Journal of Pragmatics 197 (2022) 137–158

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

If and only if people were logical! The effect of pragmatic enrichment on reasoning with abstract and realistic materials

Joanna Blochowiak [a,*], Thomas Castelain [b], Odir Antonio Rodriguez-Villagra [c], Julien Musolino [d]

[a] Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland
[b] Serra Húnter Fellow, Department of Psychology, Institute of Educational Research, University of Girona, Spain
[c] Instituto de Investigaciones Psicológicas, Centro de Investigación en Neurociencias, Universidad de Costa Rica, Costa Rica
[d] Psychology Department, Rutgers Center for Cognitive Science, Rutgers University, USA

ARTICLE INFO

Article history:
Received 2 February 2021
Received in revised form 23 May 2022
Accepted 25 May 2022
Available online 28 June 2022

Keywords:
Conditional
Biconditional
Causal conditional reasoning
Pragmatic enrichment
Invited inferences

ABSTRACT

Human beings often fall prey to fallacious reasoning. One influential view holds that reasoners endorsing fallacies do not commit logical errors but rather that they endorse such fallacies on pragmatic grounds, i.e. by tacitly considering a conditional premise (*If you wash my car, I will give you $10*) a biconditional one (*I will give you $10, if and only if you wash my car*). If such an operation – known as invited inference or pragmatic enrichment – does occur, the acceptance of an invalid argument becomes legitimate. This paper seeks to find out if pragmatic enrichment happens for conditionals and, if so, under what circumstances. To address this issue, we conducted two complementary experiments in which we manipulated the type of material used in reasoning (abstract vs realistic) and the nature of the major premise (conditional vs biconditional). Our results indicate that both of these factors do indeed affect reasoning processes and performance. Overall, our findings suggest that only conditionals undergo pragmatic enrichment albeit not systematically as enrichment takes place only with a subclass of realistic materials.

1. Introduction

A well-known finding from the literature on conditional reasoning is that people often produce and accept fallacious arguments (see Evans, 1993; Girotto et al., 1997, for summaries). Does it mean that the precepts of classical logic are accessible only to a small group of reasoners explicitly trained to master them, such as logicians, mathematicians or philosophers? The question is not new and it is still extensively studied in psychology, philosophy, and linguistics from various perspectives. One of them – which is actually intensively investigated – appeals to Bayesian networks which examine the dependencies described by conditionals in terms of probabilities (cf. Oaksford and Chater, 2007, 2013, 2016). Another perspective, largely understudied, emphasizes the importance of the linguistic means used to formulate arguments. This paper intends to fill this gap by focusing on the linguistic side, and more particularly, on the pragmatic derivation of the meaning of logical words used to form major premises in conditional reasoning.

A proper treatment of this issue involves a careful examination of the linguistic components that enter into the relevant inferential schemes, namely, expressions such as logical connectives (e.g., *and*, *or* and *if …then)* and quantifiers (e.g., *some* and *all)*.

* Corresponding author. Université de Genève, UNI MAIL, 40, boulevard du Pont-d'Arve Genève 4, CH-1211, Switzerland.
E-mail addresses: joanna.blochowiak@unige.ch (J. Blochowiak), thomas.castelain@udg.edu (T. Castelain), odir.rodriguez@ucr.ac.cr (O.A. Rodriguez-Villagra), julienm@ruccs.rutgers.edu (J. Musolino).

https://doi.org/10.1016/j.pragma.2022.05.013

These words have a well-defined meaning in classical logic (semantics) but their interpretation in natural language (pragmatics) often strays from their rigid logical sense. According to one of the prevailing views in pragmatics – the standard Gricean view – (since Grice, 1975), the semantic meaning of logical words is equivalent to their logical meaning whereas their pragmatic meaning is derived as a result of rational cooperative behaviour among the participants of the conversation. For Grice, the principle of cooperation is at the heart of conversation: speakers and hearers cooperate in order for a conversation to be successful. Grice's main point is that this assumption of cooperation serves as the basis for hearers' pragmatic inferences which allow them to arrive at the speaker's intended meaning, the "pragmatic meaning". To grasp the pragmatic meaning of the speaker, the hearer-comprehender – assuming speakers' cooperative behaviour – goes beyond the semantic meaning of words in order to recover pragmatic interpretations that are the most informative. In sum, pragmatic meaning is inferred through the interaction of the principle of cooperation and some more specific conversational rules such as the maxim of quantity: *Make your contribution to the conversation as informative as necessary!* For example, in the case of logical words, the semantic meaning of *and* is pragmatically enriched to *and then*, the meaning of *or* is narrowed down by a pragmatic enrichment to its exclusive version *or but not both*, the pragmatic meaning of the quantifier *some* (which is logically compatible with *all*) corresponds to *some but not all*. Finally, the conditional expression *if… then* (IF) is pragmatically interpreted as the more specific *if and only if* (IFF). The precise meaning and the process of pragmatic enrichment associated with conditionals such as *if …then* are our primary focus in this paper.

2. A hypothesis of perfection of conditionals

In this Section we present a dominant hypothesis about the meaning of conditionals – the hypothesis of the *perfection of conditionals* (Geis and Zwicky, 1971). We start with theoretical discussion in Section 2.1 and we focus on its experimental investigation in Section 2.2.

2.1. Theoretical background

An influential proposal concerning the interpretation of conditionals goes back to Geis and Zwicky (1971) who put forth the hypothesis of the *perfection of conditionals*, also known as *invited inferences*. Similar ideas have been proposed in linguistics (Van der Auwera, 1997; Cornulier, 1985; Horn, 2000) and also in the psychology of reasoning (Rumain et al., 1983; Braine and O'Brien, 1991). According to the original hypothesis, the meaning of conditionals in natural language is regularly "perfected" to biconditionals via a process of invited inference as in the example below where (1a) invites the inference in (1b), which leads the conditional statement to be understood as the biconditional in (1c).[1]

(1) (a) If you wash my car, I will give you $10.
(b) If you don't wash my car, I will not give you $10.
(c) I will give you $10 *if and only if* you wash my car.

The characterisation of the pragmatic mechanism behind conditional perfection varies. Some treat the inferential move from *if* to *if and only if* as an I-implicature (Atlas and Levinson, 1981; Horn, 1984, 2000; Levinson, 2000) and others as a scalar implicature (for instance, Horn, 1972; Matsumoto, 1995; Van der Auwera, 1997). Franke (2009) proposes two different mechanisms to derive conditional perfection readings. The first appeals to normality assumptions and the second to the topic of the conversation. The latter concerns cases where perfection of the conditional is tightly related to a current question under discussion (see also von Fintel, 2001). The former is more relevant for our discussion since it is applicable to causal conditionals. According to Franke's proposal, in a typical context of utterance, a conditional like the one in the example reported below (Franke, 2009: 241), will get "perfected" to a biconditional because of the assumption concerning a *normal course of events*.

(2) (a) If John leans out of that window any further, he'll fall.
(b) If John doesn't, he will not fall.
(c) John will *normally* not just fall out of the window.

More precisely, the normality assumption works as a shared presupposition rooted in world knowledge about the normal causal development of a given situation, and this assumption, together with the conditional makes the biconditional true.

Being able to determine whether the process of perfection of conditionals takes place is crucial because the meaning of conditionals is directly related to the logical schemas determining the conditions under which the inference from premises to conclusions is logically legitimate, i.e. which arguments are valid, is directly related to the meaning of conditionals. For instance, (3) is an example of a valid inference scheme, the *Modus Ponens* (MP), while (4) is a classical example of a logical fallacy, the *Affirmation of the Consequent* (AC).

[1] It should be noted that from a logical perspective, the enrichment of a conditional to a biconditional could also happen with *only if* expressions coupled with presence/absence of negation (i.e., *Only if you wash my car, I will give you $10* or *Only if you do not wash my car, I will not give you $10*). For more details, see van Canegem-Ardijns and van Belle (2008) and the discussion in Franke (2009, Chapter 5).

(3) Modus Ponens (MP)

If P, then Q	*Major premise*
P	*Minor premise*
Therefore Q	*Conclusion*

(4) Affirmation of the Consequent (AC)

If P, then Q	*Major premise*
Q	*Minor premise*
Therefore P	*Conclusion*

If the major premise is formulated with a biconditional (IFF) instead of a conditional (IF), as in (1c), the *Affirmation of the Consequent* becomes a simple *Modus Ponens*. In other words, biconditionalization has the power to convert an invalid argument into a valid one. Thus, the investigation of the conditions under which a conditional can be transformed into a biconditional, and whether this process is akin to pragmatic enrichment, is essential for our understanding of reasoning mechanisms that lead to fallacy acceptance during reasoning. A word of caution should be stated here. It is important to point out that there is a paucity of studies that explicitly address the question of how we understand the words IF and IFF. One rare and rather old study is Paris, 1973's developmental paper, in which he shows that children, adolescents and young adults achieve similar accuracy rates for both logical connectives across the four truth forms of their respective logical truth-tables. Hence, our study will further verify whether adult reasoners differentiate between the two logical words using both accuracy and reading time measures. Obtaining the same results for both IF and IFF would show that adult reasoners do not make a difference between the meaning of the two logical words, at least in the case of MP and AC type of arguments. If confirmed with a wider range of arguments, such a finding would have far-reaching consequences because it would demonstrate that the long-lasting hypothesis of *perfection of conditionals* has been constructed on the basis of an artefact, from a cognitive point of view.

2.2. Previous experimental testing of conditional perfection

Before we turn to conditionals, a first general question that arises is to know whether people perceive logical meaning at all. A growing body of evidence suggests that people are indeed sensitive to the purely logical meanings of logical words. For example, numerous studies have shown that children tend to interpret logical words as they are defined in classical logic (Noveck, 2001; Papafragou and Musolino, 2003; Guasti et al., 2005; Pouscoulous et al., 2007). A recent study on the acquisition of logical connectives, including conditionals, by Chinese children confirms that the logical senses are acquired first (Su, 2014). What's more, for adults, the ‘non-enriched’ semantic readings require less effort in treatment than ‘enriched’ pragmatic readings, something which has been measured for instance in sentence processing tasks (see Breheny et al., 2006 for *or*; Blochowiak and Castelain, 2018 for *and*, see Noveck and Reboul, 2008 for a summary).[2] It is important to note that context also plays a role in people's logical-semantic/pragmatic interpretations: the theoretical literature indicates that pragmatic enrichment occurs in non-downward entailing contexts but is suspended in downward entailing contexts (Horn, 1972, 1989; see Chierchia, 2013 for an overview).

The phenomenon of *invited inferences* has also been tested experimentally in order to determine whether the process of perfection of conditionals postulated by Geis and Zwicky (1971) does indeed take place and, if so, how and when it happens. Recent experimental studies (Bonnefond and Van der Henst, 2009, 2013; Bonnefond et al., 2012, 2013; Bonnefond et al., 2014) suggest that such pragmatic enrichment does not occur for conditionals, at least in a similar way to the kind of pragmatic enrichment that can be observed for other logical words such as *and*, *or,* or quantifiers (see Noveck and Reboul, 2008). For instance, Bonnefond et al. (2012) – using both reading times and evoked related potentials (EEG) measures – show that the process of biconditionalization does not operate on the conditional itself (which corresponds to the major premise of an argument) as revealed by the fact that there is no extra cognitive cost observed at that moment of processing. Instead, there is a slow-down effect on the minor premise when it refers to the consequent, that is, in the case of an AC argument, as in (4) (see also Barrouillet et al., 2000 for a similar result). This slow-down effect is associated with a N200 wave, suggesting that reasoners are surprised when they see the consequent as minor premise in the course of their reasoning process. Moreover, two groups of reasoners have been identified: those who reject AC arguments (Rejecters) and those who accept them (Endorsers). As for the question why people fall into these two groups, Bonnefond et al. (2012) adopted an answer which is inspired by Evans's dual-systems suppositional theory that considers endorsements of AC as a sign of shallow processing of the major premise and rejections of AC as a sign of deeper processing (e.g., Evans et al. 2007). The interesting observation here is that the slow-down effect and the N200 wave were equally present in these two groups. These results have led Bonnefond et al. (2012) to conclude that conditionals do not undergo enrichment but, rather, that they come with a built-in expectation concerning the antecedent, which would explain the effect of surprise when the minor premise is encountered (Noveck et al., 2011). However, it should be noted that a possible drawback of these studies is that they restricted their investigation to

[2] However, it should be also pointed out that such processing delays are not always associated with calculation of pragmatic meaning, as it has been demonstrated in several eye-tracking studies investigating scalar implicatures associated with *some* (Degen and Tanenhaus, 2016; Foppolo and Marelli, 2017; Grodner et al., 2010; Ronai and Xiang, 2021). It should be also noted that not all pragmatic types of inferences cause a delay in reading time (e.g., metaphors, irony, idioms).

reasoning performed on abstract materials, such as sentences containing only letters (e.g. *If there is an A, then there is a U. There is an A. Therefore, there is a U*) and to the major premises of conditionals (IF).

The present research directly addresses these concerns by extending the investigation of invited inferences in two directions. First, we consider two types of materials known to affect people's reasoning performance. The first focuses on abstract or ‘contentless’ materials (i.e. sentences containing only letters), as in Bonnefond et al.’s (2012) study, and the second involves more realistic, concrete or ‘contentful’ materials (i.e. sentences describing real situations, e.g., involving causality). Second, in addition to conditionals, we also tested biconditionals in order to be able to directly compare the two sides of the invited inference equation, namely, the IF (conditional) and IFF (biconditional) formulations of the major premise.

