Reasoning from an incompatibility: False dilemma fallacies and content effects

Janie Brisson¹ · Henry Markovits¹ · Serge Robert¹ · Walter Schaeken²

Published online: 23 March 2018
© Psychonomic Society, Inc. 2018

Abstract

In the present studies, we investigated inferences from an incompatibility statement. Starting with two propositions that cannot be true at the same time, these inferences consist of deducing the falsity of one from the truth of the other or deducing the truth of one from the falsity of the other. Inferences of this latter form are relevant to human reasoning since they are the formal equivalent of a discourse manipulation called the false dilemma fallacy, often used in politics and advertising in order to force a choice between two selected options. Based on research on content-related variability in conditional reasoning, we predicted that content would have an impact on how reasoners treat incompatibility inferences. Like conditional inferences, they present two invalid forms for which the logical response is one of uncertainty. We predicted that participants would endorse a smaller proportion of the invalid incompatibility inferences when more counterexamples are available. In Study 1, we found the predicted pattern using causal premises translated into incompatibility statements with many and few counterexamples. In Study 2A, we replicated the content effects found in Study 1, but with premises for which the incompatibility statement is a non-causal relation between classes. These results suggest that the tendency to fall into the false dilemma fallacy is modulated by the background knowledge of the reasoner. They also provide additional evidence on the link between semantic information retrieval and deduction.

Keywords Deductive reasoning · Incompatibility · False dilemma · Counterexamples · Information retrieval

Deductive inference, i.e. the ability to draw conclusions from given premises, is a key element of advanced human cognition. It requires generating a conclusion that follows “logically” from the premises, i.e. a conclusion that is necessarily true, provided that the premises were true. When a conclusion follows possibly but not necessarily from the premises, its truth value is uncertain even if the premises are true. Distinguishing between certain and uncertain conclusions is thus a key ability of deductive reasoning. Many logical mistakes consist of being certain about a conclusion that is in fact uncertain. Certain types of discourse manipulation, called informal fallacies, can lead an interlocutor to overlook possible conclusions so that she accepts a favored one as certain. For example, shortly after

9/11, when President Bush declared war on terror, he addressed the Congress by saying, “Either you are with us or you are with the terrorists.” This is an example of an informal fallacy, often found in politics or advertising, called the false dilemma or false dichotomy. This fallacy consists of reducing many options to only two dichotomous ones, thus forcing a choice between them (Hurley, 2014). One version of a false dilemma highlights the two ends of a continuum and discards all of the “in between” situations. Another version, which will be the major focus of this paper, discards alternative options, i.e. situations where a third option is true and the two presented options are false. In the example above, the fallacy is of this latter version and is used as a persuasion technique. It presents an option that one wants to be chosen by the interlocutor (to support the USA in their invasion of Iraq) and another one known as undesirable (to sympathize with terrorism), and lets the interlocutor reason, “I don’t support terrorism, therefore I must support this war.” Of course, alternative options that are neither one nor the other are possible: supporting economic, legal, or educational interventions are all peaceful ways to stand against terrorism.

✉ Janie Brisson
brisson.janie@gmail.com

¹ Departments of Psychology and Philosophy, Université du Québec à Montréal, C.P. 8888, Succ A, Montréal, QC H3C 3P8, Canada

² Department of Psychology, Katholieke Universiteit Leuven, Leuven, Belgium

From a logical point of view, the false dilemma consists in falsely presenting a premise as an exclusive disjunction between two propositions. On one hand, when the fallacy discards “in between” situations, the propositions are presented as being bound by an exclusive disjunction (Either P or Q), which allows only for one proposition to be true and the other one false, while they are bound by an inclusive one (P or Q, or both), which allows for both propositions to be true (Van Eemeren & Grootendorst, 2016). Whether the basic interpretation of a disjunction is inclusive by default (Barrett & Stenner, 1971; Crain & Khlentzos, 2007; Grice, 1989; Kamp & Reyle, 1993; Newstead & Griggs, 1983) or depends on the semantic interpretation of the premise (Bauer & Johnson-Laird, 1993; Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne & Schaeken, 1992) is an open question. However, previous research has shown that disjunctive reasoning can be modulated by the content of the premises, with performances sometimes showing an exclusive interpretation of the disjunction (Quelhas & Johnson-Laird, 2017; Newstead, Griggs, & Chrostowski, 1984; Roberge, 1977). This research suggests that reasoners might fall into the disjunctive version of the fallacy and that this tendency might present content-related variability, but further studies would be needed to investigate this question.

On the other hand, when the fallacy discards alternative options, the two propositions are in fact bound by the incompatibility connective (P is incompatible with Q), also called the NAND connective (Not both P and Q) or the Sheffer stroke. The incompatibility does not allow for both premises to be true at the same time, but is consistent with any other situation. Therefore, both the exclusive disjunction and the incompatibility connective are true when one proposition is false and the other one is true, but the incompatibility connective is also true when both propositions are false (Rautenberg, 2006). A logically equivalent way of describing this version of the false dilemma fallacy would be to say that the two propositions are presented as contradictories, i.e. an opposition where the propositions necessarily have opposite truth values, while in fact they are contraries, i.e. where the propositions could also be simultaneously false. Therefore, when two propositions are incompatible, the truth of one necessarily leads to the falsity of the other, but critically, the falsity of one does not necessarily lead to the truth of the other. In the example above, it would mean that if one supports terrorism, one is against the US war, but if one is against the war, one may or may not be on the terrorists’ side. Being able to fully understand what is entailed by a situation of incompatibility may help to suspend judgment when a situation may seem like a dichotomy, but in reality is not. This ability is an important part of human reasoning in daily life, and will accordingly be the major focus of the following studies.

Deductive reasoning from an incompatibility statement

While many inferential patterns can be generated from the same logical connective, deductive reasoning has most often been investigated through four particular ones. These inferences present two premises and a conclusion. The major premise bears a binary connective between two propositions, the minor premise affirms or denies one of these propositions and the conclusion affirms or denies the other proposition involved in the major premise. Following this pattern, deductive reasoning from an incompatibility statement can be investigated through inferences where a major premise presents two incompatible propositions, combined with one of four possible minor premises that affirms or denies one of the incompatible propositions. Within the four inferential modes generated, two are logically valid and two are invalid. The two valid ones affirm the first or the second proposition in the minor premise. We will refer to them as the Affirm First and the Affirm Second inference forms. The Affirm First inference involves reasoning with the premises “P is incompatible with Q, P is true” and leads to the logically correct conclusion, “Q is false” (e.g. “Being in Montreal is incompatible with being in Paris, John is in Montreal. Therefore, John is not in Paris”). The Affirm Second inference involves the premises “P is incompatible with Q, Q is true” and leads to the logically correct conclusion, “P is false” (“Being in Montreal is incompatible with being in Paris, John is in Paris. Therefore, John is not in Montreal”). The two invalid forms deny the first or the second proposition in the minor premise. We will refer to them as the Deny First and the Deny Second inferences. The Deny First inference involves reasoning with the premises “P is incompatible with Q, P is false.” Although these premises suggest the putative conclusion “Q is true,” this is a fallacy since the premises do not lead to a necessary conclusion. For example, the premises “Being in Montreal is incompatible with being in Paris, John is not in Montreal” suggest that “John is in Paris.” However, since John could be elsewhere, e.g. in Rome, one cannot be certain about the invited conclusion. The Deny Second inference involves the premises “P is incompatible with Q, Q is false,” and also has no certain conclusion.

