

Testing the influence of QUDs on the occurrence of Conditional Perfection

Britta Grusdt, Mingya Liu, & Michael Franke*

Abstract. In natural language conversations, speakers often communicate ‘if and only if’ when they say ‘if’. The reasons why in some circumstances, yet not all, conditionals receive a biconditional interpretation remain under investigation. Von Stechow (2001) proposed an account where the interpretation of a conditional (“if p, then q”) is predicted to depend on the focus of the conversation which may either lie on the conditions that make the consequent, q, true or on the consequences following when the antecedent, p, is true. To test this account, we present two novel behavioral experiments with non-text based stimuli that take advantage of participants’ intuitive understanding of physics. We find some supporting evidence for the tested account that is not conclusive but suggests that other aspects, like the nature of potential alternative causes for the consequent to become true (e.g., with or w/o the influence of an external variable), also play a role for the interpretation of the conditional.

Keywords. conditional perfection; indicative conditionals; pragmatics; question under discussion; exhaustivity; experimental

1. Introduction. A longstanding subject of research in the context of indicative conditionals — natural language expressions of the form “if p, q” ($p \rightarrow q$) where p is referred to as antecedent and q as consequent— is their interpretation as biconditionals, a phenomenon known as *conditional perfection* (CP) (Geis & Zwicky 1971). A conditional is said to be perfected when ‘if’ communicates ‘if and only if’, leaving unaddressed how exactly this inference comes about, e.g. through one of the inferences ‘only if p, q’, ‘if not p, not q’ or ‘if q, p’ (see Van Canegem-Ardijns & Van Belle (2008) who propose different types of conditional perfection). For some conditionals, a perfected interpretation seems to be readily endorsed, even when no concrete context is given; an eminent example are promises and threats communicated with conditionals like (1) below. Here, the biconditional interpretation is forthcoming naturally: the speaker seems to communicate to scratch the addressee’s back if and only if the addressee scratches the speaker’s back.

(1) If you scratch my back, I’ll scratch yours.

In the literature on conditional reasoning, participants are usually tested on four inferences, *Denying the antecedent* (DA), *Affirming the consequent* (AC), *Modus Ponens* (MP) and *Modus Tollens* (MT) to investigate their interpretation of conditionals.¹ Since, contrary to MP and MT, DA and AC are only logically valid in the case of biconditionals, high endorsement rates of the latter two inferences suggest a biconditional interpretation; Fillenbaum (1986) reports endorsement rates of DA for conditional premises and threats between 80 and 90%, whereas they tend to

*We would like to thank Malin Spaniol and Josefine Zerbe for their great support in implementing the experiments. Authors: Britta Grusdt, Universität Osnabrück (britta.grusdt@uni-osnabrueck.de) & Mingya Liu, Humboldt Universität zu Berlin (mingya.liu@hu-berlin.de) & Michael Franke, Eberhard Karls Universität Tübingen (michael.franke@uni-tuebingen.de).

¹Tabular overview of the four classical conditional reasoning inferences: ... continues on next page ...

be (much) lower for other types of conditionals (e.g. see Evans et al. 1993; Chapter 2 for a summary of studies). Another factor that has been shown to influence participants’ interpretation of a conditional as biconditional is the availability of alternative causes or disabling conditions for the consequent, making a biconditional interpretation more likely when fewer alternative causes are conceivable (Cummins et al. 1991, Markovits 1986). This finding connects with the hypothesis of how CP arises that we aim to test in this paper, as we will see shortly.

In the linguistic literature, CP-readings of conditionals have often been treated as conversational implicatures (Atlas & Levinson 1981, Van der Auwera 1997, Horn 2000). While Geis & Zwicky, who (re)initiated the debate about CP among linguists,² argue that conditionals are quite commonly attributed a CP-reading, this regularity has been questioned by others providing various counterexamples (e.g., Lilje 1972, de Cornulier 1983). Von Stechow (2001) goes one step further by making a proposal of when exactly a conditional is interpreted as biconditional and when it is not, which had not been precisely formulated in previous accounts. Similar to de Cornulier (1983), Von Stechow refers to exhaustivity: he argues that a biconditional interpretation arises when the antecedent of a conditional is interpreted as exhaustive list of *conditions for the consequent* whereas it is not triggered when the conditional is interpreted as exhaustive list of *consequences of the antecedent*. In other words, when the speaker is required to provide an exhaustive list of conditions for the consequent (q) and only mentions a single condition (p) the listener will infer that p is a sufficient and necessary condition for q, corresponding to a CP-reading. On the other hand, mentioning p as single condition in the antecedent does not trigger a CP-reading when the speaker is asked to provide an (exhaustive) list of consequences of p. According to Von Stechow (2001), a (possibly implicit) question under discussion (QUD) determines whether the conditional targets the conditions of the consequent (e.g., under which conditions q?) or the consequences of the antecedent (e.g., what follows from p?). To illustrate the hypothesized effect of the QUD, consider the conditional in (2), inspired by an example from Lilje (1972):

(2) If a ball bounces off the table, it is a foul.

In a situation where a person A explains the rules of the game pool to a person B who has no experience with this game and where B asks A ...