We begin by discussing the two major elements that need to be taken into account in a comprehensive investigation of the potential pragmatic enrichment of conditionals: (i) the difference between abstract and realistic conditionals and their relevance for reasoning capacities in general (Section 3.1) and (ii) finer distinctions within the class of realistic conditionals of causal type (Section 3.2), which we chose for our study as an example of meaningful materials. Section 4 presents the results of two complementary experiments: (i) *Contentless experiment*, dealing with abstract material with conditional (IF) and biconditional (IFF) major premises and (ii) *Contenful experiment*, involving realistic material with conditional (IF) and biconditional (IFF) major premises. For each experiment, we performed three types of measures: (i) reading times on minor premises in the two types of arguments: *Modus Ponens* (MP) and *Affirmation of Consequent* (AC), (ii) acceptance rate in endorsing the conclusion from MP and AC arguments, and (iii) response time to endorse or reject the conclusion from MP and AC arguments (the details of the analyses and the results for the conclusions are available in the supplementary materials). We report on these results in Sections 4.1, 4.2. and 4.3. Section 5 offers a general discussion of the main findings.

3. What matters for pragmatic enrichment of conditionals?

In this Section we discuss why the type of material used in reasoning tasks is important for an analysis of pragmatic enrichment for conditionals. We begin in Section 3.1 with a discussion of reasoning with abstract vs realistic conditionals as major premises. In Section 3.2, we discuss in more depth causal conditional reasoning for which two finer-grained categories important for pragmatic enrichment are identified. Finally, in Section 3.3 we summarize the main hypotheses and predictions that will be tested in the experimental part of the paper.

3.1. Abstract and realistic conditionals

Typically, studies on conditionals investigate reasoning with two types of materials. First, abstract or arbitrary premises (containing letters, numbers or other abstract material) are constructed and then serve as experimental items in the study, as in example (5) below.

(5) a. If there is an A, then there is a D.
b. If there is a circle, then there is a triangle.

With this type of materials, it is hard to identify a relation that could link the antecedent to its consequent.

Second, experimental items can also be created with meaningful premises, in which participants have to evaluate a real-world relation between the antecedent and the consequent. The first type is sometimes called “basic conditionals” and the second type “thematic conditionals” (Evans et al., 1996). The most common thematic conditionals are causal and deontic conditionals, as in (6a) and (6b) respectively, but the list is not exhaustive.

(6) a. If you turn on the radiator, then you will feel heat.
b. If you do your homework, then I will let you go to the movies.

On the one hand, pragmatic theories of language tacitly assume that the perfection of conditionals happens in the environment of natural language, that is, in everyday meaningful contexts, such as in (6) rather than (5). On the other hand, Bonnefond and Van der Henst (2009) observed that in the domain of the psychology of reasoning the two most popular Models of deduction – *Mental Logic* (Braine and O'Brien, 1998) and *Mental Model Theory* (Johnson-Laird and Byrne, 1991; Johnson-Laird, 2001) – treat reasoning as an uniform system, that is, a system that works in a similar way independently of the type of premises given as input. However, as Bonnefond and Van der Henst (2009) point out, fMRI studies have shown that distinct brain networks are involved depending on the content of the premises in a given reasoning task. In particular, reasoning with abstract materials generates neural activity in the parietal-frontal network while reasoning with realistic materials prompts neural activity in the temporal-frontal network (see Goel, 2007, 2009). These findings indicate that the type of materials plays an important role in the cognitive processes involved in reasoning tasks.

More precisely, it has been claimed (Goel et al., 2000) that the use of contentless materials activates the network involved in the manipulation of spatial information. This pattern is also similar to the one found for some type of mathematical reasoning, in particular approximation of numerical quantities. This suggests that reasoners build a spatial representation out of abstract arguments and solve them by means of spatial models (such as Venn Diagrams, Euler Circles, or others). On the other hand, when meaningful material is present, the language network involving both semantic and syntactic processing is actively engaged. In sum, the presence of semantic content activates the language system while the absence of semantic content engages the spatial system in otherwise identical reasoning task.

Given the presence of different signatures linked to reasoning with abstract and realistic materials, we expect to find distinct reasoning profiles depending on the material used. More precisely, our hypothesis is that abstract reasoning is not a good candidate for finding the pragmatic enrichment of conditionals. If such enrichment exists, it should be looked for while participants perform reasoning tasks with meaningful, realistic material, since the language system is engaged therein.

In the present study, we therefore decided to use both abstract and realistic materials. As examples of major premises with realistic content, we chose causal conditionals. We follow previous studies in the psychology of reasoning which used this type of materials to gauge participants' performance during reasoning with valid and non-valid arguments (Cummins et al., 1991; Cummins, 1995 and subsequent work; Markovits and Quinn, 2002; De Neys et al., 2005a, b). This research aimed at identifying the specific factors that play a role in participants' performance on solving *if …then* problems (see Markovits and Barrouillet, 2002 for an overview). One such crucial factor is the retrieval of counter-examples from background knowledge. In particular, two types of counter-examples have been shown to matter for conditional causal reasoning: alternative causes and disabling conditions. We discuss them in the next section.

3.2. Causal conditionals: alternative causes and disabling conditions as grounds for pragmatic enrichment

A whole line of research on causal conditional reasoning, stemming from seminal work by Cummins (1995), has demonstrated the importance of the type of causal link contained in the major premise in causal conditional reasoning, also called naïve causal deduction. This kind of deduction is a form of reasoning in which participants draw conclusions about the occurrence of a cause or an effect on the basis of a conditional (also called conditional rule) describing a causal link between two events or two types of events.[3] The examples below, borrowed from Cummins (1995, p. 648), illustrate this type of reasoning.

(7)	If the brake was depressed <cause>, then the car slowed down <effect>.	*Major premise*
	The brake was depressed <cause>.	*Minor premise*
	Therefore, the car slowed down <effect>.	*Conclusion*
(8)	If the brake was depressed <cause>, then the car slowed down <effect>.	*Major premise*
	The car slowed down <effect>.	*Minor premise*
	Therefore, the brake was depressed <cause>.	*Conclusion*

In (7), on the basis of the occurrence of a cause, the reasoner can draw a conclusion about the occurrence of the corresponding effect using the rule of *Modus Ponens* (MP) whereas in (8), based on the occurrence of an effect, the reasoner can infer by *Affirmation of the Consequent* (AC) the occurrence of the corresponding cause.

Note that the links between causes and effects can vary in strength. This, in turn, has been shown to have an impact on causal reasoning as far as the order of presentation of causes and effects is concerned (Moeschler et al., 2006; Blochowiak et al., 2010). Cummins (1995) further demonstrated that the evaluation of the strength of a given causal link hinges on the availability of two types of counter-examples.

The first type of counter-example involves *alternative causes*. An alternative cause is a cause that could have given rise to the effect instead of or in addition to the cause explicitly mentioned in the rule. For instance, in the example provided above, the question would be what else beyond depressing the brakes could have caused the car to slow down. One could think of going uphill or some engine problems as potential alternative causes slowing down the car.

The second type of counter-example involves *disabling conditions*. A disabling condition is a set of events that, if they had unfolded, would have prevented an effect from occurring in the presence of a potential cause. In our example, we could ask what conditions should be present in order for a car not to slow down despite the fact that the driver depressed the brake. In this case, one could think of circumstances where the road is icy or the brake lines are severed as possible conditions that would prevent a car from slowing down even when the brakes have been depressed.

As Cummins and colleagues (1991), Cummins (1995) argue, the number of counter-examples, i.e. alternative causes and/or disabling conditions, that a given causal conditional rule could potentially admit matters. There are cases for which one can come up with many possible counter-examples, as in the car example. There are other cases for which it is difficult to find possible counter-examples. One of the classical examples provided by Cummins is the following: *If John took a glass with his bare hands, then he left his fingerprints on it.* Indeed, it is difficult to think of situations in which John took a glass with his bare hands but did not leave his fingerprints on it, or situations in which he left his fingerprints on a glass but did not take it with his bare hands.

As Cummins (1995) pointed out, this observation has consequences for reasoning with causal conditionals since the availability of counter-examples directly influences the possibility of transforming a conditional causal statement into a biconditional one. In general, the presence of few counter-examples for a given conditional statement would encourage reasoners to transform a conditional into a biconditional, as in the fingerprint example: *If and only if John took a glass with his*

[3] Causal conditional reasoning has been also reanalyzed in terms of causal Bayes nets involving probabilistic inferences (Fernbach et al., 2010; Rottman and Hastie, 2014). However, the probabilistic type of analysis seems to underestimate the impact of alternative causes (cf. Cummins, 2014). Also, some studies showed that deductive reasoning in general cannot be fully explained by probabilistic inferences (Markovits et al., 2015). Hence, we stick here to the first version of the analysis of the causal naïve deduction which is sufficient to illustrate the two main types of counter-examples.

bare hand, then he left his fingerprints on this glass. By contrast, the presence of more potential counter-examples would be less conducive to this transformation from conditional to biconditional. For example, the statement *If and only if the brakes were depressed, then the car slowed down* does not sound very convincing because comprehenders could easily imagine the car slowing down without the brakes having been depressed (see also Blochowiak, 2019).

Now, if reasoners take into account the number of counter-examples, we can expect differences in the reasoning process when the major premise refers to a conditional with *few* vs *many* potential counter-examples (Verschueren et al., 2004). It has been indeed observed that reasoning with causal premises is in general more difficult compared to other type of reasoning (such as category-based reasoning) precisely because of the process of generation of alternatives involving counter-examples (Markovits, 2017). More precisely, in the case of reasoning with a conditional major premise with *few* counter-examples, we would expect the acceptance rate of conclusions to be higher than in cases in which *many* counter-examples are possible. Indeed, this prediction has been verified in adults (Cummins, 1995, 2014) as well as in adolescents (De Neys and Everaerts, 2008). Hence, the *few-many* counter-examples dimension was integrated into our experimental design in the *Contentful experiment*.

We think that the pragmatic mechanism behind causal conditional reasoning is very similar to the one proposed by Franke (2009) described earlier, i.e. the mechanism which appeals to normality assumptions about the causal development of events. Recall that according to the normality assumption, *normally* a given event would not occur in the absence of other unexpected intervening events for a given causal situation. Such 'intervening events' can be seen as counter-examples in Cummins terms.[4] The normality assumption, that is the absence of other intervening events, holds with *few* counter-examples, allowing for biconditionalisation. On the other hand, when *many* counter-examples are present, the normality assumption does not hold, preventing the construction of a biconditional meaning. In sum, while the normality assumption is taken to be true by comprehenders, the biconditional is taken to be true as well whereas when the normality assumption is not shared, the biconditional interpretation does not arise.

Summing up, in accordance with the literature on causal conditional reasoning, our hypothesis is that reasoners are more likely to enrich conditionals with *few* counter-examples than ones with *many* counter-examples and this enrichment is linked to the commonly shared normality assumption (or a lack thereof) which constitutes the natural background knowledge of comprehenders.

3.3. Summary of hypotheses and predictions

The theoretical work and earlier research presented in the previous sections give rise to a series of predictions for our empirical studies, in which we compare reasoning with abstract versus realistic materials and two types of major premise: conditional versus biconditional. First, concerning the Endorsers-Rejecters split, we expect to replicate results of previous studies and to find this opposition while participants reason with abstract materials, that is, in the *Contentless experiment*. Concerning realistic reasoning, we can speculate that, when exposed to concrete materials, reasoners would behave more uniformly because they are driven by the content of the argument − beyond its pure form − as it is the case with abstract reasoning. For this reason, we would not expect to observe the Endorsers-Rejecters split in the *Contentful experiment.*

Concerning the proper enrichment, the most general prediction stems from the original Geis and Zwicky's (1971) thesis according to which conditionals (IF) are regularly enriched to biconditionals (IFF). As we saw, several studies have already demonstrated that the claim about a regular perfection of conditionals appears to be too strong. Thus, given the results of previous studies on conditionals as well as other logical words, our assumption is that the pragmatic enrichment of conditionals, if it occurs at all, needs not arise automatically, that is, IF is not systematically enriched to IFF. Our main hypothesis is that pragmatic enrichment of conditionals occurs only with realistic materials because reasoners enrich IF only when they are prompted to do so and possible triggers are found in meaningful linguistic environments. Based on this hypothesis, we predict to find pragmatic enrichment in the *Contentful* rather than the *Contentless experiment* and, by consequence, we do not expect to observe a difference between IF and IFF in the *Contentless experiment*.

Our third hypothesis concerns realistic materials and is based on two observations. First, it is well known that conditional reasoning does not induce similar cognitive treatment in every type of argument: MP arguments are usually not problematic compared to AC arguments. So, in the case of MP arguments, there is no incentive to enrich and it is only in the case of AC arguments that participants have a reason to consider enrichment. Second, according to the literature on naïve causal deduction (cf. Section 3.2), the causal conditionals that are more likely to undergo biconditionalisation are the ones which admit *few*, as opposed to *many*, counter-examples. Given these points, we can formulate our third hypothesis according to which pragmatic enrichment of causal IF can be observed with AC arguments when *few* counter-examples are present. Hence, we predict longer reading times for the minor premises of conditionals with *few* counter-examples in AC compared to MP. By contrast, this difference should be less marked in the minor premises of conditionals with *many* counter-examples, since the pragmatic grounds for enrichment are weaker in that case. Such differences should not be observed in reasoning with IFF because no biconditionalisation is needed in this type of reasoning.

