Working Memory Capacity and the Nature of Generated Counterexamples.

Niki Verschueren (Niki.Verschueren@psy.kuleuven.ac.be)
Wim De Neys (Wim.Deneys@psy.kuleuven.ac.be)
Walter Schaeken (Walter.Schaeken@psy.kuleuven.ac.be)
Géry d'Ydewalle (Géry.Dydewalle@psy.kuleuven.ac.be)
Laboratory of Experimental Psychology
University of Leuven
Tiensestraat 102, 3000 Leuven, Belgium

Abstract

This article presents a taxonomic system for generated disablers (based on Elio, 1998) and generated alternatives. Based on the taxonomy, we distinguish three different types of knowledge that are advocated during generation tasks (1) situations that are semantically strongly related to the content of the premises (2) more remote situations and (3) the invalid or low quality counterexamples. Second, we look at the effect of working memory capacity on the nature of generated counterexamples. We found that participants with a high working memory capacity can generate more counterexamples and are flexible in their search process. Participants with low working memory generate less counterexamples and restrict themselves to the first type of counterexamples.

Introduction

Deductive reasoning with causal propositions is one of the core activities of human cognition. The prototypical causal rule is formulated as an 'if-then' sentence. The if-part of the conditional expresses the cause and the then-part contains the effect. The four reasoning problems that are traditionally used to investigate causal reasoning are (1) modus ponens - MP: does the effect follow when the cause is present (2) denial of the antecedent - DA: does the effect follow in absence of the cause (3) affirmation of the consequent - AC: did the cause occur when the effect is observed (4) modus tollens - MT: did the cause occur although the effect did not occur. The answers participants produce to these problems are classically discussed in terms of conditional answers. Schematically, the reasoning problems and answers look as follows. The conditional sentence is: 'If cause, then effect'

	Categorical premise	Conditional answer
MP	The cause occurs.	The effect follows.
DA	The cause does not occur.	The effect does not follow.
AC	The effect occurs.	The cause preceded.
MT	The effect does not occur	The cause did not precede.

Cummins (Cummins, Lubart, Alksnis, & Rist, 1991; Cummins, 1995) found that the tendency to deduce AC and DA is related to the number of alternative causes the reasoner can activate from background knowledge. The number of disabling conditions, on its turn influences the making of MP and MT. Alternative causes is a cause other than the one given, that is capable of evoking the effect. Disabling conditions is an event that can prevent an effect from occurring in the presence of the given cause.

For each of these four reasoning problems Markovits (2000) gives a detailed description of the underlying cognitive mechanism. His theory is based on the mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). The mental model theory assumes that reasoners build internal models representing the premise content, and through manipulation and extension of these models they generate a conclusion. We will briefly discuss Markovits' (2000) account of how the four reasoning problems are solved (applied to causal reasoning).

The mental model theory assumes that (1) reasoners start by representing the content of the conditional sentence in an economical way, for instance '*cause→effect'*. This model represents a possible situation and is often called the initial model. Other possible models of situations are left implicit. When they are asked what follows from the categorical premise reasoners verify whether they can produce a conclusion based on the initial model. In case of MP they can initially conclude that the effect occurs, and for the AC they can conclude that the cause preceded. For the other two reasoning forms, there is no explicit information regarding the absence of effect or cause, so no conclusion can initially be generated. (2) In case of DA and MT, reasoners create explicit models of other possible situations. According to Markovits (2000) the preferred second model is '*no cause→no effect'*. Based on this extra model it is possible to generate an initial conclusion for DA (the effect does not follow) and for MT (the cause did not precede). At this point, a first conclusion is formulated for all four reasoning problems; this conclusion corresponds to the conditional answer. (3) Most reasoners will then validate their initial conclusion by searching for possible counterexamples. For MP and MT, the falsifying model is '*cause →no effect'* (disabler). For AC and DA, the falsifying situation is '*no cause→effect'* (alternative). If a counterexample is found, reasoners become aware that there is more than one conclusion possible, and reject the initial conclusion. When no counterexamples are found, reasoners give conditional answers to all four reasoning forms. Hence, the probability of finding counterexamples informs us about the probability of giving conditional answers.