- (i) what happens if a ball bounces off the table (QUD: *if-p*)
- (ii) which actions count as foul / whether there are actions that count as foul (QUD: *when-q*)

the same answer — the conditional in (2) — seems to be interpreted as biconditional only when the conversation is guided by the question in (ii). Given the context provided by the QUD *if-p* (i), the speaker is not expected to mention all possible actions that are considered a foul and, thus, CP does not arise in this case.

premise 1	premise 2	conclusion	inference	logically valid
$p \rightarrow q.$	$p.$	$\therefore q.$	Modus Ponens (MP)	✓
$p \rightarrow q.$	$\neg q.$	$\therefore \neg p.$	Modus Tollens (MT)	✓
$p \rightarrow q.$	$q.$	$\therefore p.$	Affirming the Consequent (AC)	✗
$p \rightarrow q.$	$\neg p.$	$\therefore \neg q.$	Denying the Antecedent (DA)	✗

²As noted by Van der Auwera (1997), CP had already been discussed before Geis & Zwicky (1971), e.g. by Ducrot (1969).

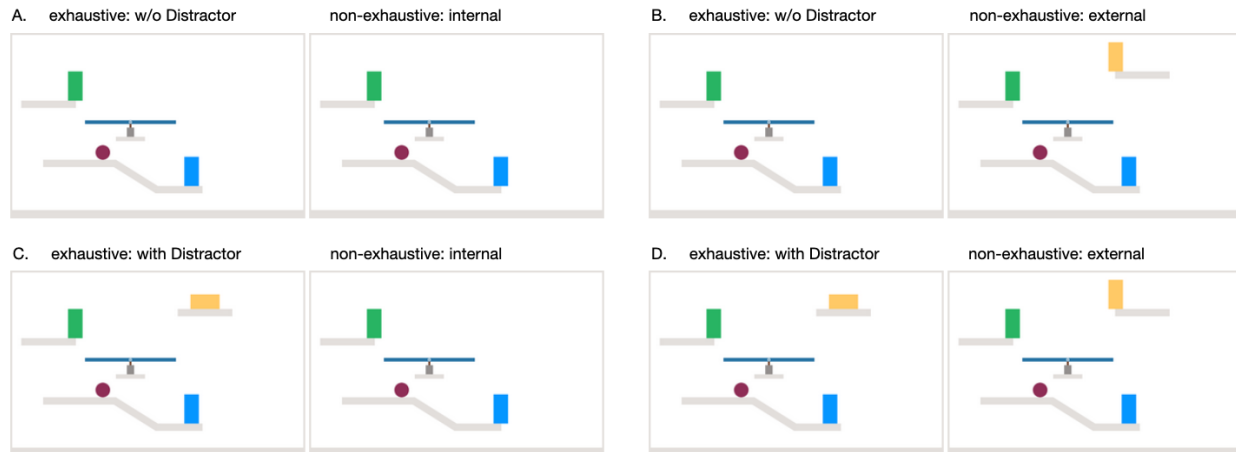


Figure 1: Exhaustive and non-exhaustive situations of critical trials in Experiment 2. The scenes in Experiment 1 were identical except for the yellow distractor block in stimuli C and D which was also centered on the platform, but standing on its short side. For simplicity, in all pictures shown here, the ant-block is green and thus, the cons-block blue; in the experiments, the color of the antecedent- and the cons-block was randomly chosen for each participant and trial.

In this paper, we present data from two novel behavioral experiments that we designed to investigate the influence of QUDs on participants’ interpretation of conditionals as biconditionals. More precisely, we aim to test whether a QUD that puts the focus of the conversation on the antecedent of the conditional, by asking about the conditions for the consequent (QUD: *will-q*), positively influences a biconditional interpretation of the conditional, as compared to a QUD that puts the focus of the conversation on the consequent by asking about the consequences of the antecedent (QUD: *if-p*). In both experiments, participants are shown scenes of toy blocks together with a dialog between two characters that consists of a question, the QUD, and a conditional answer. Participants’ task is to select the scene(s) that they believe to be best described by the conditional. The set of scenes among which participants have to choose contains at least two scenes, an *exhaustive*, and a *non-exhaustive* situation, as we call them. Both situations respectively contain (possibly among others) a blue and a green block, one in the upper left, the other in the lower right of the scene, where the falling of the upper block causes the lower block to fall as well; see Figure 1 for the critical situations from Experiment 2. Since the conditional answer is always the same — “If the upper left block falls, the lower block will fall” where ‘upper left’ and ‘lower’ are replaced by the respective color (green or blue) — we refer to the upper left block as the ‘ant-block’, mentioned in the antecedent, and to the lower block as the ‘cons-block’, mentioned in the consequent. The two situations are manipulated with respect to the number of conceivable causes for the cons-block to fall. While, in both situations, the cons-block will fall if the ant-block falls, only in the non-exhaustive situation there is a second possible reason for the cons-block to fall: either because of its own position on the edge of the platform (condition *internal*, Figure 1 stimulus A/C) or because of the falling of an additional block (condition *external*, Figure 1 stimulus B/D). The idea is that when participants interpret the conditional “If the ant-block falls, the cons-block

will fall” as biconditional, they should prefer to select the exhaustive situation as better described by the conditional. The difference between the two experiments concerns the concrete setup as we will explain below.