[4] It should be noted that some studies differentiate between what they call *semantic* (retrieved from long-term memory) and *pragmatic* (more contextual) counter-examples (Verschueren et al., 2006). We do not make this distinction in our study because we are not able to track what type of counter-examples participants appeal to during their reasoning task. Thus, we consider equally any counter-example participants may have come up with as 'an intervening event' for the causal scenarios at hand.

The three main hypotheses concerning the pragmatic enrichment of conditionals and their respective predictions can be summarized as follows:

Hypothesis 1. Endorsers-Rejecters split is present only with abstract materials regardless of IF and IFF.

Prediction 1: Difference in Acceptance rate of the conclusion of AC between Endorsers and Rejecters is expected to be observed in the *Contentless experiment* for IF and IFF.

Prediction 2: Difference in Acceptance rate of the conclusion of AC between Endorsers and Rejecters is not expected to be observed in the *Contentful experiment*, whether it is for IF, or for IFF.

Hypothesis 2. Pragmatic enrichment of IF happens only with realistic materials.

Prediction 3: Pragmatic enrichment expected only in the *Contentful experiment*: a difference of Reading time and Acceptance rate between IF and IFF will be present.

Prediction 4: No differences in terms of Reading time between IF and IFF in the *Contentless experiment*.

Hypothesis 3. Pragmatic enrichment of causal IF occurs with *few* counter-examples.

Prediction 5: In the case of IF, longer reading times on the minor premise in ACfew compared to ACmany condition are expected.

Prediction 6: In the case of IFF, longer reading times on the minor premise in ACfew compared to ACmany are not expected.

Regarding the acceptance rate of conclusions, it is worth noting that while the process of pragmatic enrichment itself is expected to be associated with longer reading times on minor premises, the consequences of pragmatic enrichment can be further observable on the acceptance rates of conclusions. In general, there should be significantly higher acceptance rates for conclusions when the major premise involves a causal conditional with *few* compared to *many* counter-examples, regardless of the conditional and argument type. We also expect to confirm the common finding that participants accept more conclusions from MP than AC type of arguments. Finally, if reasoners differentiate between IF and IFF, we expect to find higher acceptance rates for IFF compared to IF, assuming that IFF does not need any form of pragmatic enrichment.

Concerning the reading time for the conclusion, we do not assume that the effects we predict for the minor premise appear on the conclusion, since we expect pragmatic enrichment to be typically associated to the minor premise. It should be noted however that we do not exclude that the differences between AC and MP will be present on the conclusion in relation to the inherent difference observed between the two types of argument (AC is usually more difficult to participants than MP). To be complete, we should note that we did not take into account the reading times of the major premises because they are exactly the same for the abstract materials and they are not comparable (due to the differences in their length) for the concrete materials.

4. Experimental investigation

In order to test the predictions outlined above, we carried out two self-paced reading experiments. The first one tested reasoning with abstract, contentless material – *Contentless experiment*. Two versions of the same experiment were created: one involving conditionals as major premises (IF) and the other using biconditionals (IFF). The other experiment tested reasoning with realistic, contentful materials – *Contentful experiment*. The realistic materials involved the kind of causal reasoning described in Cummins et al. (1991) and Cummins (1995). As with the *Contentless experiment*, one version of the *Contentful experiment* tested reasoning with conditionals (IF) and the other with biconditionals (IFF). Moreover, the *Contentful experiment* examined the effect of causal rules that admit few counter-examples (*few*) and many counter-examples (*many*). Section 4.1 and 4.2 present the methods, procedures, results, and discussions of the *Contentless* and *Contentful* experiments, respectively. Section 4.3 focuses on a comparison between the results of both experiments. Finally, Section 5 offers a general discussion and concluding remarks.

4.1. Contentless experiment

4.1.1. Participants

One hundred (31 females, $M_{age} = 18.83$, $SD_{age} = 1.41$, [18–26 y.o.]) native English speakers participated in this experiment. Participants were recruited from a pool of students at Rutgers University majoring in Psychology. They all received credit for their participation. Participants were randomly assigned to one of the two versions of the experiment: conditional or biconditional. Eighteen additional participants could not be included in the analyses based on their low level of accuracy (in the MP-Compatible, MP-Incompatible and AC-Incompatible conditions, see details in Section 4.1.2), reflecting a lack of comprehension, attention or motivation during the completion of the task.

4.1.2. Materials

Participants were presented with 48 conditional or biconditional arguments involving letters of the alphabet. For instance, for the conditional argument, the problems had the form *If there is a C, then there is a U*, and for the biconditional arguments, *If and only if there is a G, then there is a P*. We tried as much as possible to avoid joining letters that could have some unintended meaning, such as C and D (for CD-rom) or F and B (for Facebook).

The stimuli were created so that the problems were divided between compatible and incompatible *Modus Ponens* (MP-Compatible and MP-Incompatible) and compatible and incompatible *Affirmation of the Consequent* (AC-Compatible and AC-

Incompatible). A stimulus is incompatible if the conclusion does not correspond neither to the antecedent nor to the consequent of the conditional presented as the major premise (see Table 1). The same stimuli and partition were applied to biconditional problems leading to the following formulation of the problems: *If and only if there is a C, then there is a U; There is a C//There is a U.*

Table 1
Example of the problems used in the *Contentless experiment*, the case of IF. The same problems were used for the biconditional experiment.

Condition	Major	Minor	Conclusion	Number
MP-Compatible	If there is a C, then there is a U	There is a C	There is a U	16
MP-Incompatible	If there is a M, then there is a P	There is an M	There is a V	8
AC-Compatible	If there is a J, then there is an A	There is an A	There is a J	16
AC-Incompatible	If there is a B, then there is a K	There is a K	There is a O	8

4.1.3. Procedure

The two experiments (*Contentless* and *Contentful*) were designed using the E-Prime 2.0 software (Schneider et al., 2002) and followed the same procedure, which was adapted from Bonnefond et al. (2012) to allow comparisons with the present research.

Participants were first presented with the instructions and told that they would be asked to consider a series of short puzzles, each composed of three parts: a Rule, a Fact, and a Conclusion. Then, they were introduced to the self-paced procedure and told that each element would appear one at a time and that in order to move from one screen to the next, they needed to press the space bar. When presented with the conclusion, they were instructed to decide whether it was "true" or "false" by pressing the corresponding key. They were also told to answer as quickly and as accurately as possible and to maintain their fingers on the keyboard throughout the experiment. Participants were then introduced to a training phase composed of six problems sampling the different conditions they would be exposed to during the experimental phase. No feedback was provided during the training phase.

Each element of the problems was presented at the center of the computer screen (font type: Times New Roman, bold; font size: 14; color: white; background: black). The major and minor premises (i.e. the Rule and the Fact, respectively) were preceded by a fixation cross that lasted 600 ms (200 ms for the cross and 400 ms for a pause). A numeral sign ("#") that lasted 1900 ms (1500 ms for the sign and 400 ms for a pause) preceded the conclusion, so it could be easily distinguished from the premises. Finally, a delay of 400 ms separated each problem.

4.1.4. Data treatment and statistical methods

Data processing, analyses and plotting were conducted with R in Rstudio (R Development Core Team, 2020; RStudio Team, 2020) using the following packages: tidyverse (Wickham, 2017), lme4 (Bates et al., 2015), sjPlot (Lüdecke, 2020a), sjlabelled (Lüdecke, 2020b), sjmisc (Lüdecke, 2018), coefplot (Lander, 2018), Lattice (Sarkar, 2008).

We fitted generalized linear mixed-effects models (GLMMs) to examine the effects of Argument (MP coded as −1 and AC coded as 1) as within-subjects factor, Conditional (IF coded as −1 and IFF coded as 1) and Style (Endorsers coded as −1 and Rejecters coded as 1) as between-subjects factors, on the Acceptance rate of participants evaluating the conclusion (1 accept the conclusion and 0 reject the conclusion). We also used linear mixed-effects models (LMMs) to assess the effects of the same variables (using the same contrasts coding scheme) on the reading time of the Minor (RT Minor) and the response time of the Conclusion (RT Conclusion). These contrasts were coded as orthogonal custom contrasts (i.e., planned comparisons) in which the fixed effects estimated the differences between conditions and the intercept estimated the grand mean of dependent variables. All GLMMs and LMMs included the intercept of participants and items as crossed random effects, and the models were fitted via the maximum likelihood estimation procedure.

Our models were implemented following a step-up strategy. This strategy has been favored because of the nature of our predictions. Indeed, while some of them were theoretically grounded or based on previous research, others were more exploratory. Thus, our model selection strategy allows us to test our predictions but also new hypotheses as well; and thus, to select the model that best represents the data. We started with the implementation of a null model, which included the intercept as a fixed effect, and the intercept of items and persons as crossed-random effects. Then, we successively included the fixed effects. Finally, we incorporated the interactions and third order interactions. Only the models that, according to the algorithms used in the lme4 package, successfully converged were included.

For the model selection, we used multiple criteria: the log-Likelihood statistic, along with Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). The log-likelihood values cannot be used alone as an index of fit because they are a function of sample size, but they can be used to compare the fit of different models; the higher the value, the better the model. This statistic does not take under consideration the number of parameters in a model. The AIC and BIC permit a comparison between models of different complexity and are derived from the log-Likelihood statistic. The model with the lowest AIC and lowest BIC is preferred. To disambiguate the model selection process, we included the AIC and BIC differences (i.e., ΔAIC, ΔBIC).[5] According to Burnham and Anderson (2002, p. 70) values of ΔAIC between 0 and 2 indicate little support to discriminate between models,

[5] The procedure to compute the ΔAIC differences implicates to select the model with the lowest AIC; that is, preliminary, a winning model. Then, the differences between the preliminary winning model and the remaining models are computed. After obtaining these differences the winning model adopts a value of 0 and the other models adopt the differences. The ΔBIC differences are computed following the same procedure.

Table 2
Descriptive statistics for Acceptance rate (percentage) and Reading/Response Time (milliseconds) for the Contentless and the Contentful experiments.

	Contentless				Contentful							
	MP		AC		MPfew		MPmany		ACfew		ACmany	
	IF	IFF	IF	IFF	IF	IFF	IF	IFF	IF	IFF	IF	IFF
Acceptance Rate (percentage)	97.92 (36.53)	95.79 (40.37)	43.52 (44.44)	46.06 (49.73)	98.13 (19.80)	99.26 (12.70)	97.84 (21.19)	98.38 (15.74)	90.37 (26.59)	93.68 (24.08)	84.48 (33.02)	92.50 (26.12)
Minor (Reading time)	822.29 (646.02)	813.25 (597.87)	965.40 (717.29)	949.40 (755.86)	939.40 (460.98)	1006.70 (515.24)	994.06 (487.51)	1005.16 (486.68)	1069.88 (558.82)	1076.57 (558.86)	1021.75 (539.19)	968.37 (503.02)
Conclusion (Response Time)	751.98 (673.10)	772.28 (778.44)	855.79 (975.74)	867.88 (1034.99)	976.66 (504.01)	988.16 (504.12)	964.45 (500.85)	1030.87 (621.20)	1169.22 (684.29)	1168.89 (671.97)	1172.18 (738.95)	1139.92 (647.18)

Note. AC = Affirmation of the Consequent, *ACfew* = Affirmation of the Consequent with few counter-examples, *ACmany* = Affirmation of the Consequent with many counter-examples, *IF* = Conditional, *IFF* = Biconditional, *MP* = Modus Ponens, *MPfew* = Modus Ponens with few counter-examples, *MPmany* = Modus Ponens with many counter-examples. In brackets the respective standard deviation of acceptance rate (percentage) and reading/response time (milliseconds).

values of ΔAIC from 4 to 7 indicate less support for the model with higher AIC, and values of ΔAIC > 10 suggest no support for the model with the higher AIC. Wasserman (2000) suggests that ΔBIC may be interpreted using the same guidelines.

4.1.5. Results

4.1.5.1. Descriptive statistics. Table 2 summarizes the descriptive statistics of the *Contentless* and *Contentful* experiments: Acceptance rate (percentage) of the Conclusion, Reading time for the Minor (milliseconds), and Response time for the Conclusion (milliseconds) as a function of the Argument.

4.1.5.2. Reasoning style. In the *Contentless experiment*, acceptance rates for AC were not distributed evenly across participants. Indeed, as reported in Bonnefond et al. (2012), participants can be divided between two different 'styles' of reasoning: *Endorsers* and *Rejecters*. The former typically accept the conclusion (at rates that differed from chance predictions; binomial test) while the latter systematically reject it. We found that 40 participants could be categorized as Endorsers (40%), 52 as Rejecters (52%) and 8 participants were qualified as Mixed because they showed a pattern of responses that was not consistent with either style. The Mixed participants were excluded from the following analysis.

4.1.5.3. Acceptance Rate. Table 3 gives the ΔAIC and ΔBIC values for each model for Acceptance rate (see Top panel of Table 3). The ΔAIC values exclusively supports model 5. The ΔBIC values indicate that models 2 and 5 appear to be the closest accounts for the data and it is difficult to discriminate between the goodness-of-fit of these models. Inspection of parameters estimates for model 5 indicates that, with the exception of the estimate of Conditional, all model parameters are different from zero (see the 95% confidence intervals at the top panel of Table 4). Inspection of Fig. 1 clearly suggests a higher proportion of Acceptance rate for the MP condition. However, the interaction term between Argument and Conditional is not easy to visualize in this figure.