Thus, according to standard logic, the invited conclusions of the Affirm First and Affirm Second inferences should be endorsed as certain provided that the premises were true, while the invited conclusions of the Deny First and Deny Second inferences should be rejected. The false dilemma fallacy occurs when the invited conclusion for the Deny First and Deny Second inferences are incorrectly accepted as certain. What then determines when this fallacy will be produced? The key concept is the idea that for every pair of incompatible situations (P, Q), there exist others for which both P and Q are false. These situations are thus counterexamples to the false dilemma fallacy and are sometimes referred to as

Tertium Quid or Third options, that is, an option that exposes a dichotomy as false. When an incompatibility is presented as an exclusive disjunction in a discourse manipulation, the goal is for these Third options to be discarded out of hand. However, the aim of the present studies is to investigate under what conditions they can be overlooked even when the major premise is presented explicitly as an incompatibility statement. Our basic hypothesis is that the effects of Third options to the false dilemma fallacy can be understood as a variant of the effect of background knowledge on reasoning that has been documented with respect to conditional reasoning. More specifically, we hypothesized that the more readily Third options can come to mind, the more reasoners will tend to avoid the false dilemma fallacy (for an exposition of the formal links between counterexamples on incompatibility and conditional inferences, see Robert & Brisson, 2016).

Content effects and conditional reasoning

Many studies have shown the great variability of human reasoning with inferences of the same logical form but differing in content (e.g., Cummins, Lubart, Alksnis, & Rist, 1991; Markovits & Vachon, 1990; Thompson, 1994) and a large part of the literature on human reasoning is an attempt to explain such variability. Much focus has been placed on the effects of content on conditional (if-then) reasoning. Such reasoning has been mostly investigated through the four inferential modes explained above. According to the logical definition of the conditional connective, also called the material conditional, two of these inferences are valid and two are invalid. The *Modus Ponens* inference (If P then Q, P is true, therefore Q is true), referred to as MP and the *Modus Tollens* inference (If P then Q, Q is false, therefore P is false) referred to as MT are both valid and lead to necessary conclusions. By contrast, the Affirmation of the Consequent inference (If P then Q, Q is true, therefore P is true) referred to as AC, and the Denial of the Antecedent inference (If P then Q, P is false, therefore Q is false) referred to as DA, are both invalid since their putative conclusion doesn't follow necessarily from the premises.

Content effects with these four forms of inferences have been well documented. Content-related variability can be understood under a general perspective called the “semantic memory framework” (De Neys, Schaeken, & d’Ydewalle, 2002) where the retrieval of stored knowledge impacts reasoning with meaningful premises. The impact of information retrieval on conditional reasoning has been mostly observed through the effect of potential counterexamples to a putative conclusion. For the AC and DA inferences, such counterexamples are alternative antecedents, i.e. antecedents that differ from P but imply the consequent Q. For the MP and MT inferences, counterexamples are disabling conditions, i.e. a condition that prevents the antecedent P from implying the

consequent Q. Many studies have shown that the number of potential counterexamples (Cummins, 1995; Cummins et al., 1991; Thompson, 1994) or the strength of association between them and the premise (De Neys, Schaeken, & d’Ydewalle, 2003; Quinn & Markovits, 1998) determines the rate of approval of the four forms of conditional inferences. For example, with the premise “If a rock is thrown at a window, then the window will break,” reasoners will tend to accept the AC inference (a window is broken, therefore a rock was thrown at it) less often than with the premise “If a finger is cut, then it will bleed” (a finger bleeds, therefore it has been cut). The reason is that the former premise contains many alternative antecedents, such as throwing a chair, a car accident, a tropical storm, etc., that are counterexamples to the putative conclusion, while the latter contains fewer of such antecedents (a finger is crushed, etc.).

As explained above, the Deny First and Deny Second inferences are formal equivalents of the false dilemma fallacy. Just like the AC and DA inferences, these can have a variable number of potential counterexamples, depending on the content of the major premise involved. Thus, based on the effects of content on conditional reasoning, we make the hypothesis that the more Third options are available in memory, the more reasoners will tend to be uncertain about the Deny First and Deny Second inferences. In other words, the more reasoners can think of situations outside of a dichotomy, the less likely they will be to fall into the false dilemma fallacy.

Study 1

In this study, we used premises based on a causal relation expressed as an incompatibility statement. We started with well-known causal conditional premises that have many and few alternative antecedents and translated them into incompatibility statements. For example, we started with the conditional premise “If one has a car accident, then one will be late for a meeting.” This premise has many alternative antecedents (getting up late, being stuck in traffic, helping a friend, etc.) which are counterexamples to the conclusion of the AC and DA inferences. We constructed the following incompatibility statement: “Having a car accident is incompatible with being on time at a meeting.” While presented in a different way, the possibilities that were alternative antecedents for the conditional are now Third options for the incompatibility statement, i.e. a situation where neither P nor Q are the case. This transformation was based on a law of classical logic that states the equivalence between a conditional statement and an incompatibility between the antecedent and the denied consequent, i.e. $(P \rightarrow Q) \equiv (P \uparrow \neg Q)$ (Sheffer, 1913). Simply put in terms of causality, this law expresses the idea that a causal relation is equivalent to one where you cannot have both the cause and the absence of its effect. From this law, we can also translate

alternative antecedents into Third options and disablers into Exceptions to an incompatibility statement. Provided that an alternative antecedent is a situation where P is false and Q is true, the law indicates that this same situation would be one where both premises, i.e. P and Not-Q, are false. This is precisely the definition of a Third option to a false dilemma. Moreover, provided that a disabler is a situation where P is true and Q is false, the same reasoning leads to indicating a situation where both premises, i.e. P and Not-Q, are true as an Exception, which would indeed undermine the incompatibility statement between these premises.

