Conditional inference and access to alternatives: The role of QUD and speaker knowledge in conditional perfection

Ebru Evcen[1], Alan Bale[2], and David Barner[3]

[1]Department of Linguistics, UC San Diego

[2] Linguistics Program, Concordia University

[3]Department of Psychology, UC San Diego

Author note. This manuscript has been accepted for publication in Journal of Semantics. This version is a preprint and may differ slightly from the final published version.

Contact:

Ebru Evcen

Department of Linguistics, UC San Diego

9500 Gilman Drive, La Jolla, CA, 92093

eevcen@ucsd.edu

Abstract

Conditional statements often imply meanings that extend beyond their literal content, influenced by contextual factors such as the Question Under Discussion (QUD) and a speaker's epistemic state. One such inference is Conditional Perfection (CP), wherein a statement like, "If *p* then *q*," is strengthened to imply that the antecedent *p* is the only contextually salient condition under which the consequent *q* holds (e.g., "*p* if and only if *q*"). While previous research has debated the role of QUDs in CP, the role of speaker knowledge has yet to be systematically tested. In three experiments, we investigated how these contextual factors influence conditional reasoning. In Experiment 1, participants were more likely to derive CP when the conditional statement was an answer to a QUD that focused on the antecedent compared to those that were answers to either a neutral QUD or one that focused on the consequent. Experiment 2 tested which answers count as relevant to an explicit QUD, and found that overly informative answers were also treated as relevant alternatives for exhaustification. In Experiment 3, CP was more likely when the speaker was knowledgeable about all alternative antecedents than when they had partial knowledge. These findings support the view that CP arises through a process of exhaustification, wherein the conditional statement is strengthened by negating relevant alternatives. More broadly, they suggest that conditional inferences depend on reasoning about alternative utterances, paralleling other forms of pragmatic enrichment.

Keywords: conditional perfection, implicature, question under discussion, speaker knowledge, pragmatics, alternatives, conditional reasoning

Conditional statements often imply meanings that go beyond their literal content, influenced by context and the alternative utterances that listeners consider relevant. For example, in a context where tasks like mowing the lawn, doing the dishes, or cleaning the pool are each associated with a reward, a speaker might say, "If Mary mows the lawn, she will receive \$5" implying that mowing the lawn is the only way for Mary to earn \$5. Thus, $p \,\square\!\!\rightarrow q$ implies that other salient antecedents r do not license q . This reasoning, where a conditional statement is understood exhaustively[1] to mean "only if", is often referred to as "Conditional Perfection" (CP, Geis & Zwicky, 1971). Schematically, $p \,\square\!\!\rightarrow q \Rightarrow r \,\square\!\!\rightarrow \neg q$, for any salient alternative antecedent r (where $\square\!\!\rightarrow$ is shorthand for modal implication, without commitment to a specific semantic framework). For concreteness, we follow Stalnaker's (1968) analysis, in which $p \,\square\!\!\rightarrow q$ is true in a world w if and only if the closest p-world to w is a q-world. Under this interpretation, $\neg(r \,\square\!\!\rightarrow q)$ entails $r \,\square\!\!\rightarrow \neg q$.[2] The empirical arguments in this paper, however, do not hinge on this choice: the same reasoning applies under a Lewis-style account (Lewis, 1973) with some innocent auxiliary assumptions.[3]

Most prior analyses treat conditional perfection as a type of strengthening (Cornulier, 1983; Geis & Zwicky, 1971; Horn, 2000; Levinson, 2000; van der Auwera, 1997; von Fintel,

---

[1] We use "exhaustification" in a theory-neutral sense to mean *strengthening*. Our use of the term does not presuppose a grammatical implementation of exhaustivity (e.g., via an *Exh* operator), but is compatible with both grammatical and Gricean accounts.

[2] In Stalnaker (1968), a modal statement such as $p \,\square\!\!\rightarrow q$ ('p > q' in Stalnaker's preferred notation) is true in a world w iff q is true in the closest accessible p-world from w, where *closeness* is measured via a strict ordering of worlds relative to the appropriate accessibility relation (e.g., deontic, epistemic, etc.). If we assume that for any world w and any proposition q, either q is true in w or false in w, then for any proposition r, $\neg(r \,\square\!\!\rightarrow q)$ entails $r \,\square\!\!\rightarrow \neg q$.

[3] In a Lewis-inspired analysis (Lewis, 1973), the implication from $\neg(r \,\square\!\!\rightarrow q)$ to $r \,\square\!\!\rightarrow \neg q$ requires additional assumptions, such as homogeneity (e.g., give a set W of accessible close worlds to the world of evaluation, either every w in W is a q-world, or every w in W is a ¬q-world). See the discussion in von Fintel (2001).

2001), often grounded in Gricean theories of pragmatic reasoning (e.g., Horn, 1972; Grice, 1975; Gazdar, 1979; Geurts, 2009) or similarly inspired grammatical algorithms (e.g. Chierchia, 2004; Fox, 2007; Chierchia et al., 2012). These accounts, while differing in their details, converge on the idea that CP arises when the condition stated in the antecedent *p* is understood to exhaust the set of conditions sufficient for the consequent *q* (von Fintel, 2001). For example, in a context where it is known that Peter, Paul, and Mary are mowing the lawn, the statement in (1a) often communicates not only that Mary mowed the lawn but that Peter and Paul did not. These implications can be analyzed as a type of quantity implicature, where asserting (1) implies the falsity of the contextually relevant *ad hoc* alternatives, such as those in (1b) and (1c).[4]

(1) a. Mary mowed the lawn.

b. Peter mowed the lawn.

c. Paul mowed the lawn.

According to von Fintel (2001), perfection can be derived in a similar way, where the assertion in (2a) implies the falsity of the *ad hoc* alternatives in (2b) and (2c).[5]

(2) a. If Mary mows the lawn, she will receive $5.

b. If Mary does the dishes, she will receive $5.

c. If Mary cleans the pool, she will receive $5.

Such inferences are generally thought to be sensitive to contextual factors that limit or strengthen the salience or relevance of alternatives (see Hirschberg, 1985; Matsumoto, 1995;

---

[4] The implication can be derived either by negating the non-weaker alternative in (2) or the following stronger alternative: i) Paul and Mary mowed the lawn, ii) Peter and Mary mowed the lawn, iii) Peter, Paul and Mary mowed the lawn (see Fox, 2007).

[5] As with the Peter, Paul, and Mary example, the same implications could also be derived from the following stronger alternatives: i) If you do the dishes or you mow the lawn, you will receive $5, ii) If you clean the pool or you mow the lawn, you will receive $5, iii) If you do the dishes or you clean the pool or you mow the lawn, you will receive $5.

Groenendijk & Stokhof, 1984; von Fintel, 2001, *inter alia*). For example, whether or not the alternatives in (1b) and (1c) are negated depends on the Question Under Discussion (QUD; Roberts, 1996/2012). If (1a) were uttered as an answer to the question in (3a), it would strongly imply that neither Peter nor Paul left. In contrast, as an answer to either question in (3b), (1) provides no information about the status of Peter or Paul. Critically, the sentences in (1b) and (1c) are possible answers to the question in (3a) but not to those in (3b) or (3c). Exhaustive interpretations seem to only exclude alternatives that are possible answers to the QUD.

(3) a. Who mowed the lawn?

b. What did Mary do after doing the dishes?

c. Did Mary mow the lawn?