Table 3
Models testing sequence for Acceptance rate and for Acceptance rate as a function of the Style for the Contentless experiment.

#	Fixed effects	Df	AIC	ΔAIC	BIC	ΔBIC	Log-Lik
Acceptance rate							
1	Int.	3	1852.8	207.726	1871.0	189.512	−923.41
2	Int., Argument	4	1658.1	13.043	1682.4	0.902	−825.06
3	Int., Conditional	4	1854.3	209.215	1878.6	197.073	−923.15
4	Int., Argument, Conditional	5	1660.0	14.925	1690.4	8.854	−825.00
5	Int., Argument * Conditional	6	1645.1	0.000	1681.5	0.000	−816.54
Acceptance rate as a function of the Style							
1	Int.	3	1614.4	578.15	1632.3	554.20	−804.19
2	Int., Argument	4	1415.3	379.06	1439.2	361.10	−703.64
3	Int., Conditional	4	1615.9	579.71	1639.9	561.75	−803.97
4	Int., Style	4	1404.4	368.14	1428.3	350.18	−698.18
5	Int., Argument, Style, Conditional	6	1209.7	173.45	1245.6	167.47	−598.84
6	Int., Argument * Style, Conditional	7	1036.2	0.00	1078.1	0.00	−511.11
7	Int., Argument * Conditional, Style	7	1207.8	171.59	1249.7	171.59	−596.91
8	Int., Conditional * Style, Argument	7	1208.0	171.78	1249.9	171.78	−597.00
9	Int., Argument * Style * Conditional	10	1038.6	2.41	1098.5	20.37	−509.32

Note. Argument = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Style* = Rejecters (R) and Endorsers (E); *Df.* = degree of freedom; *AIC* = Akaike's Information Criterion; *ΔAIC* = differences between the winning model and remaining models, *BIC* = Bayesian information Criterion, *ΔBIC* = differences between the winning model and remaining models, *Log-Lik* = Log-Likelihood.

Table 4
Parameters estimates of Models Acceptance rate and Acceptance rate as a function of the Style for the Contentless experiment.

Effect	Parameter	Estimate	95% CI	Standard error	*Z* (*p*-value)
Acceptance Rate[a]					
Intercept	*Intercept.*	2.5986	[2.02, 3.21]	0.2994	8.836 (<0.001)
Slope of *Argument*	*AC-MP*	−2.9046	[−3.16, −2.67]	0.1222	−23.761 (<0.001)
Slope of *Conditional*	*IFF-IF*	−0.3633	[−0.96, 0.22]	0.2964	−1.226 (0.22)
Interaction of *Argument * Conditional*	*AC-MP*IFF-IF*	0.4679	[0.24, 0.71]	0.1183	3.956 (<0.001)
Acceptance rate as a function of the Style[b]					
Intercept	*Intercept.*	1.7563	[1.51, 2.04]	0.1355	12.965 (<0.001)
Slope of *Argument*	*AC-MP*	−1.9436	[−2.18, −1.73]	0.1143	−17.011 (<0.001)
Slope of *Style*	*R-E*	−1.4992	[−1.77, −1.25]	0.1324	−11.327 (<0.001)
Slope of *Conditional*	*IFF-IF*	−0.2035	[−0.46, 0.04]	0.1254	−1.623 (0.105)
Interaction of *Argument * Style*	*AC-MP*R-E*	−1.5694	[−1.79, −1.36]	0.1098	−14.293 (<0.001)

Note. Argument = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Style* = Rejecters (R) and Endorsers (E), *AC-MP* = contrasts (i.e., differences) between Affirmation of the Consequent and Modus Ponens, *IFF-IF* = contrasts (i.e., differences) between Biconditional and Conditional, *R-E* = contrasts (i.e., differences) between Rejecters and Endorsers.

[a] Random effects for the Acceptance rate model. Random intercept for subjects = 2.74 (Standard deviation), 95% Confidence interval [CI] = 2.31, 3.28; Random intercept for items = 0.000, 95% Confidence interval [CI] = 0.00, 0.30.

[b] Random effects for the Acceptance rate as a function of the Style model. Random intercept for subjects = 0.7331 (Standard deviation), 95% Confidence interval [CI] = 0.43, 1.05; Random intercept for items = 0.1841 (Standard deviation), 95% Confidence interval [CI] = 0.00, 0.54.

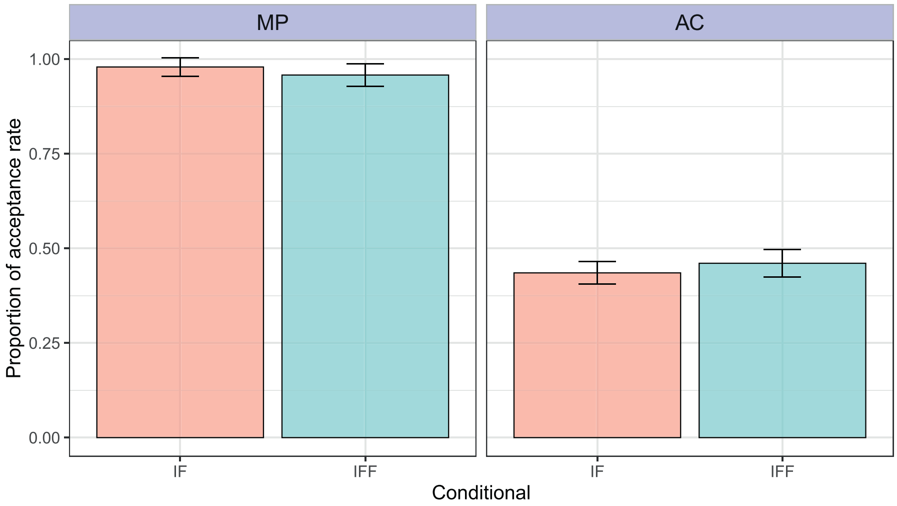

Fig. 1. Acceptance rate as a function of Conditional for the Contentless experiment. *AC* = Affirmation of the Consequent, *IF* = Conditional, *IFF* = Biconditional, *MP* = Modus Ponens. The within-subjects 95% confidence intervals were computed following the method proposed by Morey (2008) and were implemented using the R functions developed by Chang (2018).

Taking into account the considerable uncertainty of the ΔBIC values to discriminate between the goodness-of-fit of models 2 and 5, and the non-trivial effect of the interaction term in model 5, we regarded model 5 as the winning model. The top panel of Table 4 shows the parameters estimates of the winning model. The negative estimate of Argument indicates that the log-odds of Acceptance rate for the AC condition were smaller than for the MP condition. Furthermore, the interaction between Argument and Conditional suggests that, in the AC condition, the log-odds of Acceptance rate of participants in the IFF condition were higher than in the IF condition.

The bottom panel of Table 3 shows the goodness-of-fit indices for each model for Acceptance rate as a function of Argument, Conditional and Style. The ΔAIC values provide evidence to support model 6 and model 9, while the ΔBIC values exclusively support model 6.

Parameters estimates for model 9 indicate that the 95% confidence intervals of the interaction terms Argument * Conditional, Style * Conditional, and Argument * Style * Conditional include zero, suggesting a non-reliable effect on Acceptance rate (data not shown). Parameters estimates for model 6 indicate that the fixed effects Argument, Style, and the interaction term between Argument and Style are reliable predictors of Acceptance rate (see the 95% confidence intervals of this model in the bottom panel of Table 4). Due to the non-reliable parameters estimates for model 9, its complexity in terms of number of parameters, and the exclusive support for model 6 according to the ΔBIC values, we considered model 6 as the best model.

The negative estimate of the interaction term Argument*Style reveals that the log-odds of Acceptance rate of Rejecters were smaller when the Argument was the affirmation of the consequent (i.e., AC condition). Fig. 2 helps to illustrate these

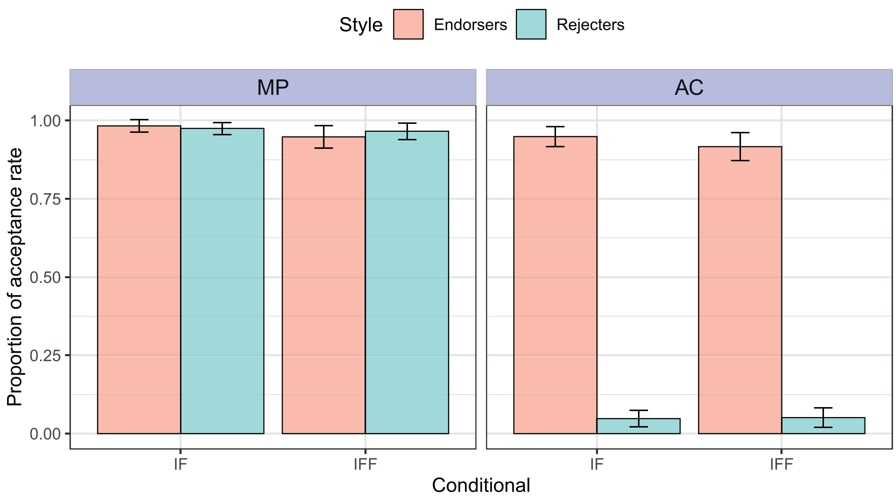

Fig. 2. Acceptance rate as a function of Conditional and Style for the Contentless experiment. *AC* = Affirmation of the Consequent, *IF* = Conditional, *IFF* = Biconditional, *MP* = Modus Ponens. The within-subjects 95% confidence intervals were computed following the method proposed by Morey (2008) and were implemented using the R functions developed by Chang (2018).

findings. As described above, the figure indicates that, regardless of Conditional, rejecters showed a smaller proportion of Acceptance rate in the AC condition.

4.1.5.4. Reading time of the minor. The top panel in Table 5 displays the models testing sequence for Reading time Minor as a function of Argument and Conditional. Although not conclusive, the ΔAIC values are more in favour of model 2 than model 4. The ΔBIC values exclusively support model 2. Examination of parameters estimates for model 5 indicates that the 95% confidence intervals of Conditional and the interaction term between Argument and Condition include zero (data not shown). We therefore took model 2 to be the bestmodel. The positive slope of Argument suggests that participants in the AC condition needed more time to read the Minor than in the MP condition (see Top panel of Table 6).

Table 5
Models testing sequence for Reading time Minor and Reading time Minor as a function of the Style for the Contentless experiment.

#	Fixed effects	Df	AIC	ΔAIC	BIC	ΔBIC	Log-Lik
Reading time Minor							
1	Int.	4	45800	40.497	45824	34.495	−22896
2	Int., Argument	5	45760	0.000	45790	0.000	−22875
3	Int., Conditional	5	45802	42.472	45832	42.472	−22896
4	Int., Argument, Conditional	6	45762	1.970	45798	7.972	−22875
5	Int., Argument * Conditional	7	45764	3.939	45806	15.944	−22875
Reading time Minor as a function of the Style							
1	Int.	4	42147	46.873	42171	30.571	−21070
2	Int., Argument	5	42111	10.382	42140	0.000	−21050
3	Int., Conditional	5	42149	48.794	42179	38.412	−21070
4	Int., Style	5	42144	43.803	42174	33.421	−21067
5	Int., Argument, Style, Conditional	7	42110	9.256	42151	10.715	−21048
6	Int., Argument * Style, Conditional	8	42109	8.459	42156	15.838	−21046
7	Int., Argument * Conditional, Style	8	42112	11.210	42159	18.589	−21048
8	Int., Conditional * Style, Argument	8	42102	1.800	42150	9.178	−21043
9	Int., Argument * Style * Conditional	11	42100	0.000	42166	25.139	−21039

Note. Note. Argument = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Style* = Rejecters (R) and Endorsers (E), *Df.* = degree of freedom, *AIC* = Akaike's Information Criterion; *ΔAIC* = differences between the winning model and remaining models, *BIC* = Bayesian information Criterion, *ΔBIC* = differences between the winning model and remaining models, *Log-Lik* = Log-Likelihood.

Table 6
Parameters estimates of Models Reading time Minor and Reading time Minor as a function of the Style for the Contentless experiment.

Effect	Parameter	Estimate	95% CI	Standard error	*Z* (*p*-value)
Reading time Minor[a]					
Intercept	*Intercept.*	893.278	[883.35, 948.14]	27.736	32.206 (<0.001)
Slope of *Argument*	*AC-MP*	72.783	[54.20, 91.34]	9.331	7.801 (<0.001)
Reading time Minor as a function of the Style[b]					
Intercept	*Intercept.*	874.775	[822.87, 926.59]	26.179	33.415 (<0.001)
Slope of *Argument*	*AC-MP*	69.741	[49.66, 89.80]	10.097	6.907 (<0.001)
Slope of *Style*	*R-E*	51.541	[0.17, 102.98]	25.950	1.986 (0.05)
Slope of *Conditional*	*IFF-IF*	3.916	[-47.97, 55.74]	26.179	0.150 (0.88)
Interaction of *Argument * Style*	*AC-MP*R-E*	13.202	[-5.40, 31.80]	9.486	1.393 (0.16)
Interaction of Argument * Conditional	*AC-MP*IFF-IF*	1.176	[-18.89, 21.25]	10.097	0.116 (0.91)
Interaction of Style * Conditional	*R-E*IFF-IF*	−81.999	[-133.37, −30.57]	25.950	−3.160 (<0.01)
Interaction of Argument * Style * Conditional	*AC-MP*R-E*IFF-IF*	−21.215	[−39.81, −2.61]	9.486	−2.236 (<0.05)

Note. Argument = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Style* = Rejecters (R) and Endorsers (E), *AC-MP* = contrasts (i.e., differences) between Affirmation of the Consequent and Modus Ponens, *IFF-IF* = contrasts (i.e., differences) between Biconditional and Conditional, *R-E* = contrasts (i.e., differences) between Rejecters and Endorsers.