The probability that reasoners find a counterexample depends on the number of counterexamples that are present in semantic memory. When there are many counterexamples, the probability of retrieving at least one is higher than when there are only few counterexamples. In order to check how many counterexamples reasoners can retrace from memory,

researchers ask participants to generate possible alternatives and/or disablers for a conditional sentence. The number of counterexamples generated in this way reflects the number of counterexamples present in background knowledge, thus reflecting the probability that reasoners find at least one counterexample during reasoning.

Previous research has focussed on specific characteristics of generated counterexamples: the absolute number of counterexamples (see e.g., Cummins et. al., 1991; Cummins, 1995), the salience of counterexamples (Markovits, 2000), and the strength of association between (alternate) causes and a consequent (Quinn & Markovits, 1998). For disablers, Chan and Chua (1994) and De Neys, Schaeken & d'Ydewalle (2001) described the importance of the perceived strength of the connection between cause and effect. Dieussaert, Schaeken, & d'Ydewalle (2002) investigated the differential effect of disablers referring to the item itself or to speaker control.

Another important characteristic is the type of the generated counterexamples. It is possible that some counterexamples are considered to be of greater importance regarding their falsifying strength, than others are. A first step in this research domain is to develop a taxonomy, which enables us to distinguish different types of disablers and alternatives. Elio (1998) has proposed a taxonomy of 'disablers'. Although she constructed this taxonomy from the perspective of belief revision, we consider this taxonomy also useful for research on 'deductive' conditional reasoning. Elio (2001) herself points out that both research area's are complementary, as 'endorsement and entrenchment of a conditional are opposite sides of the same coin'. No taxonomy has been lined out for alternatives. Developing a taxonomy for alternatives will be the first aim of the present study.

Furthermore, we presume that the ability of generating counterexamples is influenced by working memory capacity. Retrieving a counterexample is considered to be a semantic search process (Markovits, Fleury, Quinn & Venet, 1998) and since the efficiency of a semantic search process is linked to working memory capacity (Rosen & Engle, 1997) we deduce that the retrieving of counterexamples is linked to working memory capacity (see also De Neys, et al., 2002). The present study will provide some preliminary data on this topic. Secondly, we will investigate whether there are differences in the nature of generated counterexamples corresponding to differences in working memory capacity.

TAXONOMY

Disablers

Elio (1998) proposed her taxonomy for disablers in the context of belief-change. She first induced a belief-state about the rule by presenting an MP problem and its conditional answer. Then, the participant finds out that this stated conclusion is contradicted by observed facts and is asked to give some sort of rationalisation. Elio (1998; 2002) distinguishes seven categories of disablers. We will illustrate these categories for the sentence: '*If a plant is watered well, the plant stays green'*. A disabler for this sentence explains why the plant doesn't stay green although it is watered well.

The first category contains the *'real' disablers* (1), e.g. 'there is no sunlight'. These answers state that normally the cause produces the effect but in the situation under description there is an extra conditon present which prevents the effect from occuring. Instances of the second category, *demote to default* (2) merely indicate that the given rule is probabilistic in nature, e.g., 'in most cases the plant stays green, but there are exceptions'. The next category contains the *missing enablers* (3), e.g., 'the plant received too little water'. These responses indicate that a condition necessary for the cause to take effect is absent. The fourth category holds *generalisations* (4) of the effect, e.g., 'the plant stays healthy'. The rationale behind this kind of disabler is that the cause produces an effect, but not specifically the effect mentioned in the rule. Another category contains responses that indicate an *invalid relation* (5) between cause and effect; e.g., 'water is not enough for the plant to stay green'. The sixth category contains *exceptional instances* (6), e.g., 'the plant is an oak with brownish leaves'. The rule remains valid, but the participants lists an instance to which the rule exceptionally doesn't apply. The last category contains answers which make reference to *intervening variables or the passage of time* (7), e.g., 'the plant was watered well until last month'. These responses indicate that the cause was indeed followed by the effect, but something happened that cancelled the effect.