What counts as an "answer" to a QUD, however, differs across theories. Some treat overly informative answers, those that provide more information than needed—i.e., a proper subset of a QUD cell, as relevant for exhaustification, while others restrict exhaustification to propositions that correspond exactly to a QUD cell (or a union of such cells; Benz & van Rooij, 2006; Benz, 2006; Groenendijk & Stokhof, 1985; Grice, 1975; Matsumoto, 1995; Lewis, 1979; Roberts, 1996/2012). To illustrate, consider a slightly modified context. Suppose the explicit QUD is "Who worked outside?", with the possible answers *Peter worked outside*, *Mary worked outside*, and *no one worked outside*. Suppose that "Mary mowed the lawn" was uttered as a response. This sentence answers the question (e.g., it tells you that Mary worked outside) but it provides more information than requested. It is therefore overly informative relative to the QUD. Whether listeners take this to imply that Mary did not do other outside chores (e.g., weeding the gardening) depends on whether such overly informative answers are treated as relevant in the discourse.

The QUD, then, determines which propositions qualify as possible answers and thus as relevant alternatives for exhaustification. Whether a given answer is expected to exclude other possibilities depends not only on whether an answer corresponds to a QUD cell, but also on how the speaker's knowledge constrains which alternatives are available for negation. On most accounts, the speaker's epistemic state plays a similar role, since exhaustification should only exclude alternatives whose status is known to the speaker. Under a Gricean-inspired analysis, for example, to derive a quantity implicature $\neg q$ from a statement expressing *p*, the listener must first identify *q* as a relevant alternative to *p* (Matsumoto 1995), which is determined by the QUD. Given that *q* is a relevant alternative, the listener can reason, via Quantity, that if the speaker knew *q* were true, they would have uttered a statement explicitly expressing q (or alternatively, a statement expressing $p \wedge q$). Since the speaker did not do so, the listener can infer that the speaker does not know that q is true (i.e., $\neg K_S(q)$). Critically, this inference does not entail that *q* is false since the speaker might simply be ignorant about the status of *q* (i.e., $\neg K_S(q) \wedge \neg K_S(\neg q)$). Thus, to conclude $\neg q$ from $\neg K_S(q)$, the listener must believe that the speaker is knowledgeable about q (i.e., $K_S(q) \vee K_S(\neg q)$) and that they would have uttered it if it were true.[6,7] Once this is established, the listener can infer that the speaker knows that *q* is false.

This analysis predicts that conditional perfection should be sensitive to contextual manipulations of the QUD or the speaker's epistemic states. For example, a conditional

---

[6] Thus, $K_S(q) \vee K_S(\neg q)) \wedge \neg K_S(q) \vDash K_S(\neg q)$, which in turn entails $\neg q$.

[7] We have framed this discussion from a Gricean perspective, though, a similar account can be given on a grammatical theory of implicature (e.g., Fox 2007, Chierchia et al. 2012). Under such an account, an exhaustification operator negates all non-weaker alternatives (e.g., $Exh(p) = p \wedge \forall q \in ALT(p).(p \nvDash q \rightarrow \neg q)$). However, ALT(p) can be redefined to ensure that it is a subset of both ANS_{QUD} and $\{q: K_S(q) \vee K_S(\neg q)\}$ (i.e., $ALT(p) \subseteq ANS_{QUD} \wedge ALT(p) \subseteq \{q: K_S(q) \vee K_S(\neg q)\}$). Under such a redefinition, the Exh operator would only negate alternatives of p that answer the QUD and whose status is known by the speaker.

statement like, "If Mary mows the lawn, she will receive $5" ($p \,\Box\!\!\rightarrow q$) should imply that Mary won't receive $5 if she washes the dishes ($\neg(r \,\Box\!\!\rightarrow q)$) only if $r \,\Box\!\!\rightarrow q$ is a possible answer to the QUD (e.g., What activities will give Mary $5?). Similarly, the inference that Mary won't receive $5 for washing the dishes hinges on the assumption that the speaker is knowledgeable about this alternative. If the speaker isn't sure whether or not washing the dishes will lead to $5, then the listener should not perfect the conditional.[8] Interestingly, while these predictions regarding CP are relatively clear, there is relatively little experimental evidence documenting such inferences, and how contextually defined alternatives impact conditional reasoning. Those studies that have explored this question have focused on testing the effects of the QUD on perfection, with some finding strong evidence that QUD impacts perfection (e.g., Farr, 2011), and others finding little to no evidence of such effects (e.g., Cariani & Rips, 2023; Grusdt et al., 2023). Meanwhile, the role of the speaker's epistemic state in the interpretation of conditionals has yet to be tested empirically.

In one study that investigated the role of QUD in perfection, Farr (2011) presented participants with a scenario in which one character, Kerstin, asked another character, Sahra, about seafood prices. The scenario presented two ways to earn 2.50 euros: selling an eel or selling a pike. Depending on the condition, Kerstin posed either a consequent-focused question[9],

---

[8] Only by assuming $K_S(r \,\Box\!\!\rightarrow q) \vee K_S(\neg(r \,\Box\!\!\rightarrow q))$, can the listener conclude from $\neg K_S(r \,\Box\!\!\rightarrow q)$ that $K_S(\neg(r \,\Box\!\!\rightarrow q))$ and hence $\neg(r \,\Box\!\!\rightarrow q)$.

[9] Different labels for these QUDs appear in the literature. Farr (2011) distinguishes between "what-if-p?" questions, which focus on variation in the consequent (referred to here as consequent-focused), and "when-q?" questions, which focus on variation in the antecedent (referred to here as antecedent-focused). Cariani & Rips (2023) introduce the terms ANT? (antecedent-directed QUDs) and CONS? (consequent-directed QUDs), depending on whether variation is constrained to the antecedent or consequent, respectively. Grusdt et al. (2023) use "if-p?" for consequent-focused QUDs and "will-q?" for antecedent-focused QUDs, also including a "neutral" category.

"What happens if I sell an eel?" or an antecedent-focused question, "When do I get 2.50 euros?" and Sahra's response was always the same: "If you sell an eel, you get 2.50 euros." Farr found that participants judged this answer to be sufficient more often when it was in response to the consequent-focused question than when it followed the antecedent-focused question, compatible with an effect of the QUD. However, challenging this conclusion, Cariani and Rips (2023) argued that participants' responses may not have actually reflected conditional perfection. Instead, because there were two ways to earn 2.50 euros, but Sahra mentioned only one, participants' dissatisfaction may have arisen from her providing incomplete information rather than a false statement under a perfected interpretation.

To address this concern, Cariani and Rips (2023) conducted seven experiments in which participants first read a scenario (e.g., John has taken a test on Chapters 4–6 that has not been graded yet) before being presented with one of two types of questions: an antecedent-focused question (e.g., You ask Mary, "What are all the ways John could manage to do well on the test?"), or a consequent-focused question (e.g., You ask Mary, "What are all the things that could happen in case John understood Chapter 5?"). Following this, participants were told that Mary responded with a conditional statement (e.g., Mary says, "If John understood Chapter 5, then John did well on the test"). Participants were then told that the consequent was true (e.g., John did well on the test) before being asked, given this information and what Mary said, whether the antecedent was also true (i.e., John understood Chapter 5). Cariani and Rips found that participants' responses did not differ depending on whether the question focused on the consequent or the antecedent, and thus concluded that explicitly requesting exhaustive answers is not always sufficient to induce perfection. However, they also acknowledged that participants may have interpreted the antecedent-focused questions, which were expected to induce conditional perfection, as open-ended and difficult to interpret exhaustively, given how easy it is to think of alternative ways the outcome could be achieved (e.g., cheating or lucky guessing).