[a] Random effects for the Reading time Minor model. Random intercept for subjects = 260.88 (Standard deviation), 95% Confidence interval [CI] = 224.04, 306.50; Random intercept for items = 19.33 (Standard deviation), 95% Confidence interval [CI] = 0.00, 53.58.

[b] Random effects for the Reading time Minor model as a function of the Style. Random intercept for subjects = 227.46 (Standard deviation), 95% Confidence interval [CI] = 192.91, 270.40; Random intercept for items = 27.68 (Standard deviation), 95% Confidence interval [CI] = 0.00, 59.81.

With respect to the Reading time Minor as a function of Style, the ΔAIC values favour models 8 and 9, and the ΔBIC values uniquely support model 2. On the one hand, data inspection of parameters for model 9 indicates that the 95% confidence intervals of the interaction between Style and Conditional and the third order interaction (i.e., Argument * Style * Conditional) do not include zero (see the bottom panel of Table 6), suggesting a reliable effect on the Reading time of the Minor. The negative estimate of the Style * Condition interaction indicates that, regardless of the Argument, Rejecters read in a slower pace in the IF condition. The negative slope of the three-way interaction indicates that Rejecters in the IF condition were the slower group reading the Minor in the AC condition.

On the other hand, Fig. 3 suggests that, on average, Rejecters need more time to read the Minor in the IF condition but the 95% confidence intervals of the former group and the Endorsers group, in the IFF condition, overlap. Although Fig. 3 does not clearly show the three-way interaction, the parameters estimates for model 9 indicate that the effect of this interaction is different from zero. Thus, we regarded model 9 as the best model.

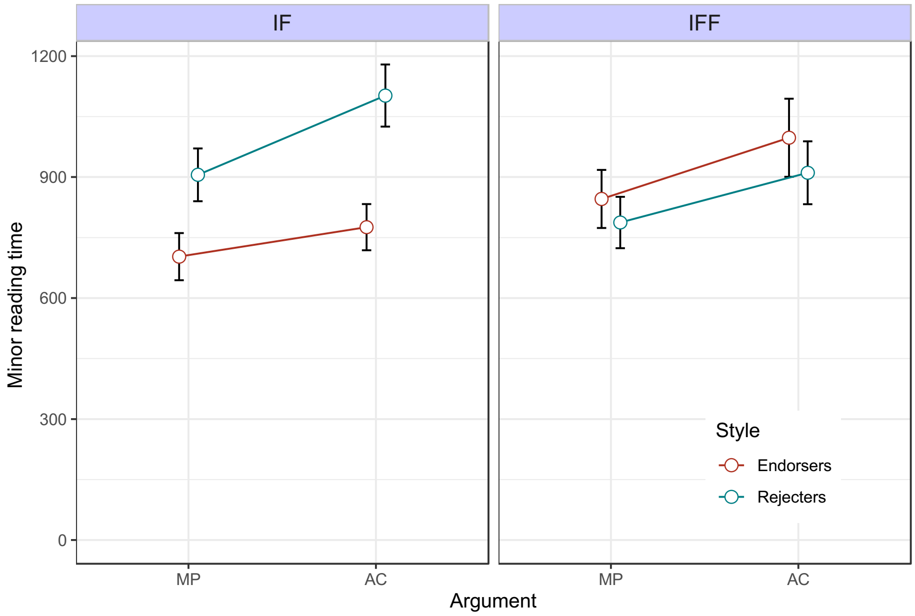

Fig. 3. Reading Time for Minor as a function of Conditional and Style for the Contentless experiment. *MP* = Modus Ponens, *AC* = Affirmation of the Consequent, *IF* = Conditional, *IFF* = Biconditional. The within-subjects 95% confidence intervals were computed following the method proposed by Morey (2008) and were implemented using the R functions developed by Chang (2018).

4.1.6. Discussion

First, it is worth pointing out that in our *Contentless experiment*, we were able to identify two groups of reasoners (Endorsers and Rejecters) as in previous studies (Bonnefond et al., 2012). The Endorsers and Rejecters do not differ in their acceptance of MP arguments but they do when it comes to AC arguments in the sense that there is a split between those who systematically accept the conclusion (Endorsers) and those who systematically reject it (Rejecters). This split in the reasoners' behavior is unsurprising with Conditional but more interestingly, we also found the same split with biconditionals.

The results from reading times reveal two interesting patterns. The first is that overall, the minor premise was read at significantly slower rates in AC compared to MP. This replicates the pattern reported in Bonnefond et al. (2012). However, this pattern was not observed when reasoning is performed with IFF. This difference shows that reasoners do distinguish between IF and IFF in abstract reasoning.

The second observation concerns the two styles that participants adopt in dealing with a reasoning task (Endorsers vs Rejecters). We found both styles independently of the form of the major premise (IF or IFF) but the behavior of Rejecters differs with respect to the treatment of the minor premise. In the case where the major premise involves IF, Rejecters take significantly more time than Endorsers to read the minor premise of AC type configurations compared to MP arguments.

4.2. Contentful experiment

4.2.1. Participants

One hundred seventy-two (83 females, M_{age} = 19.02, SD_{age} = 1.65, [18–32 y.o.]) native English speakers from a pool of students at Rutgers University majoring in Psychology participated in this study. They all received credit for their participation. Again, each of them was randomly assigned to one of the two types of conditionals (conditional or biconditional). Twenty-one additional participants could not be included in the analyses based on their low level of accuracy (in the MP-Compatible, MP-Incompatible and AC-Incompatible conditions), reflecting a lack of comprehension, attention or motivation during the completion of the task.

4.2.2. Materials

Participants were presented with 48 conditional or biconditional arguments involving everyday causal situations (i.e., realistic materials). In this case, we introduced two new classes of items: *few* for strongly related causes and consequences and *many* for weakly related ones. In order to generate our experimental items and determine whether a given causal rule admitted *few* or *many* counter-examples, we carried out an offline experiment along the lines of Cummins et al. (1991) and Cummins (1995), the *Generation Task* (see Appendix A. Supplementary materials).

4.2.3. Procedure

The *Contentful experiment* followed the exact same procedure as described in Section 4.1.3 for the *Contentless experiment*.

4.2.4. Results

4.2.4.1. Descriptive statistics. The descriptive statistics of the *Contentful experiment* are presented in Table 2 (see Section 4.1.4 Results of the *Contentless experiment*).

4.2.4.2. Reasoning style. In the *Contentful experiment*, we were not able to report the pattern of 'reasoning style' observed in the *Contentless experiment*. Most of the participants (157) showed a pattern of response similar to Endorsers (91%), 6 were described as Rejecters (3.5%) and 9 as Mixed (5.2%). The Mixed participants were excluded from the following analysis.

4.2.4.3. Acceptance rate. Table 7 shows the models testing sequence for Acceptance rate for the *Contentful experiment*. The ΔAIC values provide little support to discriminate among the goodness-of-fit of models 5, 6, 7, and 8, and provide less support for model 9. The ΔBIC values favour model 2 over model 5. Examination of parameters estimates for model 6 to 9 indicates that all interaction terms in these models include zero in their 95% confidence intervals (data not shown). Thus, the non-reliable effect of these interaction terms suggests that model 2 and 5 are two better models for the data.

Table 7
Models testing sequence for Acceptance rate for the Contentful experiment.

#	Fixed effects	Df	AIC	ΔAIC	BIC	ΔBIC	Log-Lik
1	Int.	3	1805.8	80.875	1825.7	66.991	−899.93
2	Int., Argument	4	1732.2	7.271	1758.7	0.000	−862.12
3	Int., Conditional	4	1805.3	80.354	1831.8	73.083	−898.66
4	Int., Counter-example	4	1806.2	81.182	1832.6	73.911	−899.08
5	Int., Argument, Counter-example, Conditional	6	1725.0	0.000	1764.7	5.955	−856.49
6	Int., Argument * Counter-example, Conditional	7	1726.9	1.959	1773.2	14.527	−856.47
7	Int., Argument * Conditional, Counter-example	7	1726.4	1.420	1772.7	13.989	−856.20
8	Int., Conditional * Counter-example, Argument	7	1726.7	1.670	1772.9	14.239	−856.32
9	Int., Argument * Counter-example * Conditional	10	1729.8	4.821	1795.9	37.229	−854.90

Note. *Argument* = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Counter-examples* = Many counter-examples (Many) and Few counter-examples (Few), *Df.* = degree of freedom; *AIC* = Akaike's Information Criterion, *ΔAIC* = differences between the winning model and remaining models, *BIC* = Bayesian information Criterion, *ΔBIC* = differences between the winning model and remaining models, *Log-Lik* = Log-Likelihood.

Parameters estimates for model 5 indicate that the log-odds of Acceptance rate were higher in the MP condition, the Few condition, and the IFF condition (see Table 8). Fig. 4 helps us visualize the effect of these variables on Acceptance rate: 1) a higher proportion of Acceptance for MP compared to AC; 2) a higher proportion of Acceptance when the number of counter-examples was small (i.e., Few condition) than when the number of counter-examples was large (i.e., Many condition); 3) a higher proportion of Acceptance for the IFF condition. Based on a consideration of the parameters for the different models and the inspection of Fig. 4, we determined that model 5 was the best model.

Table 8
Parameters estimates of Model Acceptance rate for the Contentful experiment.[a]

Effect	Parameter	Estimate	95% CI	Standard error	*Z* (*p*-value)
Intercept	*Intercept.*	6.03931	[5.46, 6.70]	0.31444	19.207 (<0.001)
Slope of *Argument*	*AC-MP*	−2.44766	[−2.91, −2.02]	0.22538	−10.860 (<0.001)
Slope of *Conditional*	*IFF-IF*	0.42297	[0.03, 0.83]	0.20025	2.112 (<0.05)
Slope of *Counter-examples*	*Many-Few*	−0.26072	[−0.45, −0.07]	0.09512	−2.741 (<0.01)

Note. *Argument* = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Counter-examples* = Many counter-examples (Many) and Few counter-examples (Few), *AC-MP* = contrasts (i.e., differences) between Affirmation of the Consequent and Modus Ponens, *IFF-IF* = contrasts (i.e., differences) between Biconditional and Conditional, *Many-Few* = contrasts (i.e., differences) between Many counter-examples and Few counter-examples.

[a] Random effects for the Acceptance Rate model. Random intercept for subjects = 1.9673 (Standard deviation), 95% Confidence interval [CI] = 1.62, 2.42; Random intercept for items = 0.4342 (Standard deviation), 95% Confidence interval [CI] = 0.23, 0.66.

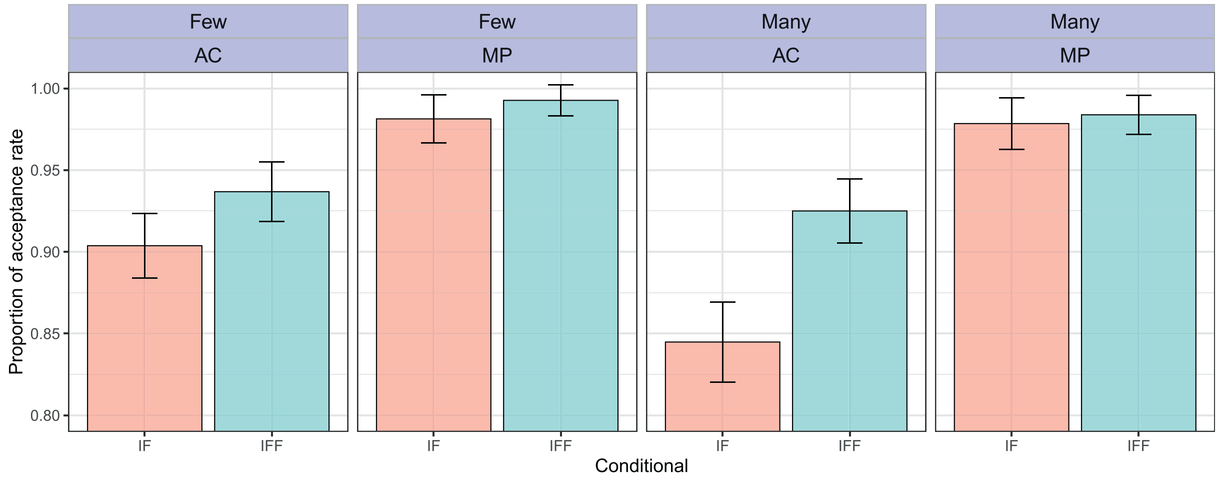

Fig. 4. Acceptance rate as a function of Conditional and Counter-examples for the Contentful experiment. *MP* = Modus Ponens, *AC* = Affirmation of the Consequent, *IF* = Conditional, *IFF* = Biconditional, *Few* = Few counter-examples, *Many* = Many counter-examples. The within-subjects 95% confidence intervals were computed following the method proposed by Morey (2008) and were implemented using the R functions developed by Chang (2018).