Previous studies have tested a QUD-effect on the occurrence of CP as proposed by Von Fintel (2001), though yielding conflicting results (Cariani & Rips 2016, Farr 2011, Cariani & Rips forthcoming). Farr’s (2011) results provide quite strong evidence for the hypothesized effect of the QUD, whereas only a minute effect, if any at all, was found by Cariani & Rips (2016). In the experiment from Farr, participants read short vignettes and were asked whether a given conditional (e.g., “If you sell an eel, you get 2.50 euros”) is a sufficient answer to a question, encoding the QUD (e.g., “What happens if I sell an eel?” vs. “When do I get 2.50 euros?”). It may be considered problematic to ask for the *sufficiency* of the conditional answer (e.g., see Cariani & Rips 2016, López Astorga 2014): the vignettes describe two possibilities to achieve the consequent (e.g., both, an eel and a pike cost 2.50), so that a no-answer to the question “Did Sahra answer Kerstin’s question sufficiently?” does not necessarily imply — even though it may strongly suggest — that participants interpreted Sahra’s conditional answer (“If you sell an eel, you get 2.50 euros”) to Kerstin’s question (“When do I get 2.50 euros?”, QUD: *will-q*) as biconditional. Cariani & Rips (2016) avoid this problem by measuring participants’ endorsement rates of the four inferences mentioned above (MP, MT, AC, DA) to investigate the degree to which participants’ interpreted a conditional as biconditional. However, the experimental stimuli from Cariani & Rips (2016) — again short vignettes — come along with world knowledge that is hard to control for. For instance, in one of their trials participants learned that ‘John has taken a test on Chapters 4-6 that has not been graded yet’. They were then asked whether the conditional (that they were told was true) ‘If John understood Chapter 5, then John did well on the test’ implies that ‘John understood Chapter 5’ when they also know that ‘John did well on the test’ (testing AC). As Cariani & Rips note themselves, participants might assume that the conditional simply does not state all conceivable conditions for the consequent; in this example, it is hard *not* to think of other reasons why John could do well on the test without having understood Chapter 5 (e.g. by cheating), which may have influenced participants’ responses.

An advantage of our non-text-based stimuli is that they allow to control participants’ elicited beliefs about the situations at hand much better. By showing participants animations of how the blocks behave, we hope to reduce the influence of additional world knowledge further. Also, our measurement for how the conditional is interpreted does not involve a direct question about the sufficiency of the conditional as answer to the QUD; we make participants select the situation in which they believe the conditional to be more appropriate.

To anticipate our results, we find some evidence for an influence of the QUD in the predicted way, yet the QUD cannot fully explain the observed data. The results are nonetheless interesting as they suggest other aspects to play a role for the occurrence of CP, like the nature of the potential alternative causes (external vs. internal), leading to different sets of alternative utterances as well as different possible types of interpretations (causal vs. epistemic).

2. Experiment 1. We preregistered the experiment based on a pilot study, in which we collected and analyzed data from 25 subjects. The code and the preregistration report are available on OSF.³

³<https://osf.io/47w85>, <https://tinyurl.com/255yaztv>

Participants A total of 300 participants were recruited via the online platform Prolific, including the 25 participants from our pilot study.⁴ All of them were self-reported native English speakers, at least 18 years old and had an approval rate on Prolific of at least 80%. The cleaned data (see below) comprises 282 participants (103 male, 175 female, 3 other, 1 not specified) with a mean age of 32.8 years (range 18 – 84). For their participation, each participant received £1.25.

Setup & Materials The experiment consisted of a training phase with 7 trials and a testing phase with 18 trials. 12 of the test trials were critical trials, the remaining 6 were control trials including 3 attention-checks. In the training phase, participants saw animations of block arrangements that were created with the rigid body physics engine ‘matter.js’.⁵ The pictures shown in the test phase are screenshots (800 × 500 pixel) of the corresponding animations right before they would start.

Manipulations. The manipulated variables comprise the QUD as encoded in Ann’s question and the shown pair of situations (exhaustive/non-exhaustive). The QUD has the following three levels: *neutral*: “Which blocks do you think will fall?”, *if-p*: “What happens if the ant-block falls?” and *will-q*: “Will the cons-block fall?”. The exhaustive and the non-exhaustive situation have two levels each: the former either contains or does not contain a yellow distractor block in the upper right (*exhaustive*: with distractor, w/o distractor) and in the latter, the second cause why the cons-block might fall is either due to its own position (*non-exhaustive*: internal) or due to the falling of a third block (*non-exhaustive*: external), as shown in Figure 1.⁶

Training phase The main purpose of the training phase was to familiarize participants with the physical properties of the blocks. To induce a maximal degree of uncertainty in the critical test trials about whether the ant-block will fall, the blocks in the training trials are all positioned such that it should be quite easy to judge whether they will fall, in particular after having seen a few examples.⁷ Contrary to that, in the critical trials of the test phase the center of the ant-block lies exactly on the edge of the platform. The order of the training trials was randomized within participants, but always alternated between trials where some blocks fall and trials where nothing happens. In each training trial participants were first asked to select all blocks that they believed to fall, by clicking on buttons with the respective block icons (or saying ‘none’). Only then, they were able to click on ‘RUN’ to start the animation to see which blocks actually fall and whether their selection was correct. We explicitly asked participants to look at all blocks shown in the scene to encourage them to consider the potential influence of each block on any other block.

Test phase In the test phase participants read a dialog between two characters, consisting of a question from Ann and an answer to that question from Bob. After reading the dialog, participants were shown two pictures of block arrangements and were asked to select the one that they rated as

⁴Note that, since two prolific ids were erroneously recorded twice, we eventually recorded 302 instead of the initially planned 300 participants such that all 300 data sets are ensured to come from 300 distinct participants. From the two data sets that were associated with the same prolific id, the one with the later time stamp was excluded.