Alternatives

In line with the categories proposed by Elio for disablers, we can construct a taxonomy for alternatives. We will use the same sentence to illustrate the different types of alternative causes*, 'If a plant is watered well, the plant will stay green'.* An alternative explains why the plant stays green, even when he is not watered well.

A first category contains the '*real' alternatives* (1), e.g., 'the plant receives a lot of fertilizer'. These are causes, which can also produce the effect, even when the given cause is absent. The second category is called *demote to default* (2), e.g., 'normally the plant needs water to stay green, but not always'. This category contains answers that point out that normally the cause produces the effect, but there are some exceptions, which are not explicitly mentioned. The third category contains *non-missing enablers* (3), e.g., 'the plant needs practically no water'. This category mirrors the missing-enabler category of the disabler taxonomy. The fact that the plant does not need a lot of water is no cause of the plant staying green. It just enables the effect to occur even if the required cause is absent. A fourth category contains the *generalizations* (4), e.g., 'if you take good care of a plant, the plant stays green'. Watering a plant well is an instantiation of the superordinate category 'taking good care of a plant'. The fifth category is the *invalid rule* (5); e.g., 'a plant does not need water to stay green'. This 'alternative' cancels the stated relation between antecedent and consequent. The sixth category contains the *exceptional instances* (6), e.g., 'the plant is a Mexican cactus'. Instances of this category point out that the conditional sentence is valid, but for this

particular example of a plant, the rule does not apply. Finally, the seventh category contains alternatives referring to intervening variables or passage of time (7), e.g., 'after a while the plant learned to live on little water'.

In addition to these 7 parallel categories, we distinguish three extra categories. The first extra category contains answers referring to luck or magic (8), e.g., 'Harry Potter came by and the plant turned green'. The second category contains answers for which the conditional sentence is read in its non-literal meaning (9), e.g., 'the plant sees that other plants receive water and turns green with envy'. A last category is reserved for invalid answers (10), e.g., 'the plant stays green by its photosynthesis', 'the plant does yoga', ... In most experiments where participants are asked to generate disablers or alternatives, these answers are excluded from the analysis. From the perspective of building a taxonomic system, we preferred to put them in a special category. As with Elio's (1998) taxonomy for disablers, we assume that some of these categories have fuzzy boundaries. The category 'luck or magic' is related to 'demote-to-default' and 'exceptional instances'.

The extra three categories are also valid for disablers. They cannot be reduced to one of the seven categories Elio proposed, so for sake of completeness, we will add them to her taxonomy. Since the taxonomy for disablers then fully parallels the taxonomy for alternatives, we can compare the distribution of the answers.

Overall, for disablers as well as for alternatives, we can say that the categories labeled 'disablers' and 'alternatives' contain the '*real' counterexamples*. Instantiations of this category appear to be semantically closely related to the content of the premises. The categories 'demote to default', '(non)missing enabler', 'generalization', 'invalid rule', 'time' and 'exceptional instance' are more *remote*. They either refer to exceptional situations or some of the basic assumptions of the conditional sentence are denied. The categories of 'luck or magic', 'non-literal interpretation' and 'invalid answers' contain counterexamples that can be given to any kind of sentence, regardless of the exact semantic content. We consider these counterexamples to be of *low quality*.

Experiment

Applying the Taxonomy

First of all, we will apply the two taxonomic systems on generated counterexamples. This way we can get some indication of which type background knowledge participants use when asked to produce counterexamples.