4.2.4.4. Reading time of the minor. Table 9 displays the model testing sequence for Reading time Minor for the *Contentful experiment*. The ΔAIC values provide support for model 6 and 9. The ΔBIC values provide support for model 1 and 2. The parameters estimates for model 9 show that the effect of the interaction between Argument, Counter-example, and Conditional is not different from zero (data not shown). However, based on Fig. 5, we can see that the MP and AC conditions differ only in the Few counter-examples case, and when problems with few counter-examples were presented to participants in the IF condition. Fig. 5 therefore hints in favour of the three-way interaction supported by the ΔAIC values. Taken together, these findings support our predictions (presented in Section 3.2). We implemented a simple effects analysis to break down the three-way interaction included in model 9. To this end, four contrasts were computed. The first contrast (i.e., ACFewIF vs MPFewIF) compared the Reading time Minor for participants in AC, Few and IF conditions (coded as 1) against participants in MP, Few and IF conditions (coded as −1). The second contrast (i.e., ACManyIF vs MPManyIF) compared the Reading time Minor between participants in the AC, Many, and IF conditions (coded as 1) and participants in MP, Many, and IF conditions (coded as −1). The third and fourth contrasts were similar to the previous two, but they uniquely included participants in the IFF condition (for details see the results displayed in Table 10). Table 10 presents the results of a model that includes the contrasts described above. As we expected, the first contrast (i.e., ACFewIF vs MPFewIF) is the only one of four contrasts that had an effect on Reading time Minor. The positive estimate of this contrast suggests that participants were slower at reading the AC arguments when they were exposed to problems that generate few counter-examples in the IF condition. Also, as anticipated, there were no differences in Reading time Minor in the remaining conditions.

Table 9
Models testing sequence for Reading time Minor for the Contentful experiment.

#	Fixed effects	Df	AIC	ΔAIC	BIC	ΔBIC	Log-Lik
1	Int.	4	77659	5.7563	77686	0.000	−38826
2	Int., Argument	5	77656	2.4965	77689	3.283	−38823
3	Int., Conditional	5	77661	7.6794	77694	8.466	−38826
4	Int., Counter-example	5	77660	6.3544	77793	7.141	−38825
5	Int., Argument, Counter-example, Conditional	7	77659	4.8981	77704	18.770	−38822
6	Int., Argument * Counter-example, Conditional	8	77654	0.000	77706	20.415	−38819
7	Int., Argument * Conditional, Counter-example	8	77658	4.2581	77710	24.673	−38821
8	Int., Conditional * Counter-example, Argument	8	77659	5.2444	77711	25.660	−38821
9	Int., Argument * Counter-example * Conditional	11	77655	1.0910	77727	41.135	−38816

Note: *Argument* = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Counter-examples* = Many counter-examples (Many) and Few counter-examples (Few), *Df.*: degree of freedom, *AIC*: Akaike's Information Criterion, *ΔAIC*: differences between the winning model and remaining models, *BIC*: Bayesian information Criterion, *ΔBIC*: differences between the winning model and remaining models, *Log-Lik*: Log-Likelihood.

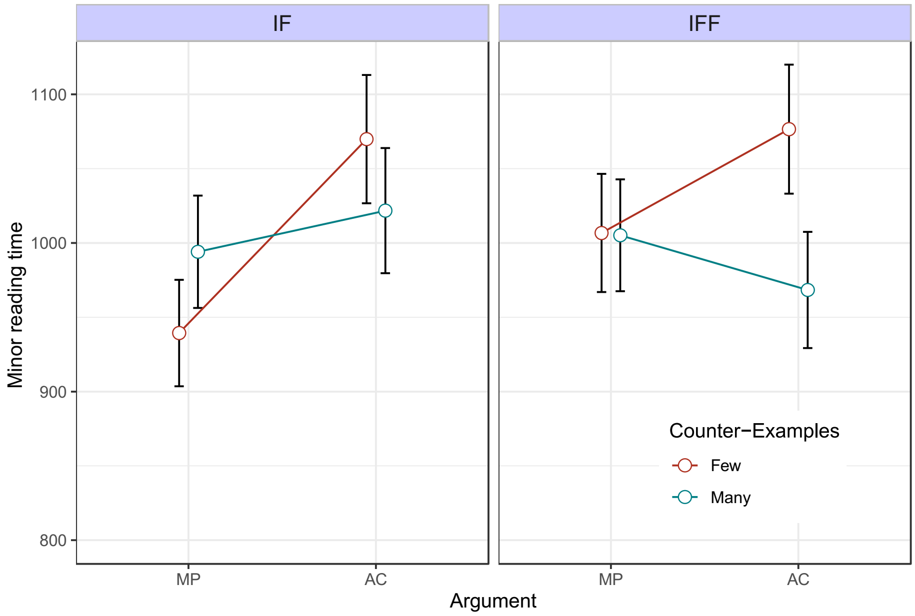

Fig. 5. Reading Time for Minor as a function of Conditional and Counter-examples for the Contentful experiment. *MP* = Modus Ponens, *AC* = Affirmation of the Consequent, *IF* = Conditional, *IFF* = Biconditional. The within-subjects 95% confidence intervals were computed following the method proposed by Morey (2008) and were implemented using the R functions developed by Chang (2018).

Table 10
Simple effects analysis for the three way interaction of model 9 (Reading time Minor for the Contentful experiment).[a]

Effect	Parameter	Estimate	95% CI	Standard error	*Z* (*p*-value)
Intercept	*Intercept.*	1004.92	[966.50, 1043.27]	19.48	51.576 (<0.001)
Slope of *Conditional few counter-examples*	*ACfewIF-MPfewIF*	69.73	[29.33, 110.14]	63.11	3.435 (<0.01)
Slope of *Conditional many counter-examples*	*ACmanyIF-MPmanyIF*	16.34	[−24.14, 56.81]	63.59	0.803 (0.42)
Slope of *Biconditional few counter-examples*	*ACfewIFF-MPfewIFF*	36.76	[−3.60, 77.16]	62.99	1.812 (0.07)
Slope of *Biconditional many counter-examples*	*ACmanyIFF-MPmanyIFF*	−20.34	[−60.76, 20.07]	63.19	−1.002 (0.32)

Note. Argument = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Counter-examples* = Many counter-examples (Many) and Few counter-examples (Few), *AC-MP* = contrasts (i.e., differences) between Affirmation of the Consequent and Modus Ponens, *IFF-IF* = contrasts (i.e., differences) between Biconditional and Conditional, *Many-Few* = contrasts (i.e., differences) between Many counter-examples and Few counter-examples.
[a] Random effects for the Reading time Minor model. Random intercept for subjects = 217.74 (Standard deviation), 95% Confidence interval [CI] = 192.98, 246.77; Random intercept for items = 63.84 (Standard deviation), 95% Confidence interval [CI] = 47.12, 83.86.

4.2.5. Discussion

The first important difference between reasoning with abstract and realistic materials revealed in our investigation is that the division between those who systematically reject (Rejecters) and those who systematically accept (Endorsers) a conclusion from AC arguments is only seen in abstract reasoning. This finding can be taken as a first indication that pragmatic enrichment is not to be sought in abstract contexts.

Concerning acceptance rates, first, we found, as expected, that reasoners accepted significantly more conclusions from MP than AC arguments. Second, conclusions based on major premises involving *few* counter-examples were accepted at a significantly higher rate compared to those involving *many* counter-examples.[6] This result can be seen as a consequence of pragmatic enrichment which is operational when *few* counter-examples are available. Finally, participants accepted more conclusions when the major premises were formulated with IFF compared to IF. This finding shows again that reasoners do differentiate between IF and IFF.

Regarding reading time, recall that according to our main hypothesis, reasoners are not expected to enrich IF unless they are compelled to do so. As discussed earlier, in the case of MP type arguments, there is no incentive to enrich, as reasoning is usually not problematic. Specifically, the conditions required to enrich are met when the major premise admits *few* counter-

[6] It should be noted that we did not observe an effect of counter-examples on MP while there are studies that reported such an effect (conclusions from MPfew were accepted at higher rate than conclusions from MPmany) (De Neys et al., 2002). This discrepancy might be due to design differences: while we asked participants to make a binary decision to accept or reject the conclusion, De Neys and colleagues provided their participants with a scale on which they had to evaluate how confident they were (Very Sure-Sure-Somewhat Sure) to accept or reject a given conclusion.

examples, contrary to *many* where enrichment is blocked by the high number of counter-examples. On the one hand, regarding IF (see Table 10), our results indeed reveal a slow-down on the minor premise during processing of AC arguments with *few* counter-examples (ACfew) compared to MP arguments with *few* counter-examples (MPfew). As expected, there is no significant difference between the reading times on the minor premise of AC and MP arguments when the major premise admits *many* counter-examples (ACmany vs MPmany). On the other hand, as far as IFF is concerned, we should not expect to observe a slow-down on the minor premise because the major premise is already in a biconditionalized form (i.e., IFF). Indeed, we do not observe differences between the reading times on minor premises of MPfew and ACfew with IFF.

4.3. Contentless vs contentful experiments

4.3.1. Results

Table 11 shows the models testing sequence for Acceptance rate for the *Contentless* versus *Contentful* experiments. The ΔAIC values support model 9 while the ΔBIC values support model 6. With the exception of the effect of Conditional on Acceptance rate, the remaining parameters of model 9 do not include zero in their respective 95% confidence intervals (see Table 12). For this reason, we regarded this model as the best model. The three-way interaction suggests that the Acceptance rate for Contentful was higher in the IFF condition than in the IF condition, albeit only when participants evaluated the affirmation of the consequent (i.e., AC condition). Furthermore, the Acceptance rate of Contentful was higher than Contentless when participants evaluated the affirmation of the consequent. Finally, the negative estimate of Argument indicates a higher Acceptance rate for the MP condition. Fig. 6 helps understanding the parameters estimates for model 6.

Table 11
Models testing sequence for Acceptance rate for the Contentless vs. Contentful experiments.

#	Fixed effects	Df	AIC	ΔAIC	BIC	ΔBIC	Log-Lik
1	Int.	3	37.09.3	321.58	3730.5	281.272	−1851.7
2	Int., Argument	4	35.47.8	160.04	3576.0	127.262	−1769.9
3	Int., Conditional	4	3711.2	323.44	3739.4	290.663	−1851.6
4	Int., Content	4	3686.8	299.13	3715.1	266.352	−1839.4
5	Int., Argument, Content, Conditional	6	3483.7	95.93	3526.1	77.299	−1735.8
6	Int., Argument * Content, Conditional	7	3399.3	11.56	3448.8	0.000	−1692.6
7	Int., Argument * Conditional, Content	7	3481.7	93.96	3531.2	82.400	−1733.8
8	Int., Conditional * Content, Argument	7	3481.8	94.09	3531.3	82.524	−1733.9
9	Int., Argument * Conditional * Content	10	3387.7	0.00	3458.4	9.651	−1683.9

Note. *Argument* = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Content* = Contentful (Ful) and Contentless (Less), *Df.* = degree of freedom, *AIC* = Akaike's Information Criterion, *ΔAIC* = differences between the winning model and remaining models, *BIC* = Bayesian information Criterion, *ΔBIC* = differences between the winning model and remaining models, *Log-Lik* = Log-Likelihood.

Table 12
Parameters estimates of Model Acceptance rate for the Contentless vs. Contentful experiments.[a]

Effect	Parameter	Estimate	95% CI	Standard error	*Z* (*p*-value)
Intercept	*Intercept.*	3.81945	[3.45, 4.22]	0.19670	19.417 (<0.001)
Slope of *Argument*	*AC-MP*	−2.05047	[−2.22, −1.89]	0.08408	−24.386 (<0.001)
Slope of *Conditional*	*IFF-IF*	0.04614	[-0.31, 0.40]	0.17966	0.257 (0.80)
Slope of *Content*	*Ful-less*	1.25129	[0.90, 1.62]	0.18437	6.787 (<0.001)
Interaction of Argument * Conditional	*AC-MP*IFF-IF*	0.26846	[0.11, 0.43]	0.08153	3.293 (<0.001)
Interaction of Argument * Content	*AC-MP*Ful-Less*	0.84290	[0.68, 1.01]	0.08227	10.246 (<0.001)
Interaction Conditional * Content	*IFF-IF*Ful-Less*	0.38289	[0.03, 0.74]	0.18047	2.122 (<0.05)
Interaction Argument * Conditional * Content	*AC-MP*IFF-IF*Ful-Less*	−0.17799	[−0.34, −0.02]	0.08143	−2.186 (<0.05)

Note. *Argument* = Affirmation of the Consequent (AC) and Modus Ponens (MP), *Conditional* = Biconditional (IFF) and Conditional (IF), *Content* = Contentful (Ful) and Contentless (Less), *AC-MP* = contrasts (i.e., differences) between Affirmation of the Consequent and Modus Ponens, *IFF-IF* = contrasts (i.e., differences) between Biconditional and Conditional, *Ful-less* = contrasts (i.e., differences) between Contentful and Contentless.

[a] Random effects for the Acceptance rate model. Random intercept for subjects = 2.4142 (Standard deviation), 95% Confidence interval [CI] = 2.12, 2.76; Random intercept for items = 0.3494 (Standard deviation), 95% Confidence interval [CI] = 0.20, 0.50.