In another study, Grusdt et al. (2023) tested the effects of QUD on conditional perfection in contexts designed to minimize prior beliefs and world knowledge. Participants were shown scenes depicting simple physical interactions, such as an upper block colliding with a ball and causing a lower block to fall. In each trial, one character asked a question focused either on the consequent ("What happens if the upper block falls?") or the antecedent ("Will the lower block fall?"), and another character, Bob, responded with a conditional statement like, "If the upper block falls, the lower block will fall." Participants were then asked to choose a scene that best matched Bob's description, selecting between either an exhaustive scenario (where only the upper block's fall causes the lower block to fall) or a non-exhaustive scenario (where other causes could also lead it to fall). While a modest QUD effect was observed, with a slight preference for exhaustive interpretations when the QUD focused on the antecedent, this effect was inconsistent.

While some previous studies have investigated the role of the QUD in perfection, albeit with inconclusive results, no previous studies have explored the role of speaker knowledge. This is important because, in other domains, speaker knowledge appears to play a strong role in modulating the availability of conversational implicatures (e.g., Bergen & Grodner, 2012; Breheny et al., 2013; Hochstein et al., 2018; Goodman & Stuhlmüller, 2013; Bale et al., 2024). For example, in one study, Bergen and Grodner (2012) investigated whether listeners consider the speaker's knowledge state when interpreting scalar implicatures, with *some*. To do this, they designed a self-paced reading experiment where participants read statements under different knowledge conditions. In one condition, the speaker was presented as having complete knowledge by stating, "At my client's request, I meticulously compiled the investment report." In contrast, the partial knowledge condition stated, "At my client's request, I skimmed the investment report" suggesting that the speaker might not know all the details. Following this, in both conditions, participants read a target sentence, such as, "Some of the real estate investments lost money," which was immediately followed by a continuation sentence, "The rest were successful despite the recent economic downturn." Bergen and Grodner reasoned that when the speaker was fully knowledgeable, *some* should trigger a strong implicature (*some, but not all*), making "*the rest…"* easier to process since this was already expected. However, if the speaker had only partial knowledge, *some* might be interpreted as *some, and possibly all*, causing "*the rest…"* to require extra processing and leading to slower reading times. Consistent with their predictions, they found that reading times for *the rest* were faster when the speaker had full knowledge, suggesting that the implicature had already been computed. In contrast, they found slower reading times in the partial-knowledge condition suggesting that no implicature was expected, making the final sentence more effortful to integrate.

In the present study, we report data from three experiments that investigate the role of contextual alternatives in conditional perfection. In Experiment 1, we tested how participants interpret conditionals across three different types of QUDs: Antecedent-focused QUDs (i.e., QUDs whose answers permit variation in the antecedent but not the consequent), consequent-focused QUDs (i.e., QUDs whose answers permit variation in the consequent but not the antecedent), and neutral QUDs (i.e., QUDs whose answers permit variation in the consequent and antecedent). As in Grusdt et al. (2023), we tested participants with contexts that minimized the influence of prior beliefs and world knowledge. However, instead of using abstract symbols or causal interactions between arbitrary objects, we presented participants with simple back-andforth social interactions. Specifically, participants watched videos in which a character, Mary, pressed three buttons (red, blue, and orange), which each produced an animal sound audible only to her (because she was wearing headphones; see Figure 1). In each video, another character asked one of three questions: a consequent-focused question (e.g., "What will happen if I press the blue button?"), an antecedent-focused question (e.g., "Which button will play a dog sound?"), or a neutral question (e.g., "What will happen if I press the buttons?"). Mary always responded with the same conditional statement: "If you press the blue button, it will play a dog barking." The type of question determined the relevant set of alternatives: consequent-focused questions put the focus on different sounds (e.g., "it will play a cat meowing", "it will play a lion roaring"), while antecedent-focused questions targeted different buttons (e.g., "if you press the red button", "if you press the orange button").

For Experiment 1, we predicted that if perfection is sensitive to QUD, then exhaustive interpretations should arise under the antecedent-focused question because the alternatives to the antecedent must be considered. In particular, since the speaker explicitly mentioned the blue button, participants might infer that only the blue button plays the dog sound, as other buttons would have been mentioned otherwise. In contrast, for the consequent-focused questions, where the alternative utterances contain variations in the consequent, this exhaustive inference should not arise if perfection is sensitive to the QUD. Meanwhile, neutral questions were inherently broad and did not specify which button was being pressed (i.e., blue button) or what kind of outcome was sought (i.e., dog sound), thereby failing to constrain the set of possible antecedents and consequents. As a result, a strengthened meaning would lead to contextual inconsistencies, making an exhaustive interpretation unlikely.

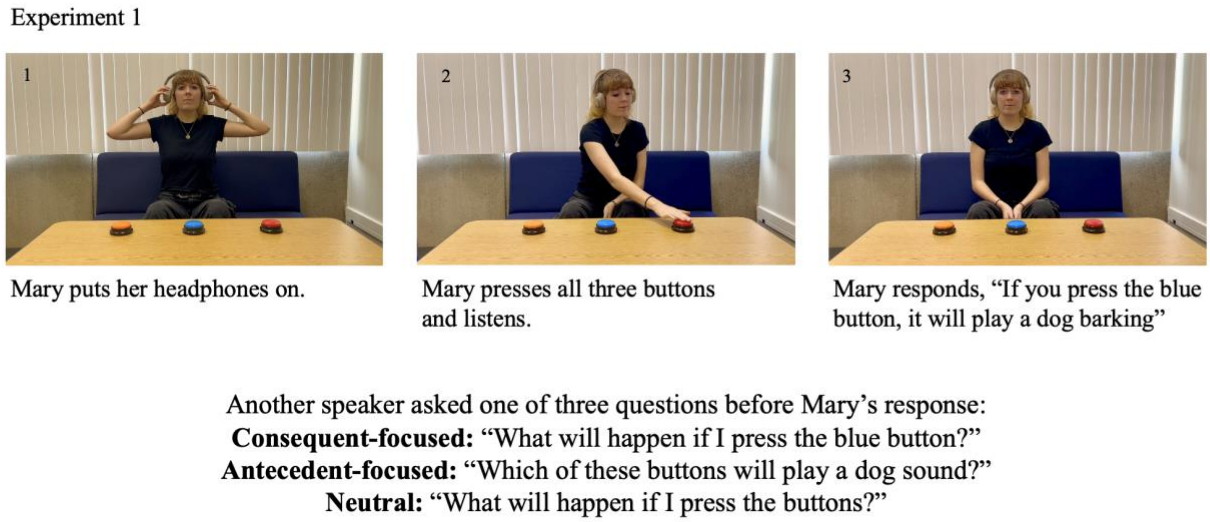

Figure 1. Still images from video trials in Experiment 1, depicting Mary putting on headphones, listening to the buttons, and responding with a conditional statement

In Experiment 2, we further explored which answers count as relevant to an explicit QUD – whether overly informative answers are treated as viable alternatives and get perfected, or whether exhaustification excludes such answers (Groenendijk & Stokhof, 1985; Grice, 1975; Matsumoto, 1995; Benz et al., 2006; Benz, 2006; Lewis, 1979). To test this, we presented participants with short videos, similar to Experiment 1, showing a speaker interacting with two shapes, triangles and squares, each appearing in two distinct colors. In each trial, another speaker asked an explicit question, “Which of these shapes, triangles or squares, will play a dog barking?” Mary then responded with a conditional statement that was either an “optimally informative” answer, such as, “If you press the triangles, it plays the sound of a dog barking,” or an “overly informative answer” such as, “If you press the blue square, it plays the sound of a dog barking.” We reasoned that if overly informative answers are treated as relevant, then Mary’s utterance should be interpreted relative to all specific alternatives, including other shapes and colors. For example, hearing, “If you press the blue square, it plays the sound of a dog barking” should lead listeners to consider alternative antecedents such as, “If you press the red square,” “If you press the yellow triangle,” and “If you press the green triangle.” Exhaustifying the statement under this interpretation should generate the inference that only the mentioned button (the blue square) produces the sound, and conditional perfection should be computed relative to these overly informative alternatives. However, if overly informative answers are not treated as relevant, then exhaustification should be restricted to alternatives that correspond to the cells of the QUD itself, such as, “If you press the triangles, it will play a dog barking” or “If you press the squares, it will play a dog barking.” In this case, the conditional statement “If you press the squares, it will play a dog barking” is a weaker alternative than “If you press the blue square, it will play a dog barking” and thus is not a viable candidate for negation. However, the alternative “If you press the triangles, it will play a dog barking” is not weaker and therefore is a viable candidate for negation. Negating this alternative implies that none of the triangles play the sound of a dog barking, but it leaves open the possibility that pressing the other square buttons does. Under this view, CP should only be computed relative to alternatives that correspond to a QUD cell, and thus there should not be any implication with respect to the other squares buttons.

