item format and task instructions were similar to those for the generation task in Experiment 1. The conditionals appeared in the same order in all booklets. The old conditionals were presented after the new ones.

Procedure. The retrieval measure was presented after all participants of a group had finished the inference study. Generation time was limited to 30 sec for the new conditionals, as in Janveau-Brennan and Markovits (1999). To account for faster reading times (and thus more retrieval time) due to previous presentation, generation time was limited to 28 sec for the old conditionals. Participants had to write down their answers. Instructions stressed the importance of writing down only the general core of the retrieved disablers, in order not to lose time because of the writing itself. The generated disablers were scored according to the list of accepted disablers as provided by the raters in our previous studies.

Results and Discussion

We disallowed 5.66% of the generated disablers in the retrieval measure, mainly because they expressed variations of the same idea according to the previous classifications. The Spearman rank-order correlation between the generated number of disablers for the new and old conditionals was rather high [r_s 5 .65, n 5 40, $t(38)$ 5 5.30, p , .001].

Acceptance ratings. We first analyzed the relation between participants' mean MP, DA, MT, and AC acceptance rating in Experiment 1 and the total number of disablers they generated for the eight conditionals in the retrieval task. Spearman rank-order correlations indicated that both MP [r_s 5 2 .35, n 5 40, $t(38)$ 5 2 2.33, p , .03] and MT [r_s 5 2 .40, n 5 40, $t(38)$ 5 2 2.67, p , .02] showed the expected significant negative relation: The more disablers people could retrieve in a limited time