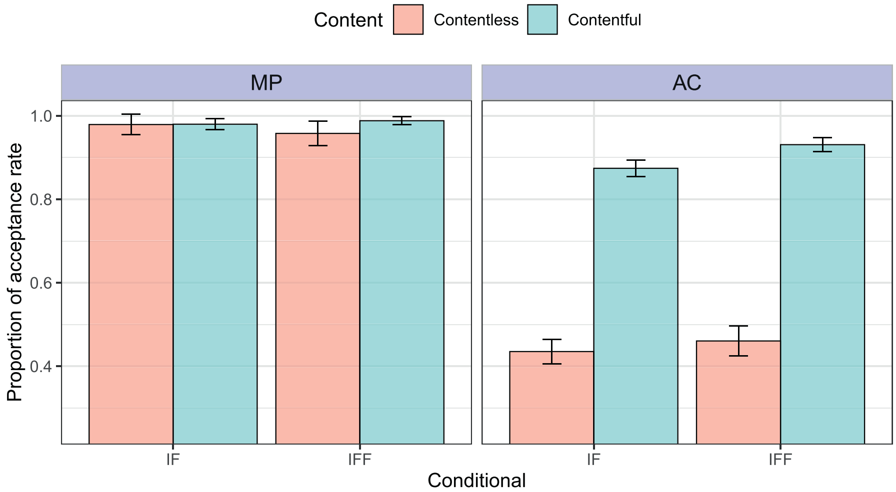

Fig. 6. Acceptance rate for the Contentless and the Contentful experiments. *MP* = Modus Ponens, *AC* = Affirmation of the Consequent, *IF* = Conditional, *IFF* = Biconditional. The within-subjects 95% confidence intervals were computed following the method proposed by Morey (2008) and were implemented using the R functions developed by Chang (2018).

4.3.2. Discussion

Acceptance rates for conclusions in *abstract* and *realistic* materials showed that overall, reasoners accept more conclusions in AC-type configurations when the reasoning is performed on abstract compared to realistic materials. We also observed significantly higher acceptance rates of AC conclusions when the major premises were formulated with IFF compared to IF. This difference was present in realistic but not abstract materials. These findings confirm our hypothesis that the pragmatic enrichment of conditionals is to be expected in reasoning with realistic materials, as in the case of causal naïve deduction.

5. General discussion and conclusion

The two self-paced reading experiments presented in this paper shed new light on a series of issues regarding the pragmatic enrichment of conditionals.

First and foremost, our findings indicate that we do not reason in a uniform manner in abstract vs realistic situations. On the one hand, reasoning with abstract materials (cf. *Contentless experiment*) is characterized by the existence of two different groups of reasoners. Endorsers systematically accept the conclusions from AC arguments and Rejecters systematically reject them. As noted earlier, this pattern has been observed in prior studies (Bonnefond et al., 2012). The new insight here comes from the results of reasoning with biconditional premises where we found exactly the same Endorsers-Rejecters split. This finding indicates that, when reasoning with abstract materials, participants were primarily sensitive to the spatial configuration of the premises of the arguments, i.e., how the premises unfold during different phases of the reasoning process (cf. Bonnefond and Van der Henst, 2009 for the importance of the reasoning sequence). This agreed well with the fact discussed in Section 3.1 that reasoning with abstract material activates a network also involved in spatial manipulation and, as such, is claimed to be similar to a form of mathematical reasoning (cf. Goel et al., 2000). On the other hand, reasoning with realistic materials (cf. *Contentful experiment*) is more homogenous, in the sense that different reasoning styles (Endorsers vs Rejecters) cannot be distinguished among participants, either when reasoners were exposed to arguments with conditionals or biconditionals major premises. This result confirms our Hypothesis 1 that reasoners exhibit different behavior when reasoning with abstract vs realistic materials, independently of the logical word (IF or IFF) used to formulate the major premise.

Let us now answer our main research question concerning *pragmatic enrichment* of conditionals, that is, a move from conditional IF to biconditional IFF. First, we will consider reasoning with abstract materials and, then we will turn to reasoning with realistic materials.

Regarding ***abstract materials***, a measure that we took into account as indicative that a process of pragmatic enrichment of conditional (IF) took place was reading time on minor premises: a slow-down on the minor premise of AC compared to MP arguments was considered to be a sign of extra cognitive cost associated with the process of enriching. Another auxiliary measure was the acceptance rate of conclusions: we assumed that if a process of pragmatic enrichment of IF occurs, a higher rate of acceptance of conclusions would consequently follow. Importantly, our experimental design allowed us to go a step further because we assessed reasoning with two types of major premises: IF (*Contentless-IF*) and IFF (*Contentless-IFF*). Since IFF is a biconditional form, it does not need to be pragmatically enriched and therefore we should observe different patterns for these two types of major premises. In other words, the *Contentless-IFF* condition can be seen as a ‘control’ for the *Contentless-IF* condition. Table 13 below summarizes the main relevant points taken into account in the interpretation of the results of the *Contentless experiment*.

Table 13
Summary of the main points relevant for the analysis of the *Contentless experiment* with Endorsers-Rejecters split.

	Endorsers		Rejecters	
	IF	IFF	IF	IFF
Do we find high acceptance rate of AC conclusions?	Yes	Yes	No	No
Is it logically valid?	No	Yes	Yes	No
Is it compatible with pragmatic meaning?	Yes	Yes	No	No
Do we find a slow-down on minor of AC compared to MP?	No	No	Yes	No

Consider first the results concerning **Endorsers** (first column of Table 13). A priori, Endorsers in the IF condition could be considered as potential Pragmatic Enrichers since they display a high rate of acceptance of conclusions from AC arguments. However, we did not observe a slow-down on the minor premises, a result which would be indicative of the process of pragmatic enrichment. What is more, the same pattern is observed for IFF, that is, reasoners massively accept the conclusions from AC configurations (which are logically correct) and at the same time there is no slow-down in their reading times on the minor premises of AC.

Regarding **Rejecters** (second column of Table 13), as these reasoners overwhelmingly reject the conclusions of AC arguments with IF, they could be considered as Logicians, since their rejection of the AC conclusions with IF is in accordance with the precepts of classical logic. However, the results from the *Contentless-IFF* condition prevent us from drawing such a conclusion, since exactly the same behavior (i.e., systematic rejection of conclusions from AC type configurations) is observed with IFF, the pattern which does not fit with the rules of classical logic. Moreover, it is interesting to observe that Rejecters in the IF condition are the only ones to slow-down on minor premises of AC compared to MP, and based on this result, they could be considered as Pragmatic Enrichers. However, Rejecters happen to reject the conclusion of AC instead of accepting it, as we would expect for Pragmatic Enrichers. It is also worth noting that Rejecters do not slow-down on the minor premise of AC in the IFF condition. This finding demonstrates in addition that reasoners do distinguish between conditionals and biconditionals already at the level of abstract reasoning.

Thus, we conclude that neither the Endorser nor the Rejecter style of reasoning can be identified as being the result of pragmatic enrichment. Instead, both styles of reasoning are indicative of a strategy that reasoners adopt when presented with abstract logical problems. This strategy consists in either systematically rejecting or systematically accepting conclusions following the minor premises related to the second segment of the major premises (i.e. AC type configurations) and is present in both IF and IFF. In sum, pragmatic enrichment for abstract reasoning was not attested (cf. Hypothesis 2).

We now turn to *pragmatic enrichment* of conditionals in reasoning with ***realistic materials***. First, it is important to point out that we did find differences between conditional and biconditional reasoning with realistic materials, confirming that reasoners distinguish between the two logical words. To verify the presence of pragmatic enrichment in realistic materials, we chose causal reasoning taking into account two well-known sub-classes of causal statements: those involving *few* vs *many* counter-examples (Cummins et al., 1991; Cummins, 1995) because the two types do not display the same potential for biconditionalization: causal conditional statements with *few* possible counter-examples more naturally lead to biconditionalization than causal conditional statements for which *many* counter-examples are possible. Based on Franke (2009), we proposed that the pragmatic mechanism for enrichment of conditionals appeals to the normality assumption according to which *normally* an event would not happen in the absence of other unexpected intervening events for a given causal situation. On the one hand, the normality assumption holds in causal situations with *few* counter-examples, as no other intervening events can be thought of that would prevent a causal effect to occur. On the other hand, the normality assumption does not hold in causal situations with *many* counter-examples, as one can imagine other events that could have intervened and changed the causal chain of events. In sum, while the normality assumption is taken to be true by the comprehenders, a given conditional is enriched to a corresponding biconditional whereas when the normality assumption is not tenable, the biconditional interpretation does not arise.

Practically, the measure indicative of enrichment was the same as in the *Contentless experiment* (i.e. Reading time) except that we took into account the *few* vs *many* distinction that was not present in our abstract materials. We found that the slow-down appears with *few* but not with *many* counter-examples while reasoners process AC compared to respective MP arguments with IF (*Contentful-IF*). Such results suggest that reasoners enrich when the meaning of the major premise invites them to do so, that is, when the rule admits *few* counter-examples (cf. Hypothesis 3). Again, an independent confirmation comes from our *Contentful-IFF* condition, in which we did not find such a difference.[7]

Moreover, if pragmatic enrichment occurs, its consequences could be observed on the rate of acceptance of conclusions so that more conclusions would be accepted when the major premise had *few* counter-examples to it. Indeed, we found an overall higher acceptance rate of conclusions based on major premises being subject to *few* compared to *many* counter-examples. This result is buttressed by the fact that more conclusions have been accepted with IFF compared to IF.

[7] To complete the picture, it should be noted that the effects regarding the reading time for the minor premises were not observed for the conclusions (see supplementary materials for the detail of the results). As we pointed from the beginning, the pragmatic enrichment was expected to occur during the treatment of the minor premise.

Therefore, we can conclude that these results confirm our Hypothesis 3 that reasoners enrich conditionals when (i) the type of conditional used as major premise invites enrichment, i.e. the conditional premise that admits *few* counter-examples and (ii) they are prompted by some trigger, such as AC type of arguments.

In summary, our two experiments gave rise to a series of novel findings. Two main dissimilarities between reasoning with abstract vs realistic materials have been shown: (i) the Endorsers-Rejecters split was observed in abstract but not in realistic reasoning and (ii) differences between the two linguistic formulations of the major premise (IF vs IFF), observed both in accuracy and reading time, were present only while participants performed reasoning with realistic materials. Concerning the process of pragmatic enrichment, the findings of our study are compatible with the hypothesis that such process cannot be present in reasoning with abstract materials because the presence of meaningful information is a necessary element for reasoners to gauge the relevance of pragmatic enrichment. Indeed, realistic reasoning encouraged participants to enrich conditionals with *few* counter-examples while no enrichment has been found in abstract reasoning. Table 14 summarizes the main differences between abstract and realistic type of reasoning found in our study. In further research, it would be particularly informative to rely on the EEG technique to determine the electrophysiological signature of the pragmatic enrichment manifested in the present study and the ERP components (N2, Pb3 or otherwise) that are elicited when reasoning about realistic content with both conditional and biconditional formulations.

Table 14
Summary of the results for *Contentless* and *Contentful* experiments.

	Abstract	Realistic
Endorsers-Rejecters split	Yes for IF & IFF	No
Different treatment of IF and IFF	No	Yes
Presence of Pragmatic Enrichment	No	Yes

Funding

This work was supported by a grant [grant number IZK0Z1_175538] from the Swiss National Science Foundation (FNS) awarded to Joanna Blochowiak.

Declaration of competing interest

The authors have no competing interests to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pragma.2022.05.013.

References

Atlas, J.D., Levinson, S.C., 1981. It-clefts, informativeness and logical form: radical pragmatics. In: Cole, Peter (Ed.), Radical Pragmatics. Academic Press.
Barrouillet, P., Grosset, N., Lecas, J.F., 2000. Conditional reasoning by mental models: chronometric and developmental evidence. Cognition 75 (3), 237–266.
Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.
Blochowiak, J., 2019. Les fondements logiques de l'interprétation de la conditionnelle dans le raisonnement causal. nCLF 33, 167–190.
Blochowiak, J., Castelain, T., 2018. How logical is natural language conjunction? An experimental investigation of the French conjunction et. In: Language, Evolution and Mind: Essays in Honour of Anne Reboul. College Publications, pp. 129–157.
Blochowiak, J., Moeschler, J., Castelain, T., 2010. The impact of order and aspect in processing of causally linked sentences. In: Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics, pp. 13–16.
Bonnefond, M., Van der Henst, J.B., 2009. What's behind an inference? An EEG study with conditional arguments. Neuropsychologia 47 (14), 3125–3133.
Bonnefond, M., Van der Henst, J.B., 2013. Deduction electrified: ERPs elicited by the processing of words in conditional arguments. Brain Lang. 124 (3), 244–256.
Bonnefond, M., Van der Henst, J.B., Gougain, M., Robic, S., Olsen, M.D., Weiss, O., Noveck, I., 2012. How pragmatic interpretations arise from conditionals: profiling the Affirmation of the Consequent argument with reaction time and EEG measures. J. Mem. Lang. 67 (4), 468–485.
Bonnefond, M., Noveck, I., Baillet, S., Cheylus, A., Delpuech, C., Bertrand, O., et al., 2013. What MEG can reveal about inference making: the case of if … then sentences. Hum. Brain Mapp. 34 (3), 84–697.
Bonnefond, M., Kaliuzhna, M., Van der Henst, J.B., De Neys, W., 2014. Disabling conditional inferences: an EEG study. Neuropsychologia 56, 255–262.
Braine, M.D.S., O'Brien, D.P., 1991. A theory of if: a lexical entry, reasoning program, and pragmatic principles. Psychol. Rev. 98 (2), 182–203.
Braine, M.D.S., O'Brien, D.P., 1998. Mental Logic. Lawrence Erlbaum Associates, Mahwah, New Jersey.
Breheny, R., Katsos, N., Williams, J., 2006. Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. Cognition 100 (3), 434–463.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Inference: A Practical Information-Theoretic Approach, second ed. Springer-Verlag, New York. https://doi.org/10.1007/b97636.
Chang, W., 2018. R Graphics Cookbook, second ed. O'Reilly Media, Inc, Champaign.
Chierchia, G., 2013. Logic in Grammar: Polarity, Free Choice, and Intervention. OUP, Oxford.
Cornulier, B., 1985. Sur un si d'énonciation prétendument non-conditionnel. Rev. Quebecoise Linguist. 15 (1), 197–211.
Cummins, D.D., 1995. Naive theories and causal deduction. Mem. Cognit. 23 (5), 646–658.
Cummins, D.D., 2014. The impact of disablers on predictive inference. J. Exp. Psychol. Learn. Mem. Cognit. 40 (6), 1638–1655.