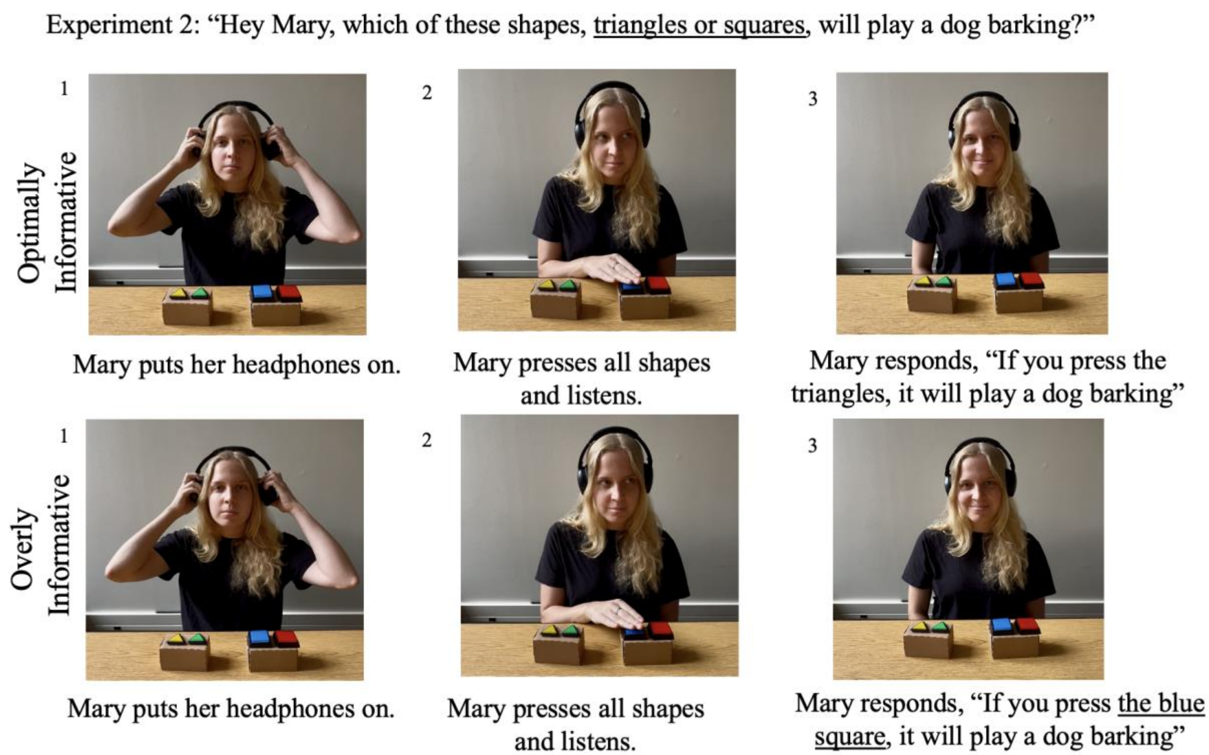

Figure 2. Still images from video trials in Experiment 2, depicting Mary putting on headphones, listening to the shapes, and responding with a conditional statement, which is either an Optimally Informative Answer given the explicit QUD (on the top) or an Overly Informative Answer (on the bottom)

Finally, in Experiment 3, using an adaptation of the paradigm used in Experiments 1 and 2, we tested how a speaker's epistemic state modulates perfection for antecedent-focused questions (e.g., "Which of these buttons will play a dog sound?"). Participants viewed trials where the speaker, Mary, pressed and listened to either two buttons (partial knowledge trials) or all three buttons (full knowledge trials) before making a conditional statement about the sounds the buttons produced, such as, "If you press the blue button, it will play a dog barking," similar to Experiment 1 (see Figure 2). We reasoned that if conditional inferences are sensitive to the speaker's knowledge state, participants should be less likely to interpret perfect conditionals when the speaker has partial knowledge than when they have full knowledge.

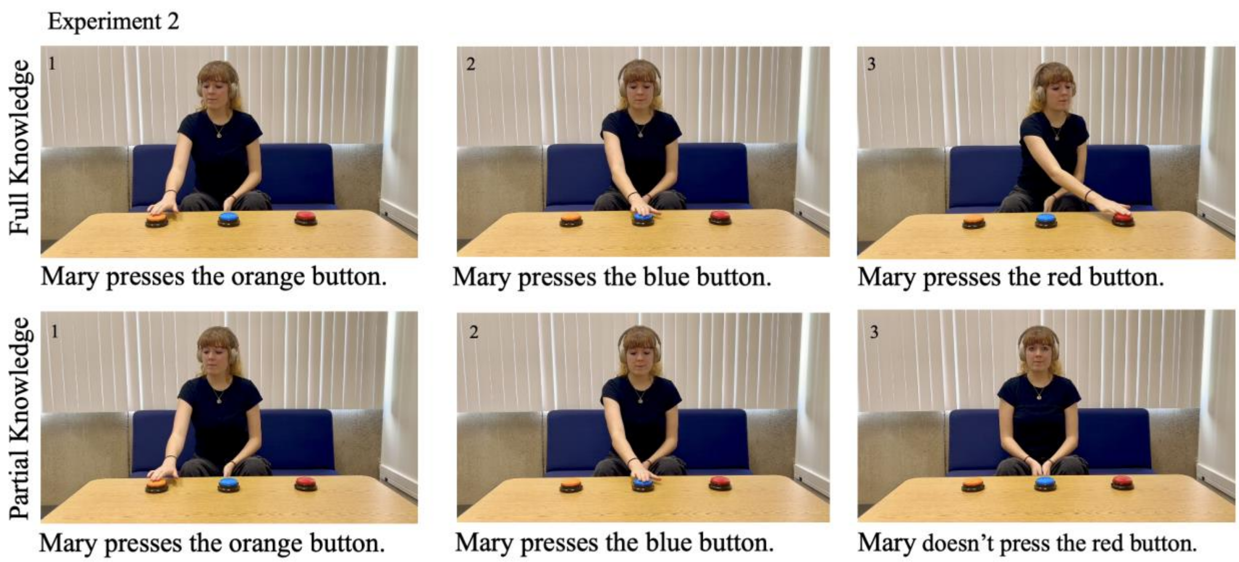

Figure 3. Still images from video trials in Experiment 2, showing Mary pressing all three buttons in the Full Knowledge condition and only two buttons in the Partial Knowledge condition.

Experiment 1

Methods

All methods and analyses were preregistered prior to data collection. The preregistration is available at https://osf.io/9tcmw/?view_only=454afc1ea72045cc80cbbd0ef4085ca1. All data and analysis code for this are available at https://osf.io/vq72b/?view_only=48081f470c6c4fc589651c85e5bfebcf.