Additionally, to avoid the use of a negation in the incompatibility statements, we identified conditionals for which the consequent is a dichotomous term that can be denied implicitly. In the example above, it means that the “not being late for a meeting” was translated into “being on time for a meeting.” The same procedure was used for causal relations with few alternative antecedents.

The causal relations that were transformed into incompatibilities were chosen to have a variable number of alternative antecedents, and relatively few disablers. We first pretested a set of such incompatibility statements to ensure that the classification into many and few alternatives was maintained.

Pretest

A total of 22 participants took part in the pretest (eight men, 14 women, average age = 33 years, 6 months), were recruited via the online platform Crowdfunder and were native English speakers. Participants were imposed no time limit and took part in the experiment individually.

We constructed 12 incompatibility statements based on causal conditional premises, of which six were chosen to have few alternatives and six to have many alternatives, which was expected to translate into few or many Third options. Moreover, all statements were chosen to have relatively few disabling conditions, which translated into few exceptions to the incompatibility. Half of the participants generated Third options for each of 12 statements, presented in a random sequence. Statements were chosen to have few exceptions, and in order to verify this, the other half generated potential exceptions for the statements.

The Third option generation task was constructed based on the methodology used by Verschueren, Schaeken, De Neys, and d’Ydewalle (2004). We presented an incompatibility statement followed by a situation where P and Q are false. We then asked participants to give as many explanations as possible for this situation, thus making them draw up a list of possible alternatives to the pair of incompatible propositions. Participants were given a limit of five explanations to prevent artificial explanations and fatigue while allowing variation between premises. Below is an example of such a task:

“Rule: Having a car accident is incompatible with being on time at a meeting.

Situation: Paul didn’t have a car accident BUT he was not on time for his meeting.

List as many explanations for this situation as possible (5 maximum).”

The same procedure was used to construct the generation task for exceptions:

“Rule: Having a car accident is incompatible with being on time at a meeting.

Situation: Paul had a car accident BUT he was on time for his meeting.

List as many explanations for this situation as possible (5 maximum).”

Responses for each premise were rated by two judges who were instructed to ignore any paraphrase of an existing response. We then calculated the mean number of Third options and exceptions generated for each of the 12 incompatibility premises. From this, we selected three premises with the highest mean number of Third options and three with the lowest mean number of Third options generated, with all having an equally low level of exceptions. The former were put in the “Many” group and the latter were put in the “Few” group. The mean number of Third options generated in the Many group was significantly higher ($M = 3.78$; $SD = 1.36$) than the mean number of Third options generated in the Few group ($M = 2.47$; $SD = 1.24$), $t(11) = 6.306$, $p < 0.001$. See Appendix A for the selected premises, the conditional version they were initially translated from and the mean numbers of Third options and exceptions for each of them.

Our main hypothesis is that the tendency to make the false dilemma fallacy is related to the number of readily available Third options. In this initial study, we used the causal premises identified in the pretest to construct inference problems with many and few Third options. We predicted that participants would produce more uncertainty responses to the Deny First and Deny Second forms for problems with many alternatives than for those with few alternatives.

Method

Participants and procedure A total of 98 students from a college in Montreal took part in the experiment (60 women, 37 men (one participant failed to indicate gender), average age: 18 years, 0 months, age range: 16–21 years). Each participant was randomly allocated one of the booklets, was told to take as much time as needed, and took part in the experiment individually.

Design and materials Four paper and pencil booklets were constructed. The first one presented three sets of inferential problems based on the major premises with many counterexamples. For each major premise, participants received problems corresponding to the Affirm First, Affirm Second, Deny First and Deny Second inferences. The order of these was randomly determined for each major premise, which were, in order:

1. Having a car accident is incompatible with being on time at a meeting;
2. Being on a plane is incompatible with sleeping well;
3. Drinking coffee in the evening is incompatible with falling asleep easily.

A second booklet presented three sets of inferential problems based on the premises with few counterexamples. For each major premise, we used the same inference order determined in the first booklet. The premises were, in order:

1. Drinking a lot of alcohol is incompatible with being sober;
2. Being in a warm environment is incompatible with feeling cold;
3. A displaced rail is incompatible with the train staying on track.

For each of these two booklets, an alternative version was constructed with the order of the major premises inverted. On the first page of each booklet, participants were asked for their gender, age, and grade level. They were then given the following instructions:

“In the following pages, we are going to show you some rules that you must **suppose to be true**. You have to assume that the rules are always true. For each rule, we are also going to show you some observations. Your task is to select the conclusion that follows logically from the rule and the given observation.”

On the top of each of the following pages, an incompatibility statement was presented. On the same page, four logical problems corresponding to the Affirm First, Affirm Second, Deny First, and Deny Second inferences were presented. For each inference, participants had to choose amongst three possible conclusions. The following is an example of such a statement and a Deny First inference problem:

Suppose that it is always true that:
Having a car accident is incompatible with being on time for a meeting.

For each of the following observations, select the conclusion that follows logically from the rule and the given observation:

Helen didn't have a car accident. One can conclude that:

1. Helen was on time for her meeting.
2. Helen was not on time for her meeting.
3. One cannot conclude whether or not Helen was on time for her meeting.

In summary, Premise-type (Many, Few) was a between-subjects variable and Logical form (Affirm First, Affirm Second, Deny First, Deny Second) was a within-subject variable.

Results and discussion

Preliminary analysis showed that some participants gave unexpected responses to either the valid or the invalid forms. That is, they neither endorsed nor suppressed valid or invalid inferences, but rather endorsed the opposite of the invited conclusion, i.e. Affirm First: “Q is true”; Affirm Second: “P is true”; Deny First: “Q is false”; Deny Second: “P is false”. Of these, three participants produced close to or more than 3 standard deviations from the mean number of unexpected responses (three or more responses out of 12). They were thus eliminated from further analysis. Of these three, two were from the Many condition and one was from the Few condition. All further analyses were conducted on the 95 remaining participants (46 in the Many, 49 in the Few condition).