Method We used twenty causal 'if-then'-sentences. The sentences covered a broad range of semantic domains. Based on previous research we choose an equal proportion of sentences for the four categories: (1) many disablers and many alternatives, (2) many disablers, few alternatives, (3) few disablers, many alternatives, and (4) few disablers and few alternatives. Our generation task was similar to the one used by Cummins (1995). First, we presented the participants with a causal rule. Subsequently we stated that the cause occurred but it did not produce the effect (disablers) or that the effect occurred in absence of the given cause (alternatives). Participants were then asked to write down as many explanations as possible (maximum 5). It was explicitly mentioned that the given explanations had to be different from the stated cause, different from each other, and that they could only give valid answers, answers such as 'the person came from Mars' are not tolerated.

Sixty-two subjects participated in the experiment as part of course requirements. Thirty-two subjects were given the disabler-generation task, while thirty other subjects received the alternative-generation task. Each participant generated either disablers or alternatives for each of the 20 sentences. The order of the sentences was randomized over participants. The participants were given 15 to 20 minutes to complete the task. For each situation that participants generated, two independent raters determined to which category type the answer belonged. Interrater reliability was .93 for the alternative and .84 for the disabler generation task.

Results and Discussion Both for the alternative and disabler generation task we first divided the sentences into two groups. For one group of sentences (n=10) there are few alternatives or disablers (dis/alt) generated, while in the other group (n=10) there are many dis/alt. The few-group contains sentences for which the total number of generated counterexamples for the sentence is less than the overall mean of all sentences. For the sentences of the many group the number of generated counterexamples for each sentence is higher than the overall mean.

For the alternatives as well as for the disablers, we determined the number of times each category type occurred, this separately for the few and the many sentences. Table 1 gives an overview of the results.

Table 1: Proportion of answers for different categories.

	Disablers		Alternatives	
Category	Few	Many	Few	Many
1. real alt/dis	76.8	81.9	74	94.5
2. demote to default	0.6	3.1	3.6	1.5
3. (non)missing enabler	12.4	7.1	1.9	-
4. generalization	0.6	2.2	2.1	3
5. invalid rule	-	-	-	-
6. exceptional instance	4.1	3.6	10.6	0.1
7. time/ intervening	3.6	1	0.4	-
8. luck/magic	-	0.8	3	0.2
9. non-literal	-	-	0.4	-
10. invalid	1.9	0.2	3.8	0.6
Total N	531	869	470	976

Within the many-group there are relatively more 'real' dis/alt generated then in the few-group. This difference is significant for disablers (p_1=.768, n_1=531 versus p_2=.819, n_1=869, p<.0209) as well as for alternatives (p_1=.74; n_1=470 versus p_2=.945, n_1=976, p<.0001). Additionally, we found that for the few disabler group, more missing enablers are generated than for the many group (dis: p_1= .124, n_1=531 vs. p_2=.071, n_1=869, p=.0008). For alternatives we observe that

in the few group participants more often list exceptional instances than in the many group (p_1=.106, n_1=470 vs. p_2= .01, n_1=976, p<.0001). All other differences between proportions are non-significant.

We assume that when participants are asked to generate dis/alts they start to look for straightforward examples, namely the 'real' dis/alts (category 1). This is because the 'real' counterexamples are semantically strongly related to the content of the conditional sentence. In addition to these 'real counterexamples' participants dispose of another pool of possible counterexamples, namely, the more remote situations. Instantiations of this type are semantically not directly linked to the premise content. They refer to exceptions to the normal situation (category 2,6) or to conversational implicatures that are suspended (category 3,4,7). The assumptions that are normally valid, such as 'promises are kept', 'birds can fly', 'coffee contains caffeine' are examined in order to account for the apparent contradicting premises. In general, this more remote category contains counterexamples sprouting from suspended conversational implicatures (Levinson, 2000).

For some assumptions you find that when the assumption not holds, the relation between antecedent and consequent changes, and can account for the apparent contradiction.

When only few dis/alt can be found, it is harder to find a full range of 'real' counterexamples. As a result, participants search also for the more remote type of disablers and alternatives.

Conclusion The two taxonomic systems can be used to categorize the answers participants give when asked to generate disablers or alternatives. Although some of the presented categories are conceptually related, the raters consistently classified the answers.