Participants

The experiment was hosted on PCIbex (Zehr & Schwarz, 2018). We recruited 104 native English speakers (N=98 post exclusions) with an IP address from the United States via the online platform Prolific. Participants received payment for their participation equivalent to California minimum wage at the time ($15.50/hour). As per our pre-registered exclusion criteria, we removed one participant due to technical problems and five participants failing to respond to attention checks.

Materials & Procedure

Participants were randomly assigned to one of three between-subjects conditions: antecedent-focused QUD, consequent-focused QUD, or neutral QUD. Each trial consisted of two videos: In the first video, a character, Mary, pressed three buttons (red, blue, and orange), each producing an animal sound audible only to her through her headphones. After pressing all three buttons, participants were asked whether Mary knew what sound a specific button made (see Table 1 for a sample trial). This question ensured that participants understood that Mary was fully aware of the sounds each button produced after testing them. Following that, in the second video, participants watched a conversation between another character and Mary. The other character posed one of three types of questions: consequent-focused (e.g., “What will happen if I press the blue button?”), antecedent-focused (e.g., “Which button will play a dog sound?”), or neutral (e.g., “What will happen if I press the buttons?”; see Table 1). Mary’s response was always the same, mentioning one button as linking a specific sound: “If you press the blue button, it’ll play a dog barking.” After each video, participants were prompted to make a guess about one of the other buttons. The question was phrased as: “Do you think the orange button plays a dog barking?” Crucially, participants never heard the sounds themselves and had to rely entirely on Mary’s descriptions. Participants could respond with “Yes” if they believed the button would produce the same sound, “No” if they believed it would not, or “Can’t tell” if they felt Mary’s response did not provide enough information to decide. Before starting the task, participants were provided with detailed instructions to ensure they understood their role in the study. The following text was displayed on the screen prior to the task:

> *This is Mary! You will watch short video clips where Mary will press the buttons, listen to the sounds the buttons play, and provide clues for you to answer some questions. Mary does not already know what animal sound the buttons play. Each time she presses a button, she will hear an animal sound in her headphones and will give you clues about the buttons. You will not be able to hear the sounds because your task is to guess which animal sound each button makes based on what Mary says. It's possible that the buttons will make the same sound, but it's also possible that each one makes a different sound. Remember, Mary is not trying to trick you; she is trying to help you. After each video, answer with "Yes," "No," or "Can't tell" based on what Mary says. If you can't tell what sound the button makes based on what Mary says, then you should click "Can't tell."*

The critical conditional statement given by Mary was consistent across trials, but the manipulation of the QUD allowed us to investigate whether participants interpreted the statement literally, resulting in more "Can't tell" responses, or pragmatically, leading to more "No" responses. Note that "Yes" responses were possible, but not expected in critical trials.

Each participant completed 7 trials, with button-sound associations and trial order pseudorandomized. Three of these trials were designed as attention checks. In these trials, Mary uttered a conditional statement using the quantifier *any* (e.g., "If you press any of the buttons, it will play a dog barking"). Participants were expected to infer that all buttons produced the same sound and respond "Yes" when asked about one of the other buttons.

Table 1. Sample trial structure with a still image of the videos

	Does Mary know what sound the **orange** button makes? (Yes/No)
QUD Type:	**Question to Mary:**
Antecedent-focused	Hey Mary, which of these buttons will play a dog sound?
Consequent-focused	Hey Mary, what will happen if I press the blue button?
Neutral	Hey Mary, what will happen if I press the buttons?
Mary's answer: If you press the blue button, it will play a dog barking.	
Critical question: Do you think the orange button plays a dog sound? (Yes / No / Can't tell)	

Results

Performance on the attention check control trials was near ceiling, with 98.63% of responses being correct, confirming that participants followed the task instructions as expected. Following our preregistered exclusion criteria, we excluded 1.62% of trials where participants failed the “speaker-knowledge-check”, indicating inattention on those trials. Additionally, we excluded trials where participants responded “Yes” to critical trials. While “Yes” responses were possible, they were not expected, as the listener did not have direct evidence of what was contained in the third button. These responses were very rare, making up 14 data points (3.53% of the whole dataset).

To investigate whether participants interpreted conditional statements differently based on the QUD Type, we built generalized logistics mixed-effects models predicting the proportion of “No” responses (vs. “Can’t tell” responses) based on fixed effects of QUD Type (neutral, antecedent-focused, consequent-focused). The model included random intercepts for participant and item. Although we initially planned to include random slopes for QUD Type by item, preliminary model fitting indicated that this addition led to overfitting without improving explanatory power, so we did not include it. The model revealed a main effect of QUD Type ($\chi^2(2) = 57.32$, $p < .001$), indicating that participants’ responses varied significantly based on the QUD. To explore these findings, we used the *emmeans* package in R to perform pairwise comparisons among the three levels of the categorical predictor variable, applying Bonferroni adjustments to control for multiple testing. We found a significant difference between consequent-focused (M=0.22, SE=0.10) and antecedent-focused questions (M=0.65, SE=0.10; ($\beta = -18.08$, SE = 2.46, z = -7.34, $p < .001$), as well as between antecedent-focused and neutral questions (M=0.29, SE=0.11; $\beta = -17.30$, SE = 2.37, z = -7.31, $p < .001$). However, there was no significant difference between neutral and consequent-focused questions ($p > .05$). Thus, the interpretation of conditional statements varied significantly depending on the type of question, with consequent-focused and neutral QUDs leading to more literal interpretations, and antecedent-focused QUDs inducing more conditional perfection.

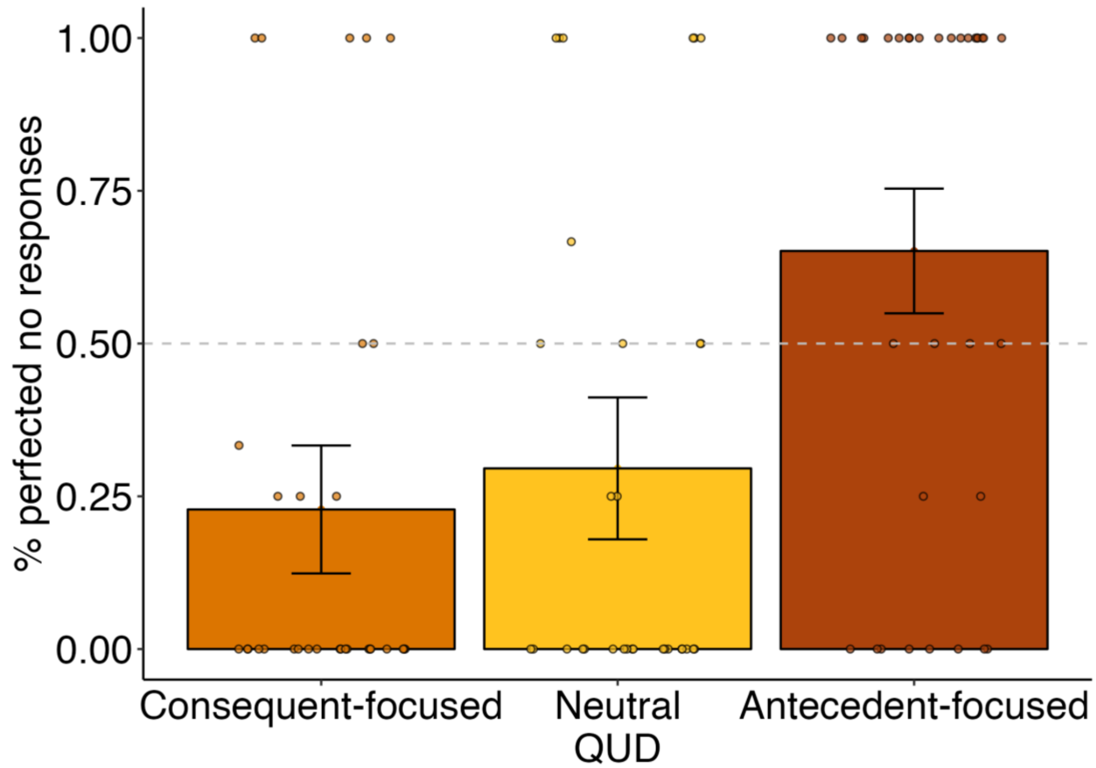

Figure 3. Percentage of perfected "no" responses by QUD Type in Experiment 1. Each dot represents a participant. Error bars represent the standard error of the mean, computed across participants