Cummins, D.D., Lubart, T., Alksnis, O., Rist, R., 1991. Conditional reasoning and causation. Mem. Cognit. 19 (3), 274–282.
De Neys, W., Everaerts, D., 2008. Developmental trends in everyday conditional reasoning: the retrieval and inhibition interplay. J. Exp. Child Psychol. 100 (4), 252–263.
De Neys, W., Schaeken, W., d'Ydewalle, G., 2002. Causal conditional reasoning and semantic memory retrieval: a test of the semantic memory framework. Mem. Cognit. 30 (6), 908–920.
De Neys, W., Schaeken, W., d'Ydewalle, G., 2005a. Working memory and everyday conditional reasoning: retrieval and inhibition of stored counterexamples. Think. Reas. 11 (4), 349–381. https://doi.org/10.1080/13546780442000222.
De Neys, W., Schaeken, W., d'Ydewalle, G., 2005b. Working memory and counterexample retrieval for causal conditionals. Think. Reas. 11 (2), 123–150. https://doi.org/10.1080/13546780442000105.
Degen, J., Tanenhaus, M.K., 2016. Availability of alternatives and the processing of scalar implicatures: a visual world eye-tracking study. Cognit. Sci. 40 (1), 172–201.
Evans, J.S.B., 1993. The mental model theory of conditional reasoning: critical appraisal and revision. Cognition 48 (1), 1–20.
Evans, J.S.B., Ellis, C.E., Newstead, S.E., 1996. On the mental representation of conditional sentences. Q. J. Exp. Psychol.: Hum. Exp Psychol 49A (4), 1086–1114.
Evans, J.S.B., Handley, S.J., Neilens, H., Over, D.E., 2007. Thinking about conditionals: a study of individual differences. Mem. Cognit. 35 (7), 1772–1784.
Fernbach, P.M., Darlow, A., Sloman, S.A., 2010. Neglect of alternative causes in predictive but not diagnostic reasoning. Psychol. Sci. 21 (3), 329–336.
Foppolo, F., Marelli, M., 2017. No delay for some inferences. J. Semant. 34 (4), 659–681.
Franke, M., 2009. Signal to Act. PhD thesis. University of Amsterdam, The Netherlands.
Geis, M.L., Zwicky, A.M., 1971. On invited inferences. Linguist. Inq. 2 (4), 561–566.
Girotto, V., Mazzocco, A., Tasso, A., 1997. The effect of premise order in conditional reasoning: a test of the mental model theory. Cognition 63 (1), 1–28.
Goel, V., 2007. Anatomy of deductive reasoning. Trends Cognit. Sci. 11 (10), 435–441.
Goel, V., 2009. Fractionating the system of deductive reasoning. In: Kraft, E., Guylás, B., Pöppel, E. (Eds.), Neural Correlates of Thinking. Springer Press, Berlin.
Goel, V., Buchel, C., Frith, C., Dolan, R.J., 2000. Dissociation of mechanisms underlying syllogistic reasoning. Neuroimage 12 (5), 504–514.
Grice, H.P., 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), Syntax and Semantics: Vol. 3, Speech Acts. Academic Press, New York, pp. 41–58.
Grodner, D.J., Klein, N.M., Carbary, K.M., Tanenhaus, M.K., 2010. "Some," and possibly all, scalar inferences are not delayed: evidence for immediate pragmatic enrichment. Cognition 116 (1), 42–55. https://doi.org/10.1016/j.cognition.2010.03.014.
Guasti, T.M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., Meroni, L., 2005. Why children and adults sometimes (but not always) compute implicatures. Lang. Cognit. Process. 20 (5), 667–696.
Horn, L.R., 1972. On the Semantic Properties of Logical Operators in English. PhD thesis. University of California, Los Angeles.
Horn, L.R., 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature (pp. 11–42). In: Shiffrin, D. (Ed.), *Meaning, Form, and Use in Context: Linguistic Applications*. Georgetown University Press, Washington.
Horn, L.R., 1989. A Natural History of Negation. Chicago University Press, Chicago.
Horn, L.R., 2000. From if to iff: conditional perfection as pragmatic strengthening. J. Pragmat. 32 (3), 289–326.
Johnson-Laird, P.N., 2001. Mental models and deduction. Trends Cognit. Sci. 5 (10), 434–442.
Johnson-Laird, P.N., Byrne, R.M.J., 1991. Deduction. Laurence Erlbaum Associates, Hillsdale, New Jersey.
Lander, J.P., 2018. Coefplot: Plots Coefficients from Fitted Models. R package version 1.2.6. https://CRAN.R-project.org/package=coefplot.
Levinson, S.C., 2000. Presumptive Meanings. The Theory of Generalized Conversational Implicature. MIT Press, Cambridge, Massachusetts.
Lüdecke, D., 2018. Sjmisc: data and variable transformation functions. J. Open Source Softw. 3 (26), 754. https://doi.org/10.21105/joss.00754.
Lüdecke, D., 2020a. sjPlot: Data Visualization for Statistics in Social Science. https://doi.org/10.5281/zenodo.1308157. R package version 2.8.3. https://CRAN.R-project.org/package=sjPlot.
Lüdecke, D., 2020b. Sjlabelled: Labelled Data Utility Functions (Version 1.1.3). https://doi.org/10.5281/zenodo.1249215. https://CRAN.R-project.org/package=sjlabelled.
Markovits, H., 2017. In the beginning stages: conditional reasoning with category-based and causal premises in 8- to 10-year olds. Cognit. Dev. 41, 1–9.
Markovits, H., Barrouillet, P., 2002. The development of conditioned reasoning: a mental model account. Dev. Rev. 22 (1), 5–36. https://doi.org/10.1006/drev.2000.0533.
Markovits, H., Quinn, S., 2002. Efficiency of retrieval correlates with "logical" reasoning from causal conditional premises. Mem. Cognit. 30 (5), 696–706. https://doi.org/10.3758/BF03196426.
Markovits, H., Brisson, J., de Chantal, P.L., 2015. Deductive updating is not Bayesian. J. Exp. Psychol. Learn. Mem. Cognit. 41 (4), 949.
Matsumoto, Y., 1995. The conversational condition on Horn scales. Linguist Philos. 18 (1), 21–60.
Moeschler, J., Chevallier, C., Castelain, T., Van der Henst, J.-B., Tapiero, I., 2006. Le raisonnement causal: de la pragmatique du discours à la pragmatique expérimentale. CLF 27, 241–262.
Morey, R.D., 2008. Confidence intervals from normalized data: a correction to Cousineau (2005). Tutor. Quant. Methods Psychol. 4 (2), 61–64.
Noveck, I., 2001. When children are more logical than adults: experimental investigations of scalar implicature. Cognition 78 (2), 165–188.
Noveck, I., Reboul, A., 2008. Experimental pragmatics: a Gricean turn in the study of language. Trends Cognit. Sci. 12 (11), 425–431.
Noveck, I., Bonnefond, M., Van der Henst, J.-B., 2011. Squib: a deflationary account of invited inferences. Belg. J. Linguist. 25 (1), 195–208.
Oaksford, M., Chater, N., 2007. Bayesian Rationality: the Probabilistic Approach to Human Reasoning. Oxford University Press.
Oaksford, M., Chater, N., 2013. Dynamic inference and everyday conditional reasoning in the new paradigm. Think. Reas. 19 (3–4), 346–379.
Oaksford, M., Chater, N., 2016. Probabilities, causation, and logic programming in conditional reasoning: reply to Stenning and Van Lambalgen (2016). Think. Reas. 22 (3), 336–354.
Papafragou, A., Musolino, J., 2003. Scalar implicatures: experiments at the semantics–pragmatics interface. Cognition 86 (3), 253–282. https://doi.org/10.1016/S0010-0277(02)00179-8.
Paris, S.G., 1973. Comprehension of language connectives and propositional logical relationships. J. Exp. Child Psychol. 16 (2), 278–291.
Pouscoulous, N., Noveck, I., Politzer, G., Bastide, A., 2007. A developmental investigation of processing costs in implicature production. Lang. Acquis. 14 (4), 347–375.
Ronai, E., Xiang, M., 2021. Pragmatic inferences are QUD-sensitive: an experimental study. J. Linguist. 57 (4), 841–870. https://doi.org/10.1017/S0022226720000389.
Rottman, B.M., Hastie, R., 2014. Reasoning about causal relationships: inferences on causal networks. Psychol. Bull. 140 (1), 109–139.
RStudio Team, 2020. RStudio. Integrated Development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com/.
Rumain, B., Connell, J., Braine, M.D., 1983. Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: if is not the biconditional. Dev. Psychol. 19 (4), 471–481.
Sarkar, D., 2008. Lattice: Multivariate Data Visualization with R. Springer, New York. http://lmdvr.r-forge.r-project.org.
Schneider, W., Eschman, A., Zuccolotto, A., 2002. E-prime (Version 2.0). Computer Software and Manual. Psychology Software Tools Inc, Pittsburgh, PA.
Su, Y.E., 2014. The acquisition of logical connectives in child Mandarin. Lang. Acquis. 21 (2), 119–155.
van Canegem-Ardijns, I., van Belle, W., 2008. Conditionals and types of conditional perfection. J. Pragmat. 40 (2), 349–376.
Van der Auwera, J., 1997. Pragmatics in the last quarter century: the case of conditional perfection. J. Pragmat. 27 (3), 261–274.
Verschueren, N., Schaeken, W., De Neys, W., d'Ydewalle, G., 2004. The difference between generating counterexamples and using them during reasoning. Q. J. Exp. Psychol. Sec. A 57 (7), 1285–1308.
Verschueren, N., Schaeken, W., Verbrugge, S., 2006. Conversational implicatures in counterexample retrieval: working memory and crystallized pragmatics. In: Proceedings of the 28th Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates, Mahwah, New Jersey, USA, pp. 2311–2316.

von Fintel, K., 2001. Conditional Strengthening – A Case Study in Implicature. Unpublished manuscript.
Wasserman, L., 2000. Bayesian model selection and model averaging. J. Math. Psychol. 44 (1), 92–107. https://doi.org/10.1006/jmps.1999.1278.
Wickham, H., 2017. Tidyverse: Easily Install and Load the 'Tidyverse'. R Package Version 1.2.1. https://CRAN.R-project.org/package=tidyverse.

Joanna Blochowiak is a senior postdoctoral researcher at the Faculty of Psychology and Educational Sciences working on reasoning with generics from theoretical and empirical perspectives. She graduated from the University of Geneva with a dissertation in formal semantics and pragmatics on 'why'-questions and their answers with 'because'. During her PhD she spent one year at Harvard University focusing on 'why' questions from formal semantics perspective. In recent years, she investigated within different postdoctoral projects a series of research topics in the field of experimental pragmatics including the study logical words such as negation and conjunction (University of Geneva), conditionals and biconditionals (RuCCS at Rutgers University), epistemic aspects of causal connectives as 'parce que', 'car' and 'donc' as well as the expression of speaker's subjectivity in argumentation (Université catholique de Louvain, Belgium).

Thomas Castelain is a Serra Húnter Fellow at the Department of Psychology at the University of Girona (Spain). His main areas of interest are reasoning, argumentation, communication and social cognition. During the last years, his work has focused on the question of the function of human reasoning through an intercultural and developmental perspective. His field of research also includes epistemic vigilance, language comprehension and processing of causality and irony.

Odir Rodríguez-Villagra is an Associated Professor at the University of Costa Rica. His main areas of interest are the study of the sources of variation (individual, group and developmental) in working memory capacity and its relationship to fluid intelligence, academic performance and reading comprehension. In addition, over the years he has also been interested in the study of executive functions. He tackles these questions through statistical modeling, mathematical modeling and, recently, using electroencephalography and eye-tracking.

Julien Musolino is an Associated Professor at Rutgers University where he currently holds a dual appointment in the Psychology Department and the Rutgers University Center for Cognitive Science. He is interested in our capacity to acquire language, to reason, and to think critically. This work, based on his training in theoretical linguistics, cognitive/developmental psychology, and speech-language pathology, takes an interdisciplinary approach and integrates concepts, methods, and experimental techniques from these different fields. The goal of this work is to understand how these capacities are acquired, represented, and put to use by both typical and atypical populations of human beings.