By applying the taxonomy we found that more 'real' alternatives and disablers were generated in the many groups than in the few groups. This difference is compensated by a shift to the more remote types. We assume that participants start searching for counterexamples from the pool of 'real' counterexamples, because these counterexamples are semantically close to the content of the premises. In addition, the search can be directed to the more remote categories.

In the second part of this experiment we will investigate whether working memory capacity affects the type of the generated dis/alts.

Working Memory Capacity

Double task experiments showed that working memory capacity puts a constraint on the ability to generate counterexamples (De Neys, Schaeken, & d'Ydewalle, 2002). First, we will investigate the effect of working memory capacity on the *number* of generated counterexamples. We expect that participants with high working memory capacity generate more counterexamples than those with low working memory capacity. We expect this difference to be larger for sentences with few counterexamples. For the many sentences we assume that the difference may be blurred due to a ceiling effect. Second, we will look at the effect of working memory on the *type* of generated counterexamples. Do differences in working memory capacity affect somehow the sort of counterexamples participants come up with?

All first year psychology students had fulfilled a Dutch version of the OSPAN test (La Pointe, & Engle, 1990; Dutch version: De Neys et al., 2002) for measuring working memory capacity. As such, we can link the number and nature of the generated answers to differences in working memory capacity.

Results and Discussion The subjects are divided in three groups depending on their working memory capacity. The high participant group consists of the top third (dis: Min: 37; Max: 54 - alt: Min: 39; Max: 54). The low group contains participants with scores of the bottom third (dis: Min: 18; Max: 24 - alt: Min: 39; Max: 54). Table 2 displays the distribution of the relative proportion of answers.

Participants with high working memory capacity generate more disablers than participants with a low working memory capacity (dis: p_1=.45; n_1=964 versus p_2=.55; n_1=964, p<.0001). This difference is not significant for alternatives.

Table 2: Proportion of generated counterexamples for each category (numbers refer to categories of Table 1). The shaded regions refer to the three types category 1 equals Type 1, category 2 to 7 corresponds to Type 2, categories 8 to 10 are labeled Type 3.

	Disablers						**Alternatives**					
	Low			High			Low			High		
Category	Few	Many	Total	Few	Many	Total	Few	Many	Total	Few	Many	Total
1	82.5	86.4	85	71.7	80.1	76.7	79.4	95.2	90.4	76.7	93	87.7
2	5.8	2.9	3.9	0.9	4.2	2.8	6.1	1	2.6	2.8	1.5	1.8
3	3.9	5.7	5.1	14.6	8.7	11.1	1.8	-	0.6	4.4	-	0.5
4	0.6	1.4	1	0.5	2.9	1.9	-	3.2	2.2	1.1	3.5	2.7
5	-	-	-	-	-	-	-	-	-	0.6	-	0.2
6	3.3	2.5	2.8	5	2.6	3.6	5.5	0.3	1.9	8.9	-	2.9
7	3.3	-	1.2	3.7	0.6	1.9	0.6	-	0.2	-	-	-
8	-	1	0.7	-	0.6	0.4	4.2	-	1.3	2.8	0.5	1.3
9	-	-	-	-	-	-	-	-	-	0.6	0.3	0.4
10	0.6	-	0.2	3.7	0.3	1.7	2.4	0.3	0.9	5	1.5	2.5
Total N	154	279	433	219	312	531	165	375	540	180	371	551

We will now discuss the effects of working memory capacity separately for the sentences with few and many disablers.

A first important result is that the observation that participants with high working memory capacity generate *more disablers* than those with low working memory capacity, is only found on the sentences with few disablers (p_1=.413; n_1=373 vs. p_2= .587; n_1=373, p<.0001). This finding can be explained as follows. For sentences with only few disablers, it is inevitably harder to generate counterexamples than for sentences with many disablers. In general, for sentences with only few disablers most participants quickly run out of inspiration (only in 5% of the trails there were more than 3 disablers given). Because participants with low working memory capacity experience more difficulty in generating disablers, we can expect that their searching process takes more time than that of participants with high working memory capacity. For sentences with many disablers, we assume that there is a ceiling effect. As participants can choose from a large pool of possible counterexamples, the differences in working memory capacity on the generated number of counterexamples does not show.