One potential concern raised by a reviewer is that the antecedent-focused question ("Which of these buttons will play a dog sound?") might trigger a uniqueness inference, biasing participants toward an interpretation where exactly one button plays the sound. To test this possibility, we conducted two follow-up experiments (each n = 32) using antecedent-focused QUDs that replaced the original question with either, "What buttons will play a dog sound?" or "Which buttons will play a dog sound?" (i.e., avoiding the phrase *of these*, in case it was interpreted as "which one of these buttons"). In each experiment, results replicated the original finding: the rate of perfection responses was comparable across QUDs ("what buttons": M= 0.86, *SE* = 0.11; "which buttons": *M* = 0.77, *SE* = 0.13; "which of these buttons": *M* = 0.65, *SE* = 0.10), and if anything was slightly lower in the *which of these* version. A generalized mixed-effects logistic regression including QUD as a fixed effect showed no reliable difference across conditions ($p = .7$) (more details are available on the OSF project page). These results suggest that the observed effects cannot solely be attributed to a uniqueness presupposition potentially introduced by the original phrasing in the antecedent-focused QUD.

Discussion

In Experiment 1, we found strong evidence that manipulating the QUD impacts the likelihood of perfecting conditional statements. In particular, when the context introduced an antecedent-focused QUD, participants were much more likely to perfect conditional statements than when the context introduced a consequent-focused QUD. We also found that perfection was unlikely in the neutral condition. This result was expected on the hypothesis that participants are sensitive to the QUD, since a question like, "What will happen if I press the buttons," is equally likely to generate alternatives that impact both the consequent and antecedent,[10] and strengthen with respect to all such alternatives would result in the implicature that the speaker did not hear *any* sounds when the other buttons were pushed - an inference at odds with contextual evidence that the speaker did in fact listen to the sounds produced by all buttons (see Ginzburg, 1995a; 1995b; Beck & Rullman, 1999 and van Rooij, 2004, for evidence that strengthening relative to a QUD is often blocked or canceled if it is inconsistent with contextual factors such as a questioner's intentions). Overall, these results are compatible with the idea that the QUD

[10] For example, "if you press the orange button, it will play a dog sound", "if you press the red button, it will play a dog sound", "if you press the orange button, it will play a pig sound", "if you press the red button, it will play a cat sound", etc..

impacts perfection by manipulating which propositions are generated as relevant alternatives in the context.

Experiment 2

Experiment 1 showed that manipulating the QUD influences whether conditionals are perfected, suggesting that the QUD constraints which alternatives are considered viable during interpretation. Experiment 2 built on this result by testing how the QUD determines relevance (Groenendijk & Stokhof, 1985; Grice, 1975; Matsumoto, 1995; Benz et al., 2006; Benz, 2006; Lewis, 1979). Specifically, we contrasted Optimally Informative Answers (i.e., responses that correspond exactly to a QUD cell or to a union of cells, such as "If you press the triangles, it plays the sound of a dog barking") with Overly Informative Answers (i.e., responses that eliminate at least one possibility but do not correspond to a QUD cell, such as "If you press the blue square, it plays the sound of a dog barking"). We asked whether overly informative answers are treated as viable alternatives relative to the explicit QUD, or whether only optimally informative, cell-level answers trigger conditional perfection. If overly informative answers count as relevant, perfection should arise across both trial types; if not, only optimally informative answers should yield strengthening.

Methods

All methods and analyses were preregistered prior to data collection. The preregistration is available at https://osf.io/t6rvw/overview?view_only=1cc4f6227e924dcf98a6e2e60a9c7654 All data and analysis code for this are available at https://osf.io/vq72b/overview?view_only=48081f470c6c4fc589651c85e5bfebcf

Participants

The experiment ran on PCIbex (Zehr & Schwarz, 2018). We recruited 56 native English speakers through Prolific, restricting participation to users with U.S.-based IP addresses. One participant was excluded for failing an attention check, leaving a final sample of 55. Participants were paid at a rate matching California's minimum wage at the time ($15.50 per hour).

Materials & Procedure

The experiment was based on the design of Experiment 1, but we introduced two button shapes, squares and triangles, each appearing in two distinct colors. This allowed us to manipulate Answer Type, with Optimally Informative Answer and Overly Informative Answer conditions presented as within-subjects.

Each trial consisted of two short videos. In the first video, a speaker pressed each of the buttons and listened through her headphones. After pressing all three buttons, participants were asked whether Mary knew what sound each shape made. Following that, in the second video, in the second video, another speaker asked an explicit, antecedent-focused QUD: "Hey Mary, which of these shapes, triangles or squares, will play a dog barking?" Mary's response varied by condition. In the Optimally Informative Answer condition, she gave a conditional response that matched the explicit QUD's partitioning, such as "If you press the triangles, it will play a dog barking." In the Overly Informative Answer condition, she provided an overinformative answer (i.e., forms a strict subset of a cell), e.g., "If you press the blue square, it will play a dog barking." After each video, participants were asked whether another button would produce the same sound. In Optimally Informative Answer trials, they were asked whether "the squares" would play the same sound, and in Overly Informative Answer trials, whether the other button of the same shape "red square" would play a dog barking.

Participants never heard the sounds themselves and relied entirely on Mary's statements. They could respond "Yes" if they believed the button would produce the same sound, "No" if they believed it would not, or "Can't tell" if the information was insufficient. Optimally Informative Answer trials served as a control condition, since utterances that completely resolve the QUD should consistently trigger conditional perfection and yield "No" responses. If overly informative answers are treated as relevant, participants should interpret them in the same exhaustive way, again responding "No." If overly informative answers are not treated as relevant, the utterance should not be exhaustified within the same shape domain, leaving open the possibility that both buttons of the same shape produce the same sound; in this case, participants should respond "Can't tell."

Before beginning, participants read detailed instructions explaining that the buttons were connected to a system that played animal sounds. They then completed comprehension questions to confirm understanding. Each participant completed eight trials in total. Button–sound pairings and trial order were pseudorandomized. Two of the eight trials served as attention checks, where Mary said, "If you press any of the shapes, it will play a dog barking." In these cases, the correct response was "Yes" for all follow-up questions.

Results

Following preregistered exclusion criteria, we excluded 2.72% of trials (12 trials total) in which participants failed the speaker-knowledge check, suggesting inattention on those trials. Performance on the attention-check control trials ("If you press any shape…") was at ceiling, consistent with our inclusion criteria and confirming that participants followed task instructions. For critical trials, we excluded 3.03% of responses in which participants responded "Yes."

Although “Yes” responses were possible, they were not expected, since listeners had no direct evidence of what the unmentioned shape would produce.