We then calculated the percentage of logically correct responses out of three inferences for each of the four logical forms (see Table 1). First, a 2 (Validity: Valid, Invalid) $\times$ 2 (Inference-type: First, Second) $\times$ 2 (Premise-type: Many, Few) $\times$ 2 (Order: First, Second) mixed-design ANOVA revealed no significant effect of Order, $F(1, 91) = 0.76, p = 0.39$. We then performed an ANOVA with Validity (Valid, Invalid) and Inference-type (First, Second) as repeated measures and Premise-type (Many, Few) as a between-subjects variable. This gave no main effect of Validity, $F(1, 93) = 0.02, p = 0.9$, Inference-type, $F(1, 93) = 3.6, p = 0.06$, nor Premise-type,

Table 1. Mean percentage of logically correct responses for the four logical forms (Deny First, Deny Second, Affirm First, Affirm Second) by Type (Many, Few) in Study 1 (standard deviations in parentheses)

Type	Deny First	Deny Second	Affirm First	Affirm Second
Many	0.69 (0.42)	0.70 (0.41)	0.59 (0.43)	0.48 (0.44)
Few	0.52 (0.33)	0.49 (0.41)	0.67 (0.33)	0.70 (0.36)

$F(1, 93) = 0.22, p = 0.64$. The results also showed a significant interaction between Validity and Premise-type $F(1, 93) = 6.62, p < 0.05, \text{partial } \eta^2 = 0.066$, a significant three-way interaction between Validity, Inference-type and Premise-type $F(1, 93) = 4.964, p < 0.05, \text{partial } \eta^2 = 0.051$ and no significant interaction between Inference-type and Premise-type, $F(1, 93) = 2.7, p = 0.11$, nor between Validity and Inference-type, $F(1, 93) = 0.56, p = 0.46$.

Post hoc comparisons used the Tukey procedure with $p = 0.05$. We first analyzed the Validity $\times$ Inference-type interaction. For the invalid forms, this showed that the number of logically correct responses was greater for the Many, $M = 0.70$; $SE = 0.06$, than for the Few premises, $M = 0.51$; $SE = 0.06$. The analysis of the three-way interaction showed that this difference was maintained between the Many, $M = 0.69$; $SE = 0.06$, and Few premises, $M = 0.52$; $SE = 0.05$, for the Deny First form as well as between the Many, $M = 0.70$; $SE = 0.06$, and Few premises, $M = 0.49$; $SE = 0.06$, for the Deny Second form. For the valid forms, there was no significant difference between the number of logical responses for the Many and the Few premises. However, analysis of the three-way interaction revealed that the difference was significant for the Affirm Second form, $M = 0.48$; $SE = 0.06$; $M = 0.70$; $SE = 0.06$, but did not reach significance for the Affirm First form.

The results of this initial study are generally consistent with our hypothesis. They show that the tendency to accept the invited conclusion for the two invalid forms, and thus fall into the false dilemma fallacy, is directly linked to the relative number of Third options. When participants have access to fewer numbers of such options, they show an increased tendency to accept the invited conclusion when this is not merited.

The results also showed an unexpected content effect for the Affirm Second form, which was rejected more often for the Many premises than for the Few premises. This is surprising since both the Many and the Few premises were chosen to have equally low numbers of exceptions. The suppression effect literature on conditional reasoning indeed indicates that such content effects on valid inferences are primarily driven by disabling conditions rather than alternative antecedents (Byrne, 1989, 1991, Cummins, 1995, Cummins et al., 1991, Thompson, 1994). Accordingly, suppression of the valid incompatibility inferences should be driven by exceptions rather than third options. However, previous research on conditional reasoning has observed an indirect impact of alternatives on the suppression of valid inferences. Some studies have shown that retrieval of alternative antecedents to conditional premises may induce retrieval of disabling conditions. Notably, the generation of alternative antecedents has been associated with more rejection of the valid MP and/or MT inferences in both children (Janveau-Brennan & Markovits, 1999) and

adults (Markovits & Potvin, 2001; De Neys et al., 2002). Similarly, the retrieval of Third options for the Many incompatibility premises might have facilitated the retrieval of exceptions. This explanation, however, remains open for debate since these suppression effects on valid inferences have not been replicated when alternative antecedents were provided explicitly to participants (De Neys, Schaeken, & d'Ydewalle, 2003). Nonetheless, the suppression effects found in this study were unclear and minor and did not affect support for our main hypothesis. Possible suppression effects on the valid incompatibility inferences thus fall outside the scope of this paper and should be investigated in further studies.

Study 2A

The results of the first study suggest that reasoning from incompatibility statements uses the same basic semantic retrieval processes that are responsible for content effects in conditional reasoning. However, the causal premises used in this study, while presented as incompatibilities, might have triggered a causal interpretation which in turn could have generated a retrieval process linked to the basic conditional relation. In that case, these results could be seen as a simple replication of the well-known content effects on conditional reasoning. We thus decided to replicate these results using premises that were not causal.

In order to do this, we started by constructing incompatibility statements using categories designed to produce variable numbers of Third options. We first constructed incompatibility statements of the form “X is A is incompatible with X is B”, where A and B are base-level categories derived from the same parent category. For example, dogs and horses are both animals. The corresponding incompatibility would be that “For an animal, being a dog is incompatible with being a horse.” We constructed an initial set using parent categories which allowed many possible base-level categories, which we refer to as Broad statements. We then constructed a set of Reduced statements, where the Reduced parent was chosen to have relatively few base-level instances, e.g., polar animals. Finally, we constructed a set of Close to binary statements, for which the parent was an action category that allowed almost no other options, e.g., from the parent voting on a bill, which allows voting for or against or abstaining. (See Appendix B for the parent and base-level categories for all of the incompatibilities that were generated by type.) It should also be noted that, while minimized, potential exceptions to the rule could be generated from the premises used in Study 1. In this study,

premises were constructed in such a way that the generation of exceptions would require the construction of imaginary possibilities, like an animal that is both a dog and a horse. These premises thus allowed for no realistic exceptions.

Pretest

Participants were asked to write as many Third options as they could for each of 13 pairs of base level categories.

A total of 19 students (five men, 14 women, average age = 21 years, 11 months) were examined for the pretest. All participants were native French speakers and volunteers, and were students at the Université du Québec à Montréal. Each participant was randomly allocated one of the booklets, was told to take as much time as needed and took part in the experiment individually.

Two paper and pencil booklets were constructed. On the first page of each booklet, participants were asked for their gender, age and grade level. They were then given the following instructions (translated from the original French):

“In the following pages, we will ask you to provide as many elements as possible belonging to a given category. You must answer the questions in the presented order and you cannot go back to the questions you have previously answered.”

On the top of each of the following pages, a short context was presented into which a pair of subcategories belonging to a particular parent category was presented. Participants were then asked to write down as many Third options as they could. The following is an example of such context for the “animal” category (translated from the original French):

“In an encyclopedia, we learn about animals. An animal can be a dog or a horse. On the following lines, make a list of all the animals that are neither a dog nor a horse. Write down everything that comes to mind.”

The rest of the page was filled with blank lines on which participants had to write down their answers.

On the first version of the booklet, questions were asked in the following order: four questions of the Close to binary content type, five questions of the Reduced type and four questions of the Broad type. A second version of the booklet was constructed in which the order of the questions was inverted.