⁵<https://brm.io/matter-js/>

⁶In the exhaustive situations, the distractor block never moves and has no influence on the falling of the other blocks. In the non-exhaustive situations, participants learn in the training phase that the cons-block falls if the antecedent- or the yellow block falls or if both fall.

⁷80% of the participants responded correctly in their respectively last trial of the training phase, whereas only about 50% gave the correct answer in the first training trial.

more likely described by Bob.

2.1. RESULTS.

Data exclusion We excluded all data from participants who did not give the correct response in: (i) all three attention-check trials or in (ii) more than one of the control trials or in (iii) the example test trial in the end of the training phase. Additionally, we excluded the data from two participants whose comments in the end of the experiment indicated that they did not do the experiment properly. Overall, the data of 282 participants remained to be included in the analysis.

Behavioral Data Figure 2 shows the proportion of participants who selected the exhaustive situation as the situation that is more likely described by Bob's conditional answer to Ann's question. Participants' choices seem to depend strongly on the stimulus (color coded); while similar results are observed for stimuli B and D on the one hand and stimuli A and C on the other hand, there is a striking difference between the results for stimuli B and D as compared to the results for A and C. The difference between stimuli A/C and B/D lies in the second cause for the cons-block to fall in the non-exhaustive situation: for stimulus B and D, it is the potential falling of an additional block whereas for stimulus A and C, it is the position of the cons-block itself that may cause it to fall. In the former two stimuli, participants show a strong preference for the exhaustive situation and in the latter two stimuli, participants seem to prefer the non-exhaustive situation (selection rates consistently below 0.5).

The difference in responses observed between QUDs is much less striking. By eyeballing the data, we observe the expected tendency for stimulus A and C towards higher selection rates of the exhaustive situation when QUD=*will-q* as compared to QUD=*if-p*; for stimulus B and D, the same tendency is observed, but much less pronounced.

Statistical model We run a Bayesian logistic regression model, using the R-package brms (Bürkner 2017), with the QUD, the stimulus (pair of two situations) and their interaction as predictors. As random effects, we include varying intercepts and slopes per participant and use brms default priors for all parameters. Only for stimulus A, there is good reason to believe that the selection rate of the exhaustive situation will be larger when QUD=*will-q* as compared to QUD=*if-p*, reaching a posterior probability of approximately 96% ($P(\beta_{\text{qudif-p}} + \beta_{\text{stimulusC:qudif-p}} < \beta_{\text{qudwill-q}} + \beta_{\text{stimulusC:qudwill-q}}) = 0.959$, 95% CI: [-1.08, -0.03]). For the remaining three stimuli the posterior probabilities and 95% credible intervals for the respective comparison of parameters are 0.71 ([-1.01, 0.48], stimulus B), 0.833 ([-0.89, 0.23], stimulus C) and 0.712 ([-0.80, 0.38], stimulus D). The estimated posterior probability for the selection rate of the exhaustive picture to be larger when QUD = *will-q* compared to when QUD = *if-p* across all four stimuli amounts to 0.953.

2.2. DISCUSSION. We found supporting evidence for the postulated effect of the QUD only for stimulus A, where the posterior probability for the exhaustive situation being selected more often when QUD=*will-q* as compared to QUD=*if-p* is reasonable large. However, the results for stimulus B,C and D also show a tendency towards this effect.

These results may be related to the unexpectedly strong difference in participants' responses between the four stimuli. Especially salient is the systematic preference for the exhaustive situation in stimuli B and D (non-exhaustive: *external*) and for the non-exhaustive situation in stimuli A and C (non-exhaustive: *internal*). Put differently, the biconditional interpretation is overall not

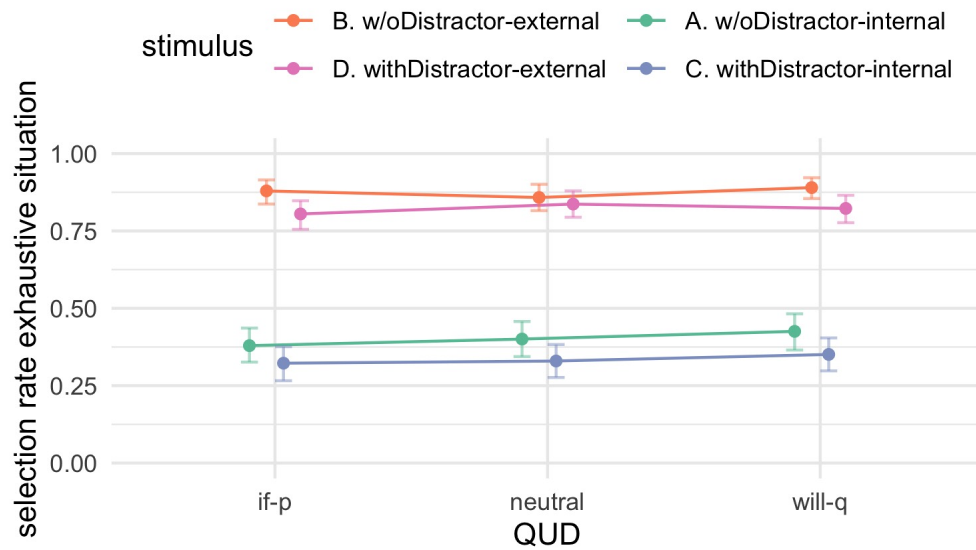


Figure 2: Relative frequency of the selection of the exhaustive situation in Experiment 1 as better described by the conditional “If the ant-block falls, the cons-block will fall”, for each QUD, color coded by stimulus. Errorbars are 95% bootstrap confidence intervals.

very prominent for stimulus A and C, but it is for stimulus B and D. One aspect that might have influenced this difference is the set of salient alternative utterances: for stimulus B and D (external cause), there is a salient alternative to describe the *dispreferred* non-exhaustive situation, namely “If the *yellow or* the ant-block falls, the cons-block will fall”, which may help explain the selection rates of the exhaustive situation in stimulus B and D close to ceiling. For stimulus A and C, for which the selection rate of the exhaustive situation is much lower throughout all QUDs, there is no similarly salient alternative for the non-exhaustive situation where the cons-block may fall without the influence of any other block.