We analyzed the remaining data using a generalized mixed-effects logistic regression predicting the probability of a “No” response (vs. “Can’t tell”) from Answer Type (Optimally vs. Overly Informative). The model included random intercepts for participants and items and random slopes for Answer Type by participant. The model revealed no significant effect of Answer Type ($\chi^2(1) = 1.93$, $p = .16$), indicating that participants did not differ reliably in how they interpreted Optimally and Overly Informative Answers.

To assess whether participants systematically strengthened conditionals, we calculated each participant’s proportion of “No” responses for each condition and compared these proportions to chance, using one-sample *t*-tests. Participants responded “No” in both conditions, with a higher mean proportion in the Optimally Informative Answer condition ($M = 0.92$, $SE = 0.09$, $t(54) = 14.71$, $p < .001$) than in the Overly Informative Answer condition ($M = 0.84$, $SE = 0.07$, $t(54) = 7.11$, $p < .001$). Thus, both answer types appeared to serve as viable alternatives within the QUD, consistent with the hypothesis that overly informative answers can be treated as relevant, exhaustive responses.

Discussion

In Experiment 2, we asked whether overly informative answers, utterances that refer to a subset of a QUD-defined category, can also trigger conditional perfection. We found that listeners perfected both optimally and overly informative answers at comparable rates, suggesting that overly informative answers can still trigger exhaustive interpretations. This suggests that once the QUD establishes a set of relevant antecedent categories (e.g., triangles vs. squares), reference to a specific subset within one category (e.g., the blue square) is sufficient to exclude alternatives within that same category (e.g., the red square). Taken together with Experiment 1, these findings suggest that perfection depends on how the QUD structures the set of alternatives, and that overly informative answers can likewise license conditional perfection.

Experiment 3

As a final test of how access to alternatives affects conditional perfection, we investigated the role of speaker knowledge. In Experiment 1, we showed that the QUD defines the relevant space of alternatives, but which of these are actually considered depends on what the listener believes the speaker knows. Building on this idea, Experiment 3 manipulated the speaker's knowledge directly, since manipulating this effectively manipulates the set of alternatives available for interpretation. Participants saw trials in which the speaker, Mary, tested either two buttons (partial knowledge) or all three (full knowledge) before making a conditional statement about the sounds produced, such as, "If you press the blue button, it will play a dog barking," similar to Experiment 1. If conditional inferences depend on the speaker's knowledge, we expected participants to perfect conditionals less often when the speaker had partial knowledge. For listeners to exhaustify a statement and derive conditional perfection, they must believe the speaker knowingly omitted relevant alternatives. However, if the speaker lacks knowledge of stronger alternatives, the listener cannot infer that those alternatives are false. This logic parallels scalar implicatures, where listeners strengthen *some* to *not all* only when they assume the speaker is knowledgeable about stronger alternatives and has chosen not to assert them (e.g., Bergen & Grodner, 2012; Breheny et al., 2013; Hochstein et al., 2018; Goodman & Stuhlmüller, 2013; Bale et al., 2024).

Methods

All methods and analyses were preregistered prior to data collection. The preregistration is available at https://osf.io/e7n2f/?view_only=d92a76c2ec1f426797890c05fff6d682. All data and analysis code for this are available at https://osf.io/vq72b/?view_only=48081f470c6c4fc589651c85e5bfebcf.

Participants

We recruited 75 native English speakers (N=72 post exclusions) with an IP address from the United States via the online platform Prolific. Participants received payment for their participation equivalent to California minimum wage at the time ($15.50/hour). As per our pre-registered exclusion criteria, we removed 1 participant for failing to respond to attention checks and 2 participants for failing to respond to knowledge-check questions accurately.

Materials & Procedure

The materials and procedure in Experiment 3 were similar to those in Experiment 1, with one critical difference: we manipulated the speaker's knowledge state as a within-subjects factor while keeping the antecedent-focused QUD constant across all trials.

In each trial, participants watched two videos, one after the other. In the first video, Mary either pressed and listened to all three buttons (full knowledge trials) or only two of the buttons (partial knowledge trials). To ensure that they were paying attention, participants were asked whether Mary knew the sound of the unpressed button. In the second video, Mary asked an antecedent-focused question as in Experiment 1: "Which of these buttons will play a dog sound?" and responded with the conditional statement, "If you press the blue button, it will play a dog barking." Participants were then asked to guess whether the orange button, which Mary either listened to (full knowledge) or did not (partial knowledge), played the same sound. As in

Experiment 1, a literal interpretation of the conditional would result in a "Can't tell" response, whereas a pragmatic interpretation would yield a "No" response.

Participants completed nine trials in total: three full-knowledge trials, three partial-knowledge trials, and three attention-check trials. The attention checks were identical to those in Experiment 1, where Mary pressed all three buttons and provided a general statement such as "If you press any of the buttons, it will play a dog barking". To control for potential order effects, two pseudo-randomized trial sequences were created. Half of the participants began with a full knowledge trial, and the other half began with a partial knowledge trial.

Results

Participants provided correct responses 98.14% of the time on attention check trials ("Yes" responses in attention-check trials). Following our preregistered exclusion criteria, we removed 3.24% of trials in which participants incorrectly identified Mary's epistemic status, as this also indicated a lack of attention. Also, as in Experiment 1, the critical trials where participants answered "Yes" were also excluded. Although "Yes" was an available response, it was not anticipated since the listener lacked direct information regarding the content of the third button. These instances were extremely rare, representing 12 data points (2.77% of the total dataset), and were subsequently removed.

Our main question was whether participants would prefect conditionals less frequently when the speaker was ignorant regarding the truth of relevant contextual alternatives, in the Partial Knowledge condition, relative to when they were fully knowledgeable, in the Full Knowledge condition. To assess this, we constructed a generalized linear mixed-effects model predicting the likelihood of "No" responses (vs "Can't tell" responses) from the contrast-coded fixed effect of Knowledge State (Full, Partial), including Knowledge State as a random intercept by Participant and random intercept by item. This model revealed a significant main effect of Knowledge State ($\chi^2(1) = 26.01$, $p < .001$), reflecting the fact that there were significantly more "No" responses in Full Knowledge condition ($M = 0.72$, SE=.13) than in Partial Knowledge condition ($M = 0.21$, SE=.12; $\beta = -11.07$, $SE = 2.86$, $z = -3.86$, $p < .001$). Thus, participants were significantly more likely to adopt a perfected interpretation of the conditional in the Full Knowledge condition compared to the Partial Knowledge condition (see Fig 4).

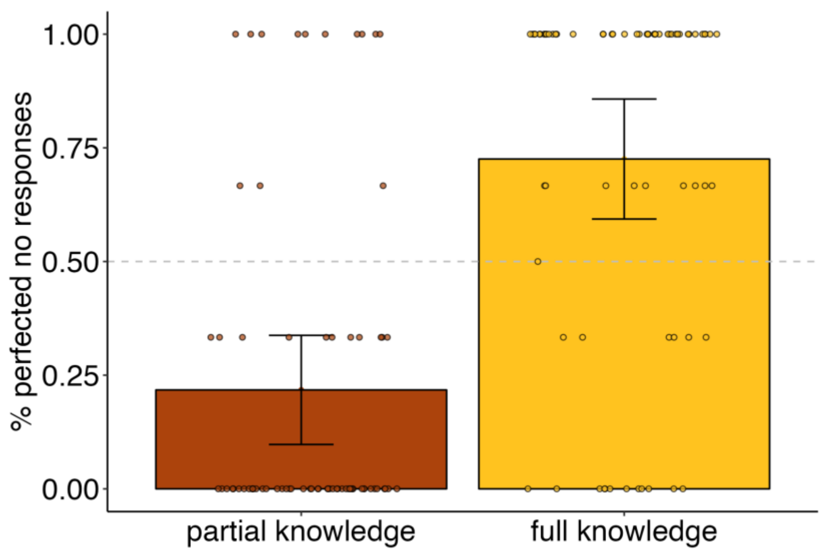

Figure 4. Percentage of perfected "No" responses by Knowledge State in Experiment 3. Each dot represents a participant. Error bars represent the standard error of the mean, computed across participants.