Responses for each item were rated by two judges who were instructed to ignore any paraphrase of an existing response. We then calculated mean numbers of Third options generated for each item (see Appendix B). Inspection of these results showed that the German city question generated a very

low mean number of Third options ($M=1.07$; $SD=1.16$), probably due to participants’ lack of knowledge. This left four items in each of the three categories, which were all used in Study 2A.

We first examined differences in the total numbers of Third options generated as a function of Category-type. We performed an ANOVA with the total number of Third options generated as a dependent variable with Category-type (Broad, Reduced, Close to binary) as a repeated measure and Order as a between-subjects independent variable. This gave only a significant main effect of Category-type, $F(2, 30) = 21.33$, $p < 0.001$, $partial\ eta^2 = 0.587$. Contrast analyses with Bonferroni corrections showed that, as expected, the total mean number of Third options generated for the Broad type ($M = 44.76$; $SD = 33.68$) was greater than the total number for the Reduced type ($M = 11.71$; $SD = 6.17$), $p < 0.01$, which was in turn greater than the total number for the Close to binary type ($M=3.06$; $SD = 2.59$), $p < 0.001$.

The items retained after the pretest can thus be classified as a function of the relative availability of Third options. If the retrieval model generalizes to these kinds of premises, then we can predict that uncertainty responses to the invalid forms should vary by category. Specifically, we predict that the number of uncertainty responses to the Deny First and Deny Second forms will be greater when premises come from the Broad category than when they come from the Reduced category, which in turn will be greater than for premises from the Close to binary category.

Method

Participants and procedure A total of 237 students (97 men, 140 women, average age: 22 years, 8 months, age range: 17–52) participated in this study. All participants were native French speakers and volunteers and were recruited in colleges or universities in Montreal. Each participant was randomly allocated one of the booklets, was told to take as much time as needed and took part in the experiment individually.

Design and materials Six paper and pencil booklets were constructed. Each booklet presented four sets of inferential problems, which consisted of the Affirm First, Affirm Second, Deny First, and Deny Second inferences. Note that given that no effect of Order was found in Study 1, we kept this order for each problem set in all six booklets. Problem sets for the first booklet were based on major premises that come from the Broad category. The major premises were, in order (translated from the original French):

1. For a vegetable, being a broccoli is incompatible with being a pepper;
2. For an animal, being a dog is incompatible with being a horse;

3. For a person, being in Montreal is incompatible with being in Paris;
4. For a fruit, being a grape is incompatible with being a strawberry.

A second booklet presented four sets of inferential problems based on the premises from the Reduced category. These were, in order (translated from the original French):

1. For a root vegetable, being a potato is incompatible with being a carrot;
2. For a polar animal, being a polar bear is incompatible with being a penguin;
3. For a fruit with pit, being a peach is incompatible with being a cherry;
4. For a dessert of the day in a restaurant; being the chocolate cake is incompatible with being the lemon pie.

Finally, a third booklet presented four sets of inferential problems based on the premises from the Close to binary category. These were, in order (translated from the original French):

1. For a player at the “heads and tails” game, betting on tails is incompatible with betting on heads;
2. For a driver at a fork on the road, taking the right road is incompatible with taking the left road;
3. For a player at the “even and odd” game, betting on an even number is incompatible with betting on an odd number;
4. For a person who votes on a bill, voting for the bill is incompatible with voting against the bill.

For each of these three booklets, an alternative version was constructed where the order of the major premises was inverted. On the first page of each booklet, participants were asked for their gender, age and grade level. They were then given the following instructions (translated from the original French):

“In the following pages, you will be presented statements that you must **suppose to be true**. Your task is to select the conclusion that follows logically from the given statements.”

On the top of each of the following pages, a short context was presented into which the incompatibility was presented. On the same page, four logical problems were presented in the following order: Affirm First, Affirm Second, Deny First, Deny Second. For each problem, participants had to choose amongst three possible conclusions. The following is an example of

such a context and an Affirm First inference problem in the Many Third options condition:

In an encyclopedia, we learn about animals. An animal cannot be both a dog and a horse at the same time. In other words, being a dog is incompatible with being a horse.

Suppose it is always true that:

Being a dog is incompatible with being a horse.

An animal is a dog. One can conclude that:

1. This animal is a horse.
2. This animal is not a horse.
3. One cannot conclude if this animal is a horse or not.

In summary, Premise-type (Broad, Reduced, Close to binary) is a between-subjects variable and logical form is a within-subjects variable.

Results and discussion

Preliminary analysis showed that seven participants failed to answer a majority of the problems. They were then eliminated from further analysis. With the remaining 230 participants, we calculated the percentage of logically correct responses (out of four inferences) for the four logical forms (see Table 2).

A 2 (Validity: Valid, Invalid) $\times$ 2 (Inference-type: First, Second) $\times$ 2 (Premise-type: Many, Few) $\times$ 2 (Order: First, Second) mixed-design ANOVA revealed no significant effect of Order, $F(1, 223) = 0.03$, $p = 0.86$. We then performed an ANOVA with Validity (Valid, Invalid) and Inference-type (First, Second) as repeated measures and Premise-type (Broad, Reduced, Close to binary) as a between-subjects variable. This gave main effects of Validity $F(1, 226) = 107.26$, $p < 0.001$, $\text{partial } \eta^2 = 0.322$ and Premise-type, $F(2, 226) = 14.57$, $p < 0.001$, $\text{partial } \eta^2 = 0.114$, and no main effect of Inference-type, $F(1, 226) = 0.05$, $p = 0.83$. The results also showed a significant interaction between Validity and Premise-type, $F(1, 226) = 107.26$, $p < 0.001$, $\text{partial } \eta^2 = 0.322$ and no significant interaction between Inference-type and Validity, $F(1, 226) = 0.2$, $p = 0.65$, Inference-type and Premise-type, $F(1, 226) = 1.76$, $p = 0.18$, nor between Validity, Inference-type and Premise-type, $F(1, 226) = 0.78$, $p = 0.46$. This showed that participants had greater accuracy on the valid, $M = 89.67$, $SE = 1.42$, than on the invalid inferences, $M = 60.88$, $SE = 2.52$.