Further, for stimulus B and C, the observed preferences (exhaustive for B, non-exhaustive for C) may be strengthened by the presence of the yellow distractor block in only one of the two shown situations; participants may have favored the situation *without* the yellow block — corresponding to the respectively preferred situations — just because it is not mentioned in the conditional. This is not per se problematic to test for an effect of the QUD, at least as long as the potential effect is not superposed completely by selection rates close to ceiling which we do observe for stimulus B.

Another possibility that may have influenced the results, in particular the preference for the non-exhaustive situation in stimuli A and C, is an epistemic instead of a purely causal interpretation of the conditional in these cases.⁸ Assuming an epistemic interpretation, the conditional is particularly true in the non-exhaustive situation — which is overall preferred in stimuli A and C.

In order to circumvent the possibility that a putative effect is not found due to participants’

⁸Epistemic interpretation in the sense that if the “nature of” the ant-block is such that it falls, the cons-block should fall as well since it is of the same “nature” as the ant-block: both blocks are identically centered on the edge of their platforms.

strong tendency to prefer either the exhaustive or the non-exhaustive situation depending on the concrete scenes, we conducted a follow-up experiment where we do not force participants to choose one among two situations but allow them to select both. When QUD=*if-p*, the conditional $p \rightarrow q$, should in fact be accepted as description for both situations since other possible reasons for the consequent are simply expected to be irrelevant and thus, the conditional is an appropriate answer in the exhaustive as well as in the non-exhaustive situation. Indeed, several participants in Experiment 1 mentioned in their comments in the end of the study that for *some* trials, both pictures were possible.

3. Experiment 2. The code and the preregistration report for Experiment 2 are available on OSF.⁹

Participants 315 participants were recruited via the online Platform Prolific, using the same eligibility criteria as for Experiment 1.¹⁰ The cleaned data (see below) comprises data from 181 participants (76 male, 103 female, 2 other) with a mean age of 37.12 years (range 18 – 68). For their participation, each participant received £1.67; on average they finished the experiment in approximately 14 minutes (range 4.5 – 46).

Setup & Materials The training phase consisted of 8 trials and was followed by the test phase consisting of 17 trials split into 3 blocks, a practice block with 4 trials followed by two test blocks with 7 and 6 trials respectively. The trials of the two test blocks alternated between filler and critical trials and included an attention check trial after the first test block. The order of trials within filler, critical and practice trials was randomized for each participant. After each block, participants had the possibility to take a break before proceeding with the next block. In the end, we further asked participants to answer a set of questions about the experiment to (i) verify that they did not ignore Ann's question and (ii) to get an idea of how certain participants had to be such that they would select only one scene.

Procedure The most important difference in the procedure of Experiment 2 compared to Experiment 1 is that in Experiment 2, participants were not forced to select a single picture. They read the same dialog as in Experiment 1 but were shown three instead of two scenes to choose from. The third picture is referred to as the *control scene* since it shows a situation that contradicts Bob's conditional answer and should, thus, never be selected.

By telling participants that Ann sees part of the scene that Bob describes, Ann's question was meant to be more purposeful: she seeks to get more information about a scene that she only has partial access to, see Figure 3 for an example trial. While the partial scene was immediately visible in each trial, Ann's question, Bob's response and the three scenes had to be revealed one by one. With this setup we hoped to enforce participants to process both, the QUD and the conditional, before making a selection.

Further, Experiment 2 was built up as a game where participants can earn points, with the aim to incentivize that participants do not always select a single situation (or always both): when selecting two scenes, they get 50 points if the correct one is among them, otherwise, they loose

⁹<https://tinyurl.com/pphck52h>, <https://tinyurl.com/pphck52h>

¹⁰We increased the number of recorded participants by 100 after having cleaned the data of the originally planned 215 participants since we had to exclude many more participants than expected who did not fulfill our predefined criteria. All data was analyzed only after all 315 participants were recorded.