A second striking finding is that participants with a low working memory seem to restrict themselves to a single type of counterexamples. Participants with a high working memory capacity generate overall more 'real' disablers, this group represents a larger proportion of the generated responses of the participants with low working memory (p_1= .85; n_1=433 vs. p_2= .77; n_1=531, p<.0013). This decrease in 'real' disablers is mirrored by a significant increase in the proportion of disablers of the 'remote' type. Participants with a high working memory score generate more missing enablers (category 3) than participants with low working memory capacity (p_1=.051, n_1=433 versus p_2=.111, n_1=531, p<.0009). As stated above we assume that participants start their search for counterexamples by checking situations, which are semantically related to the content of the premises. In addition, reasoners can check situations that are semantically more remotely related to the premise content. Thus, the results suggest that participants with a high working memory capacity can more easily shift from the straightforward type of counterexamples to the more remote type. Participants with a low working memory capacity are rather conservative in their search for counterexamples. It can be argued that the flexibility to change from one semantic domain to another yields a substantial profit in finding counterexamples. Based on our results we can add that working memory capacity is a crucial mediator of this flexibility.

The significant effects on disablers are paralleled by non-significant trends for alternatives. The absence of any significant working memory effects on alternatives can be explained with reference to the structural difference between the two types of counterexamples. When you are asked to generate a disabler, you have to find a situation in which the effect does not occur in presence of the given cause. The presence of the given cause constitutes an important element of the situation you have to generate. As a result you have to maintain two different propositions, the cause as well as the effect, in memory. For alternatives it is not necessary to maintain the given cause in memory. You just need to look for some alternate causes, and maintain the effect in memory. Markovits argues that for young children, it is harder to search for disablers than to search for alternatives (Jeanveau-Brennan, & Markovits, 1999; Markovits, 2000). We assume that his finding can be generalized to adults. No effect of working memory capacity is observed on alternatives because the generation of alternatives does not challenge working memory capacity in the way that the generation of disablers does.

General Discussion

This article addressed four main issues. First of all, we applied Elio's (1998) taxonomic system for disablers in belief-revision to data gathered in a conditional reasoning perspective. The described system was equally valid for categorizing disablers, as almost all the answers could readily be categorized. Elio (1998) pointed out that belief revision and deductive reasoning are complementary fields of research. This experiment proves that the use of her taxonomic system can be generalized to the domain of conditional reasoning.

Second, we constructed a taxonomic system for generated alternatives. By applying this categorization system we can distinguish different types and categories of alternatives. Although some categories are interrelated, only few answers were subject to discussion. Hence, we conclude that the taxonomy serves its purpose well. In line of Elio's (1998) proposal we suggest that this taxonomic system can also be used for categorizing the alternatives generated in the context of belief-revision. By applying this system researchers are able to shed light on the type of knowledge that is used during the process of belief revision, or in terms of Elio (1998), on the belief-revision operators that people use for resolving everyday contradictions.

Recent research emphasizes the importance of pragmatic and semantic aspects in theories on conditional reasoning (see e.g., Chan & Chua, 1994; Newstead, Ellis, Evans, & Dennis, 1997; Quinn & Markovits, 1998). By outlining two taxonomic systems we provide researchers with an additional methodological weapon for disclosing how the search for counterexamples takes place. These taxonomic systems can also be used for categorizing counterexamples for other types of conditionals than causal ones.