Table 2. Mean percentage of logically correct responses for the four logical forms (Deny First, Deny Second, Affirm First, Affirm Second) by Type (Broad, Reduced, Close to binary) in Study 2A (standard deviations in parentheses)

Premise-type	Deny First	Deny Second	Affirm First	Affirm Second
Broad	0.80 (0.34)	0.80 (0.35)	0.90 (0.22)	0.89 (0.24)
Reduced	0.66 (0.41)	0.67 (0.40)	0.83 (0.28)	0.86 (0.27)
Close to binary	0.37 (0.41)	0.35 (0.40)	0.95 (0.11)	0.94 (0.16)

Post hoc comparisons used the Student-Newman-Keuls procedure with $p = 0.05$. Note that, when three means are compared, as is the case here, this procedure holds the familywise error rate to .05 (Howell, 2012). We first analyzed the main effect of Premise-type. This showed that the number of logically correct responses was greater for the Broad premises, $M = 0.85$; $SE = 0.03$, than the Close to Binary premises, $M = 0.65$; $SE = 0.03$, $p(q[3, 226] = 0.17) < 0.05$. The difference between the number of logically correct responses between the Broad and Reduced premises, $M = 0.76$; $SE = 0.03$ and between the Close to binary and Reduced premises did not reach significance, $p(q[2, 226] = 0.12) > 0.05$. We then analyzed the Validity $\times$ Premise-type interaction. This showed the predicted pattern. For the invalid inferences, the number of logically correct responses was greater for the Broad premises, $M = 0.80$; $SE = 0.04$, than for the Reduced premises, $M = 0.67$; $SE = 0.05$, which was greater than the number for the Close to binary premises, $M = 0.36$; $SE = 0.27$, $p(q[2, 226] = 0.12) < 0.05$. For the valid inferences, there was no significant difference in the number of logically correct responses between the Broad, Reduced, and Close to binary premises.

These results provide a clear replication of the content effects found in Study 1. That is, the fewer Third options are available, the more reasoners tend to be certain about the Deny First and Deny Second inferences, i.e. the greater the tendency to make the False dilemma fallacy. Moreover, the premises used in Study 1 represented a causal relation, which could have facilitated a conditional interpretation of the statements. We tried to minimize this possibility here by using categorical premises bound by a commutative relation, i.e. that involved no directionality nor entailment. Again, this suggests that the retrieval process found in conditional reasoning is also at work in deductive reasoning from an incompatibility statement. Moreover, in contrast to Study 1, we did not find evidence for a content effect with the valid inferences. This was to be expected since the premises we used allowed for no realistic exceptions. This also supports the possibility that the retrieval of such exceptions underlied the suppression of valid inferences observed in Study 1. However, the relationship between the invalid inferences

and the effects of content on the suppression of valid inferences should be investigated in further studies.

Interestingly, participants performed better overall on the valid, $M = 0.90$, $SE = 0.01$, than on the invalid forms, $M = 0.61$, $SE = 0.03$, while no such difference was found in Study 1. A possible explanation for this could be that the invalid inferences contained a negated premise while the valid ones did not. In fact, previous research has found that disjunctive inferences are harder with negative categorical premises (Johnson-Laird et al. 1992; Roberge, 1976). Our results suggest a similar phenomenon in reasoning from an incompatibility.

Finally, it should be noted that pragmatic interpretation of the premises from the Close to binary content types might explain some of the high levels of False dilemma fallacies observed with these items, over and above the number of Third options. These premises being action categories, the situations may have been interpreted as ones where it is better to act rather than to abstain. For example, with the premise “Voting for a bill is incompatible with voting against it,” some participants may have interpreted this situation as one where it is better to play a role in the decision and vote, thus favoring a choice between the two propositions. However, in the Broad and Reduced premises, which are very similar object categories, this type of pragmatic effect is unlikely to account for the variation between these content types.

Study 2B

Studies 1 and 2 showed that the number of Third options available underlies the endorsement of invalid inferences with an incompatibility statement, which supports our hypothesis on the effect of content on the False Dilemma fallacy. However, one potential limitation to these studies might be that the formulation “P is incompatible with Q” might have been ambiguous to some participants and might have generated different interpretations of the major premises. In the presentation of these statements, we tried to explain the meaning of an incompatibility by using the logically equivalent conjunctive formulation “P and Q cannot be true at the same time.” An interesting question that arises from this is the extent to which the Conjunctive

and Incompatible formulations might produce different interpretations of the major premise. One task that has been frequently used to examine interpretations of statements is the truth table task (Evans & Over, 2004). This task presents subjects with a statement involving two terms, followed by the four combinations of true or false terms. Participants must indicate whether each of the combinations make the statement true, false or neither true nor false. Patterns of responses correspond to different forms of interpretation. We thus constructed truth table tasks with a sample of the premises used in Study 2A. Following an incompatibility statement, the four combinations of true or false forms of the premises were presented: $P \& Q$, $\text{not-}P \& Q$, $P \& \text{not-}Q$, $\text{not-}P \& \text{not-}Q$. Each task was presented with an Incompatible or Conjunctive formulation.

Method

Participants and procedure A total of 100 students (51 men, 49 women, average age: 21 years, 8 months, age range: 17–40) participated in this study. All participants were native French speakers and volunteers and were recruited in colleges or universities in Montreal. Each participant was randomly allocated one of the booklets, was told to take as much time as needed and took part in the experiment individually.

Design and materials Four paper and pencil booklets were constructed. Two of these presented statements of incompatibility used the Incompatible formulation, while two used the Conjunctive formulation. On the first page of each booklet, participants were asked for basic demographic information and were presented with the following instructions:

“In the following pages, you will be presented with various rules. For each rule, you will be presented with different situations. For each of these, you must indicate whether the situation shows that the rule is true, or that the rule is false, or whether it does not show that the rule is true or false. Indicate **true** if the situation shows that the rule is true, **false** if the situation shows that the rule is false, or **one cannot know** if the situation does not allow knowing if the rule is true or false.”

A first booklet was constructed for the Incompatible condition. Six statements of incompatibility between two propositions were presented, with each statement on the top of a new page. These statements were taken from Study 2A. Two of these were from the Close to binary category, which were, in order:

1. For a player at the “heads and tails” game, betting on tails is incompatible with betting on heads;
2. For a person who votes on a bill, voting for the bill is incompatible with voting against the bill.

The two following statements came from the Reduced category and were, in order:

3. For a polar animal, being a polar bear is incompatible with being a penguin;
4. For a fruit with pit, being a peach is incompatible with being a cherry.

These were in turn followed by two statements from the Broad category, which were, in order:

5. For a vegetable, being a broccoli is incompatible with being a pepper;
6. For a person, being in Montreal is incompatible with being in Paris.

Below each statement, four specific cases were presented and corresponded to all four combinations of affirming or negating either the first or the second proposition. The order of these was randomly determined for each premise. Three choices were given for each combination. For example, following the statement “For a player at the “heads and tails” game, betting on tails is incompatible with betting on heads,” the following statement, combining a false and a true proposition, was presented with the three options directly following it:

Julien bets on heads and doesn’t bet on tails.