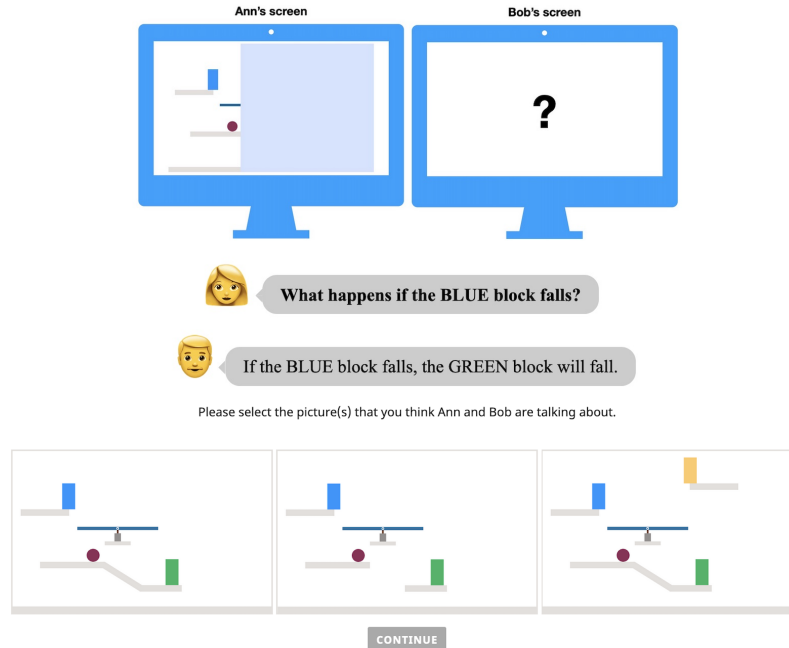


Figure 3: Critical test trial from Experiment 2 where QUD=*if-p*, non-exhaustive (right) *external* and exhaustive (left) *w/o distractor*. The picture in the middle is the control scene.

100 points. Selecting only the correct scene is awarded with 100 points, but when a single scene is selected that is *not* the correct one, participants lose 100 points. Therefore, in the long run, participants are better off to select two scenes when they are undecided.

Manipulations As in Experiment 1, we manipulate the QUD encoded in Ann’s question, but without using the *neutral* condition (“Which blocks do you think will fall?”) in the critical test trials. While the four critical stimuli (pairs of an exhaustive and a non-exhaustive situation) are the same as in Experiment 1, there are only 6 critical trials in Experiment 2: since Ann’s question, the QUD, relates to the block that is shown in the partial scene (when QUD=*if-p*, the partial scene shows the ant-block, when QUD=*will-q*, it shows the cons-block), it should be at the same position as the respective block in the three pictures of scenes among which participants make their selection; only then none of the three situations can be excluded just because it does not match the part of the scene that Ann sees. Thus, stimuli A / C are not combined with QUD *will-q*.

Training phase The animations in the training phase were the same as in Experiment 1 plus one additional trial which showed the situation that contradicts the conditional “If the ant-block falls, the cons-block will fall” (see Figure 3, middle), which is the control scene in the critical conditions of the test phase.

Test phase The test phase consists of three blocks, a practice and two test blocks. The practice block consists of 4 trials in which participants got feedback about the correct picture and the number of points they received with their selection. The main purpose of the practice trials was to demonstrate that Bob’s responses are informative and to make participants learn how their choices impact the amount of points they get. The procedure in the trials of the two test blocks was the

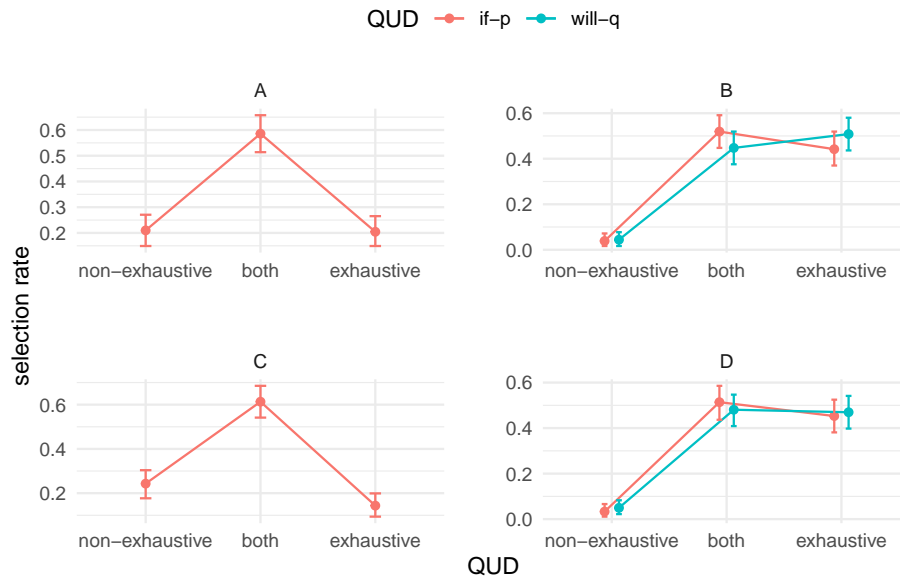


Figure 4: Relative frequency of the scene(s) participants selected in Experiment 2 based on the conditional “If the ant-block falls, the cons-block will fall”; color code represents the two QUDs, errorbars are 95% bootstrapped confidence intervals.

same as in the practice block, except that participants did not receive feedback anymore. In order to keep the character of the game up without influencing participants’ choices, they were told that they would get their final score in the end of the experiment. The filler trials in the test blocks were designed such that there is a similar number of trials where QUD=*if-p* (6) and QUD=*will-q* (5).

3.1. RESULTS.

Data exclusion We excluded all data from participants who fulfilled at least one of the following criteria: (i) they did not select the correct scene in the attention-check trial, (ii) they selected the control scene at least once in the test phase (excluding the practice block), (iii) they affirmed either that they only read Bob’s answer, but not Ann’s question, or that Ann’s question was always the same or (iv) they responded within less than 6 seconds in at least 2 of the critical trials.