Third, we used the categorization of alternatives and disablers to examine the sort of background knowledge that is advocated during the search for alternatives or disablers. Three broad types of counterexamples were distinguished. The first type contains the 'real' disablers or alternatives. These are descriptions of situations that are semantically close to the stated premises. The second type of answers is of a more 'remote' type. They include answers in which the normal conversational implicatures are suspended. Participants go beyond the usual scheme's (Chan & Chua, 1994) that the premises refer to, in order to find some condition that could not apply (enabler, truthfulness of the speaker, …). A third category contains answers that are

given just to lengthen the list of alternatives. They include answers referring to some magical interference, plain luck, a non-literal reading or just invalid responses. For a possible rule to decide to start looking in another pool, we can refer to the stopping rule proposed by Johnson-Laird (1994); when it gets too hard to generate another imagined situation, participants stop their search (for alternative stopping rules, see Elio, 2002).

Fourth, we looked at the effect of working memory capacity on the taxonomic distribution of the generated disablers and alternatives. We found that reasoners with high working memory retrieve more disablers and are more flexible in their search. They tend to retrieve different types of disablers while reasoners with low working memory capacity are more conservative. Reasoners with low working memory capacity generated more 'real' disablers then reasoners with high working memory capacity. This result suggests that reasoners with low working memory capacity start by searching the pool of semantically related disablers and are conservative in their search. In contrast, reasoners with high working memory capacity are more flexible in redirecting their search.

We like to add that asking participants to generate as many disablers or alternatives as possible could reflect a somewhat different cognitive process than the process active during the validation phase of reasoning (see also Markovits, Fleury, Quinn & Venet, 1998).

In sum, participants with a high working memory capacity can retrieve more counterexamples and are flexible in their search process. Participants with low working memory generate less counterexamples and restrict themselves to counterexamples of the first type. We find that considering qualitative aspects of generated counterexamples provide valuable information on the underlying cognitive search process. In this study we describe three different types of counterexamples. Furthermore, we argue that working memory capacity is not only a crucial mediator for maintaining and searching information but determines also a reasoner's flexibility to search different semantic domains.

Acknowledgments

This work is carried out thanks to the support of the Fund for Scientific Research Flanders.

References

Chan, D., & Chua, F. (1994). Suppression of valid inferences: syntactic views, mental models and relative salience. *Cognition, 53*, 217-238.

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

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

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2000). *Causal conditional reasoning and strength of association: the disabling condition case.* Psychological Report N°: 271. Leuven: University of Leuven, Laboratorium of Experimental Psychology.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2001). Does pure water boil, when it's heated to 100°C?': The associative strength of disabling conditions in conditional reasoning. *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society, 249-245*. Mahwah, NJ: Lawrence Erlbaum Associates.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). *Working memory capacity and causal conditional reasoning*. Manuscript in preparation.

Elio, R. (1997) What to believe when inferences are contradicted. The impact of knowledge type and inference rule. *Proceedings of the Nineteenth Annual* Conference of the Cognitive Science Society, 211-216. Hillsdale, NJ: Lawrence Erlbaum Associates.

Elio, R. (1998). How to disbelieve p=>q: Resolving contradictions. *Proceedings of the Twentieth Meeting of the Cognitive Science Society, 315-320.* Mahwah, NJ: Lawrence Erlbaum Associates.

Elio, R. (2002) *Belief revision and plausible inference*.. Manuscript submitted for publication.

Elio, R., & Pelletier, F.J. (1997). Belief revision as propositional update. *Cognitive Science, 4,* 419-460

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

Johnson-Laird P., Byrne, R., Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99,* 418-439.

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

Markovits, H., (1984). Awareness of the 'possible' as mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology, 75*, 367-376.

Markovits, H. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking and Reasoning, 6,* 335-347.

Markovits, H., Fleury, M., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development, 69*, 742-755.

Newstead, S. E., Ellis, M. C., Evans, J. St. B. T., & Dennis,I. (1997). Conditional reasoning with realistic material. *Thinking and Reasoning, 3*, 49-76.

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: strength of association as a predictive factor for content effects. *Cognition, 68*, 93-101.

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

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

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