This situation shows that the rule is:

1. True
2. One cannot know
3. False.

The three other situations were:

1. Louis doesn’t bet on heads and bets on tails.
2. Martine doesn’t bet on heads and doesn’t bet on tails.
3. Anne bets on heads and bets on tails.

A second booklet was constructed for the Conjunctive condition. This booklet was identical to the first one, excepted that the statements were presented with a Conjunctive formulation. The two statements from the Close to Binary category were:

1. A player at the “heads and tails” game cannot both bet on tail and bet on head;

2. A person who votes on a bill cannot both vote for the bill and vote against the bill.

The two following statements were from the Reduced category:

3. A polar animal cannot be both a polar bear and a penguin;
4. A fruit with pit cannot be both a peach and a cherry.

Which were followed by two statements from the Broad category:

5. A vegetable cannot be both a broccoli and a pepper;
6. A person cannot be both in Montreal and in Paris.

A second version of each booklet was constructed where the order of the premises was inverted.

Results and discussion

We first examined patterns of responses for each of the six statements. Overall, the most common patterns corresponded to two forms of incompatibility interpretations (these combined for 57% of total responses).

Incompatibility This required judging that the P & Q statements made the rule false while the not-P & Q, P & not-Q, and not-P & not-Q statements made the rule true.

Defective incompatibility This required judging that the P & Q statements made the rule false, while the not-P & Q and P & not-Q statements were judged as making the rule true or considered to be irrelevant and the not-P & not-Q statements was judged as irrelevant.

It should be noted that the same kinds of defective interpretations have been found with truth table tasks using conditional statements. The only other relatively frequent pattern that could be interpreted was the exclusive disjunction, which accounted for less than 8% of total responses.

Exclusive disjunction This required judging that the P & Q and not-P & not-Q statements made the rule false and the not-P & Q and P & not-Q statements made the rule true.

Finally, most of the other patterns were uninterpretable and these were grouped together in an “other” category along with a few other interpretable patterns which corresponded to fewer than 5% of total responses. Table 3 shows the relative frequency of these patterns as a function of Premise-type and Formulation.

In order to examine whether the two formulations differed in the extent to which they generated incompatibility interpretations, we counted up the number of

incompatibility interpretations generated for the two premises for each of the three premise types. We then performed an ANOVA with the number of incompatibility interpretations as dependent variable, with Premise-type as repeated measure and Formulation as a between-subjects variable. This showed nonsignificant effects for both Formulation, $F(1, 98) = 1.98$, $p = .16$ and the Formulation $\times$ Premise-type interaction, $F(2, 97) = .01$, $p = .98$. There was however a significant effect of Premise-type, $F(2, 97) = 12.69$, $p < .001$. Post hoc comparisons were made using paired sample t-tests with a Bonferroni correction. This showed that there was a significant difference in the mean number of incompatibility interpretations between the Broad, $M = 1.31$, $SE = 0.07$, and the Reduced premise types, $M = 0.98$, $SE = 0.09$, $t(99) = 5.06$, $p < 0.001$, with neither showing a significant difference with the Close to Binary premise types, $M = 1.17$, $SE = 0.08$, $t(99) = 1.71$, $p = 0.09$ and $t(99) = 2.07$, $p = 0.04$, respectively.

Thus, while there were differences in the extent to which the different premise types generated incompatibility interpretations, the use of the different formulations did not by themselves produce significant differences in this respect.

General discussion

The False dilemma fallacy is a common error of reasoning. It can be seen as wrongly interpreting a situation as a dichotomy between two options while in fact they are more loosely bound by a relation of incompatibility. This transformation is a rhetorical device that puts the emphasis on the fact that both options cannot be true at the same time, while placing the possibility that both could be false into the background. The strength of this effect can be seen by the results of both studies, which show overall levels of fallacious thinking as high as 65%, even with the kinds of emotionally neutral items used.

Moreover, the results showed a great deal of variability in the extent to which participants made the fallacy. Our key hypothesis here is that a major component of this variability is the result of a retrieval failure whose probability is related to the amount of information stored in semantic memory. In other words, a key factor in understanding the strength of the False dilemma fallacy is the role of semantic memory processes in reasoning (see Markovits, 2014 for a review). In the case of reasoning from an incompatibility, we have argued that the critical information concerns potential Third options, which are cases where neither of the two presented ones are true. The semantic memory model allows

Table 3. Interpretations of statements for each premise type (Broad, Reduced, Close to Binary) as a function of formulation (Incompatible, Conjunctive)

Premise type	Formulation	Interpretation			
		Incompatibility	Defective Incompatibility	Exclusive disjunction	Other
Close to Binary	Incompatible	10	44	15	31
	Conjunctive	28	35	12	25
Reduced	Incompatible	8	36	4	52
	Conjunctive	14	40	9	37
Broad	Incompatible	20	41	2	37
	Conjunctive	28	42	3	27

the general prediction that the extent to which people make the False dilemma fallacy will be inversely related to the numbers of potential Third options that are readily available in semantic memory.

The results of these studies are consistent with this prediction. Study 1 and 2A show that the tendency to be uncertain about the invalid forms grows with the number of available Third options in the major premise. In Study 1, we translated causal premises into incompatibility statements with many and few Third options. As predicted, the endorsement of the Deny First and Deny Second inference forms was greater when premises contained few Third options than when they contained many Third options. In Study 2A, we replicated these effects using premises with non-causal content, i.e. where no directionality is involved. We constructed premises that stated an incompatibility relation between two base-level categories belonging to a parent category. As regards the number of base-level categories they contained, each parent category was broad (many base-level categories), reduced (moderate number of base-level categories), or close to binary (close to two base-level categories). We used these categories to construct incompatibility statements where each base-level category outside of the statement would be a Third option. We predicted that the tendency to draw the Deny First and Deny Second invalid inferences would be inversely proportional to the number of available Third options. Again, our results showed the predicted pattern: endorsement of these inferences was highest when premises came from the Close to binary type, then decreased when premises came from the Reduced type and decreased even more when premises came from the Broad type.