Behavioral Data Figure 4 shows participants’ average responses in Experiment 2 for each of the four critical stimuli. Similarly to Experiment 1, we observe a preference away from an exhaustive interpretation in stimuli A and C, where selecting both situations is much more likely than selecting only one situation; in stimulus C, selecting only the non-exhaustive situation is even more likely than selecting only the exhaustive situation. Contrary to that, stimuli B and D again show a preference towards an exhaustive interpretation: the selection rate of only the exhaustive situation is much higher in these stimuli than it is in stimuli A and C. Selecting both situations is, however, almost equally likely for stimuli B and D compared to A and C.

By eyeballing the results in Figure 4, the QUD *will-q* shows the predicted effect for stimulus B: we observe an increase in the selection of the exhaustive situation when QUD=*will-q* compared to QUD=*if-p* and a decrease in the selection of both situations. For stimulus D, selecting both

situations is more likely than selecting only the exhaustive situation when QUD=*if-p*, but we do not observe the predicted effect of the QUD on the selection rate of the exhaustive situation, which is in fact on average more likely when QUD = *will-q* but the increase seems to be marginal.

Statistical model We run an ordinal regression model with brms (Bürkner & Vuorre 2019, Bürkner 2017) with the QUD, the non-exhaustive and exhaustive situation and the interaction between the QUD and the exhaustive situation and between the exhaustive and the non-exhaustive situation as predictors, including by-participants random intercepts and slopes for each predictor except for interactions. We chose an ordinal regression model as the three response categories reflect the degree of how exhaustive the conditional is interpreted: the selection of only the non-exhaustive situation corresponds to a maximally non-exhaustive interpretation and the selection of only the exhaustive situation to a maximally exhaustive interpretation, selecting both situations corresponds to an interpretation in between both extremes.

For stimulus B, the posterior probability that participants interpret the conditional more exhaustively when QUD=*will-q* as compared to QUD=*if-p* amounts to 93% ($P(\beta_{\text{qudwill-q}} + \beta_{\text{exhwoD}} + \beta_{\text{qudwill-q:exhwoD}} > \beta_{\text{exhwoD}}) = 0.93$, 95% CI: [-0.04, 0.58]), pointing towards the postulated effect of the QUD. As Figure 4 suggested, for stimulus D, the posterior probability is much lower ($P(\beta_{\text{qudwill-q}} > 0) = 0.56$, 95% CI: [-0.26, 0.32]) and does not provide evidence for our hypothesis.

We further speculated (a) that when QUD=*will-q*, the selection rate of only the exhaustive situation will be larger than the selection rate of both situations and (b) when QUD=*if-p*, the selection rate of both situations will be larger than of only the exhaustive situation. Our data only provides evidence for (b) in stimuli A and C (posterior probability for $P(\text{both} \mid \text{QUD}=\textit{if-p}) > P(\text{exhaustive} \mid \text{QUD}=\textit{if-p})$ is 1).

3.2. DISCUSSION. Only the data for stimulus B provides a reason to believe in a more exhaustive interpretation of the conditional when QUD=*will-q* as compared to QUD=*if-p*, for stimulus D, the QUD does not seem to have the same alleged effect.

When we look at the exact conditions in which participants responses differ between stimuli B and D, we find the strongest difference in the selection rate of the exhaustive situation for stimuli B and D when QUD=*will-q*, with a posterior probability for $P(E \mid \text{QUD} = \textit{will-q}, \text{stimulus} = B) > P(E \mid \text{QUD} = \textit{will-q}, \text{stimulus} = D)$ of 0.92. When QUD=*if-p*, the posterior probability that $P(E \mid \text{QUD} = \textit{if-p}, \text{stimulus} = B) > P(E \mid \text{QUD} = \textit{if-p}, \text{stimulus} = D)$ is 0.52. Assuming that potential alternative causes are irrelevant for the choice that participants make when QUD = *if-p*, it seems reasonable that the observed difference between participants' responses in stimuli B and D can mainly be ascribed to the condition where QUD=*will-q* since when QUD=*if-p*, the focus of the conversation lies on the consequences of the antecedent and so, participants are not expected to consider other potential causes for the consequent. Further, it may have been the case that the presence of the distractor block in the exhaustive situation of stimulus D (that is absent in B) made participants more hesitant to decide for a single scene and thus they tend to choose both situations more often in this stimulus when alternative causes are considered, i.e., when QUD=*will-q*. Participants were indeed encouraged to select a single situation only when they were very confident, which 86% of participants confirmed in the questions in the end of the experiment.

Concerning the results for stimuli A and C, we observe a tendency towards a non-exhaustive

interpretation of the conditional (selection of the non-exhaustive situation or both), similarly to what we saw in Experiment 1. In fact, for stimulus C, the posterior probability of the probability to select only the non-exhaustive situation to be larger than the probability to select only the exhaustive situation amounts to almost 0.98 ($P(\text{non-exhaustive} \mid \text{QUD}=\text{if-}p, \text{stimulus} = C) > P(\text{exhaustive} \mid \text{QUD}=\text{if-}p, \text{stimulus} = C)$). The fact that in Experiment 2 we find that if participants choose a single scene in stimulus C, they select the non-exhaustive situation significantly more often than the exhaustive situation — which is not the case for stimulus A — might help to explain why, in Experiment 1, we did not find the predicted QUD-effect for stimulus C which we found for stimulus A; a general tendency towards the non-exhaustive situation may have interfered with a putative QUD-effect which may thereby become harder to find, especially when we assume that this preference would only become stronger when the alternative causes are assumed to be particularly considered, that is, when $\text{QUD}=\text{will-}q$.

4. Conclusion. Overall, we find some evidence supporting the hypothesis that a QUD that focuses on the conditions bringing about the consequent yields a more exhaustive interpretation of an indicative conditional than does a QUD that focuses on the consequences of the antecedent. Our results are far from being conclusive, yet they show that the interpretation of conditionals as biconditionals is likely to be the result of an interplay of various factors.

Especially Experiment 1 showed that when forced to choose either an exhaustive or a non-exhaustive situation, participants showed substantially different preferences depending on the nature of the second conceivable cause for the cons-block to fall in the non-exhaustive situation. It either fell because of a third block (stimuli B/D) or because of its own position on the edge of a platform (stimuli A/C). Stimuli B/D yield a strong preference of the exhaustive-situation whereas in stimuli A/C we observed a (less strong) preference of the non-exhaustive situation — across all QUDs including a neutral question. The second cause for the cons-block to fall as it is realized in stimuli A and C, namely because of its own position on the edge, comes along with another possible interpretation of the conditional that does not apply to stimuli B and D: the conditional may receive an epistemic instead of a causal interpretation. Consider the following conditional as an example for a conditional describing a situation similar to those in stimuli A and C, receiving an epistemic interpretation: “If that guy solved the puzzle, she will solve it [too / all the more]”. It does not seem to suggest that ‘if that guy does not solve the puzzle, she will not solve it’, rather, it suggests that ‘she might solve it, while he might not, but if he does, she will as well’. And this seems to be the case even if the conditional is an answer to the question “Will she solve the puzzle?”. In other words, under an epistemic interpretation of the conditional we should not expect to see a difference between the QUDs whereas we do expect a difference when the conditional receives a non-epistemic, causal interpretation. Therefore, disentangling a causal versus an epistemic interpretation of the conditional may help to get a cleaner picture of what is going on here.

Further, the observed tight connection between conditionals and causality generally suggests that it may be worth to look at the production of conditionals in comparison to the use of causal language (e.g., “X may make Y fall” or “Y may fall because of X”) to learn more about how participants use (and interpret) conditionals.

References

- Atlas, Jay David & Stephen C. Levinson. 1981. It-clefts, informativeness and logical form: Radical pragmatics (revised standard version). In *Radical pragmatics*, 1–62. Academic Press.
- Bürkner, Paul-Christian. 2017. Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1). [10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01).
- Bürkner, Paul-Christian & Matti Vuorre. 2019. Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science* 2(1). 77–101. [10.1177/2515245918823199](https://doi.org/10.1177/2515245918823199).
- Cariani, Fabrizio & Lance J Rips. 2016. Experimenting with (Conditional) Perfection: Tests of the Exhaustivity Theory .
- Cariani, Fabrizio & Lance J Rips. forthcoming. Experimenting with (Conditional) Perfection: Tests of the Exhaustivity Theory. In Stefan Kaufmann, Over David & Ghanshyam Sharma (eds.), *Conditionals: Logic, linguistics, and psychology*, Palgrave.
- Cummins, Denise D., Todd Lubart, Olaf Alksnis & Robert Rist. 1991. Conditional reasoning and causation. *Memory & Cognition* 19(3). 274–282. [10.3758/BF03211151](https://doi.org/10.3758/BF03211151).
- de Cornulier, Benoit. 1983. ‘If’ and the Presumption of Exhaustivity. *Journal of Pragmatics* 7(3). 247–249.
- Ducrot, Oswald. 1969. Présupposés et sous-entendus. *Langue française* [10.3406/lfr.1969.5456](https://doi.org/10.3406/lfr.1969.5456).
- Evans, Jonathan St BC, Jonathan St BT Evans, Stephen E. Newstead & Ruth MJ Byrne. 1993. *Human reasoning: The psychology of deduction*. Psychology Press.
- Farr, Marie-Christine. 2011. Focus influences the presence of conditional perfection: Experimental evidence. *Proceedings of Sinn & Bedeutung* 15. 225–239.
- Fillenbaum, Samuel. 1986. The use of conditionals in inducements and deterrents. In Alice Ter Meulen, Charles A. Ferguson, Elizabeth Closs Traugott & Judy Snitzer Reilly (eds.), *On Conditionals*, 179–196. Cambridge: Cambridge University Press. [10.1017/CBO9780511753466.010](https://doi.org/10.1017/CBO9780511753466.010).
- Geis, Michael L. & Arnold M. Zwicky. 1971. On invited inferences. *Linguistic inquiry* 2(4). 561–566.
- Horn, Laurence R. 2000. From if to iff: Conditional perfection as pragmatic strengthening. *Journal of pragmatics* 32(2000). 289–326.
- Lilje, Gerald W. 1972. Uninvited inferences. *Linguistic Inquiry* 540–542.
- López Astorga, Miguel. 2014. ¿Podemos evitar la perfección del condicional enfocando el antecedente o son necesarios antecedentes alternativos? *Revista signos* 47(85). 267–292.
- Markovits, Henry. 1986. Familiarity effects in conditional reasoning. *Journal of Educational Psychology* 78(6). 492.
- Van Canegem-Ardijns, Ingrid & William Van Belle. 2008. Conditionals and types of conditional perfection. *Journal of Pragmatics* 40(2). 349–376. [10.1016/j.pragma.2006.11.007](https://doi.org/10.1016/j.pragma.2006.11.007).
- Van der Auwera, Johan. 1997. Pragmatics in the last quarter century: The case of conditional perfection. *Journal of Pragmatics* 27(3). 261–274.
- Von Fintel, Kai. 2001. Conditional strengthening. *Unpublished manuscript* .