The presented studies are in line with research on content effects in conditional reasoning. It has been shown that the endorsement of the AC and DA inferences decreases when more counterexamples are

available. It should be noted that two major classes of theories provide an account for this type of content-related variability in conditional reasoning. Mental models theory (Johnson-Laird, 2001; Johnson-Laird & Byrne, 2002) focuses on the use of information in order to generate potential counterexamples, while probabilistic theories consider that people's inferences generate estimations of the likelihood of a given conclusion (e.g., Evans, Over, & Handley, 2005; Oaksford & Chater, 2007). However, both theories adopt the assumption that the core mechanisms of human reasoning are highly sensitive to the semantic properties of the problem rather than being primarily based on formal rules analogous to those of classical logic (Braine, 1978; Rips, 1983). The present studies were conducted within the general perspective that reasoning is affected by memory processes, and adopt a neutral position as regards to the specific debate on the nature of human reasoning. The present results attempt to extend the scope of the semantic memory framework by examining a new type of inference. To our knowledge, these studies are the first to show the content effects on reasoning from an incompatibility statement. This suggests that the impact of the retrieval process on reasoning with meaningful premises goes beyond the scope of conditional reasoning and is a key component in the tendency to draw the False dilemma fallacy. When few Third options were available, reasoners tended to treat an incompatibility statement as a dichotomous one, which is precisely the desired outcome when one uses the false dilemma fallacy to force a choice between two selected options. Our findings suggest that reasoners are vulnerable to such manipulations, more so when few instances of such options can be retrieved from memory. Consequently, users of the false dilemma fallacy, like politicians or advisers, might take advantage of their audience's lack of knowledge about a topic to persuade people to agree with them while other options are possible.

Appendix A

Table 4. Mean number of Third options and Exceptions generated (maximum of 5) for each premise (translated from the original French) by Type (Many, Few)

Many Third options	Third options	Exceptions
Having a car accident is incompatible with being on time at a meeting. Conditional version: If one has a car accident, then one will be late at a meeting.	4.00	1.78
Being on a plane is incompatible with sleeping well. Conditional version: If one is on a plane, then one will have poor sleep.	3.75	1.89
Drinking coffee in the evening is incompatible with falling asleep easily. Conditional version: If one drinks coffee in the evening, then it will be harder to fall asleep.	3.58	1.89
Few Third options	Third options	Exceptions
Drinking a lot of alcohol is incompatible with being sober. Conditional version: If one drinks a lot of alcohol, then one will be drunk.	2.42	1.11
Being in a warm environment is incompatible with feeling cold. Conditional version: If one is in a warm environment, then one will feel warm.	2.5	1.78
A misplaced rail is incompatible with the train staying on track. Conditional version: If a rail is misplaced, then the train will derail.	2.5	1.44

Appendix B

Table 5. Mean number of Third options generated for each parent category and its base-level categories (translated from the original French) by Type (Broad, Reduced, Close to Binary)

Broad		
Parent category	Base-level categories	Third options
Vegetables	Broccoli; Pepper	8.41
Animals	Dog; Horse	10.71
World cities	Montreal; Paris	16.47
Fruits	Grape; Strawberry	9.18
Reduced		
Parent category	Base-level categories	Third options
Root vegetables	Potato; Carrot	2.29
Polar animals	Polar bear; Penguin	2.25
German cities	Berlin; Hamburg	1.07
Fruits with pits	Peach; Cherry	2.59
Dessert of the day in a restaurant	Chocolate cake; Lemon pie	5.65
Close to Binary		
Parent category	Base-level categories	Third options
Bet on “heads and tails”	Heads; Tails	0.75
Paths at a fork (road)	Right; Left	1.18
Bet on “even, odd” dice game	Even; Odd	0.33
Vote on a bill	For; Against	1.11

References

- Barrett, R. B., & Stenner, A. J. (1971). The myth of the exclusive 'or'. *Mind*, 80(317), 116–121.
- Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological science*, 4(6), 372–378.
- Braine, M. D. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85(1), 1–21.
- Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31(1), 61–83.
- Byrne, R. M. (1991). Can valid inferences be suppressed? *Cognition*, 39(1), 71–78.
- Crain, S., & Khlentzos, D. (2007). Is logic innate? *Biolinguistics*, 2(1), 024–056.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, 23 (5), 646–658.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & cognition*, 19 (3), 274–282.
- De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & cognition*, 30 (6), 908–920.
- De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*, 31(4), 581–595.
- De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology*, 15 (2), 161–176.
- Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.
- Evans, J. S. B., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, 112(4), 1040–52.
- Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.
- Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.
- Hurley, P. (2014). *A concise introduction to logic* (pp. 165–166). Nelson Education.
- Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology*, 35(4), 904.
- Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological review*, 99(3), 418.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale: Erlbaum.
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, 5(10), 434–442.
- Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological review*, 109(4), 646.
- Kamp, H., & Reyle, W. (1993). From discourse to logic: Introduction to model theoretic semantics of natural language, formal logic and discourse representation theory. Dordrecht: Kluwer.
- Markovits, H. (2014). Conditional reasoning and semantic memory retrieval. In A. Feeney, V. Thompson (Eds.). *Reasoning and memory*, Hove: Psychology Press.
- Markovits, H., & Potvin, F. (2001). Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals. *Memory & Cognition*, 29(5), 736–744.
- Markovits, H. & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology*, 26 (6), 942.
- Newstead, S.E., & Griggs, R.A. (1983). The language and thought of disjunction. In Evans, J.St.B.T. (Ed.) *Thinking and reasoning: psychological approaches*. (pp. 76–106.) London: Routledge, Chapman & Hall.
- Newstead, S. E., Griggs, R. A., & Chrostowski, J. J. (1984). Reasoning with realistic disjunctives. *The Quarterly Journal of Experimental Psychology Section A*, 36(4), 611–627.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.
- Quelhas, A. C., & Johnson-Laird, P. N. (2017). The modulation of disjunctive assertions. *The Quarterly Journal of Experimental Psychology*, 70(4), 703–717.
- 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 (3), B93-B101.
- Rautenberg, W. (2006). *A concise introduction to mathematical logic* (pp. 2–3). Springer Science & Business Media.
- Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, 90(1), 38–71.
- Roberge, J. J. (1976). Reasoning with exclusive disjunction arguments. *The Quarterly Journal of Experimental Psychology*, 28(3), 419–427.
- Roberge, J. J. (1977). Effects of content on inclusive disjunction reasoning. *The Quarterly Journal of Experimental Psychology*, 29(4), 669–676.
- Robert, S., & Brisson, J. (2016). The Klein group, squares of opposition and the explanation of fallacies in reasoning. *Logica Universalis*, 10(2–3), 377–392.
- Sheffer, H. M. (1913). A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society*, 14 (4), 481–488.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & cognition*, 22 (6), 742–758.
- Van Eemeren, F. H., & Grootendorst, R. (2016). *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. Routledge.
- Verschueren, N., Schaeken, W., De Neys, W., & d'Ydewalle, G. (2004). The difference between generating counterexamples and using them during reasoning. *The Quarterly Journal of Experimental Psychology Section A*, 57 (7), 1285–1308.