Discussion

In Experiment 3, we found strong evidence that the speaker's epistemic state impacts the likelihood of conditional perfection. Specifically, when the speaker had full knowledge of all relevant alternatives, participants were far more likely to perfect conditional statements, excluding unmentioned possibilities. In contrast, when the speaker's knowledge was incomplete, participants were much less likely to assume that the conditional statement was exhaustive. The broader implications of these results will be discussed further in the general discussion.

General Discussion

We investigated the role of reasoning about alternatives when interpreting conditional statements. To do so, we conducted three experiments and tested whether conditional perfection is sensitive to (1) the contextually defined QUD, (2) the scope of exhaustification with respect to the QUD (i.e., whether it provides an optimally or overly informative answer), and (3) the speaker's knowledge state, which determines whether the listener can infer the falsity of unasserted, non-weaker alternatives. These experiments revealed three main results. First, whereas previous studies find mixed results regarding the role of alternatives and the QUD in interpreting conditionals (Farr, 2011; Cariani & Rips, 2023; Grusdt et al., 2023), in Experiment 1 we found strong evidence that modifications to the QUD impact the likelihood of perfection. For example, when the speaker was asked an antecedent-focused question (e.g., "Which of these buttons will play a dog sound?") and responded with, "If you press the blue button, it will play a dog barking," participants inferred that other buttons, such as the orange one, did not produce a barking sound. In contrast, when the speaker was asked a consequent-focused question (e.g., "What will happen if I press the blue button?") or a neutral question (e.g., "What will happen if I press the buttons?"), participants were less certain about whether the other buttons made a sound. Second, Experiment 2 extended this finding by showing that participants negated both overly informative and optimally informative alternatives at similar rates, thus demonstrating that perfection operates within a proper subset of a QUD cell. Third, in Experiment 3, we found that participants were significantly more likely to perfect conditionals when speakers were knowledgeable of all contextually relevant antecedent statements, relative to when they were ignorant. For example, when the speaker was asked, “Which of these buttons will play a dog sound?” and responded with the conditional statement, e.g., “If you press the blue button, it will play a dog barking,” participants judged that another button – e.g., the orange one – did not make a barking sound in the knowledgeable speaker condition, but that they weren’t sure in the ignorant speaker condition.

These results provide converging evidence that conditional perfection is sensitive to the contextually defined availability of alternative utterances. By extension, they are compatible with the view that CP is derived by a form of quantity implicature wherein a conditional statement of the form “if p then q” is strengthened to exclude alternatives like “if s then q” and “if r then q” (see von Fintel, 2001; van der Auwera, 1997; Cornulier, 1983; Horn, 1972). As discussed by von Fintel (2001), if the computation of CP involves exhaustification with respect to such alternatives, then CP should be sensitive to grammatical mechanisms that constrain or limit alternatives, such as the QUD or speaker knowledge (see Groenindijk & Stokhof, 1984). Further, our finding that participants perfected both optimally and overly informative answers at comparable rates suggests a particular analysis of what counts as an “answer” to a QUD, according to which overly informative answers, which provide more information than needed are relevant to exhaustification (for discussion, see Groenendijk & Stokhof, 1985; Grice, 1975; Matsumoto, 1995; Benz et al., 2006; Benz, 2006; Lewis, 1979). Our finding that perfection is sensitive to the availability of alternatives, as manipulated through the QUD or speaker knowledge, mirrors empirical patterns found with other types of strengthening. First, as noted in the Introduction, multiple previous studies found effects of speaker knowledge on some/all implicatures (e.g., Bale et al., 2024; Bergen & Grodner, 2012; Breheny et al., 2013; Hochstein et al., 2018; Goodman & Stuhlmüller, 2013; Katsos et al., 2023). Similarly, previous studies have found that some/all implicatures are sensitive to QUDs. For example, how-many questions (e.g., "How many boxes have cookies?") generate more implicatures than yes-no questions (e.g., "Are there cookies in any of the boxes?") given a statement with "some" (e.g., "There are cookies in some of the boxes": see Degen, 2013; Degen & Tanenhaus, 2015; Zondervan et al., 2008; Ronai & Xiang, 2020; Kursat & Degen, 2020).

Crucially, a failure to find a sensitivity to QUD and speaker knowledge would have provided strong evidence against the hypothesis that CP is a form of implicature. Nevertheless, despite the similarities we report, there are also important differences between perfection and other putative cases of implicature, as often noted in the literature. For example, although conditional inferences can be restricted to contextually-specified domains of alternatives, as in our experiments, perfection often arises in contexts where there is no clearly defined set of alternatives, and where such a set is in principle difficult to specify. Early neo-Gricean approaches analyzed conditional perfection as a Quantity-based implicature, drawn by comparing "if p, then q" to a stronger alternative such as "if and only if p, then q" (Atlas & Levinson, 1981), but the negation of "if and only if p, then q" does not derive conditional perfection but its opposite. Given this, Horn (2000) proposed that, given a conditional statement like "if p, then q", the relevant stronger alternative is the unconditional statement "q (no matter what)." By choosing a conditional form instead of an unconditional one, the speaker implies that *q* depends on *p*. However, as pointed out by von Fintel (2001), the implicature derived by denying "q (no matter what)" is weaker than perfection, since it only implies that there is at least one condition under which *q* is false, but not that *q* is false in *all* conditions besides *p*. Given this, such an account cannot explain our finding that, after hearing the statement, “If you press the blue button, it will play a dog barking,” participants responded “no” rather than “can’t tell” when asked whether pressing the orange button would play a dog sound. Given that there were three buttons in total, Horn’s analysis derives the implication that at least one of the other two buttons doesn’t play a dog sound, but not the inference that only the blue button will.

A further potential difference between perfection and previously studied forms of scalar implicature concerns processing cost, although this was not investigated here. Whereas some studies argue that some/all implicature demands substantial cognitive effort (Bale et al., 2025; De Neys & Schaeken, 2007; Dieussaert et al., 2011; Marty & Chemla, 2013; Marty et al., 2013), comparable work on perfection suggests that it is computed with little measurable effort, perhaps as the default interpretation (Blochowiak et al., 2022; Evcen & Barner, 2025a; Marcus & Rips, 1979; van Tiel & Schaeken, 2016). Importantly, this does not mean that CP is not derived through scalar mechanisms. One explanation for this pattern, consistent with the data reported here, is that conditional inferences do not require generating or negating lexically encoded alternatives, and that instead relevant alternatives are generated by the immediate linguistic and contextual structure of the utterance, e.g., as determined by the QUD (see Meyer & Feiman, 2021, for a further discussion on processing costs for different types of implicatures).

As noted in the Introduction, past studies on perfection have reported somewhat inconsistent results, raising the possibility that it is derived less consistently than other forms of implicature. Mitigating this concern, however, previous studies find substantial variability in the rate at which implicatures are for particular scales. For example, for *some/all* implicatures, some find very high rates of implicature (e.g., Papafragou & Musolino, 2003) while others find rates closer to 50-60% (e.g., Bale et al., 2025; Dieuleveut et al., 2019; Noveck, 2001; Noveck &

Posada, 2003; van Tiel & Schaeken, 2017, etc.). Similar variability is also found for the frequently discussed case of disjunction, with some studies finding almost no implicatures at all for sentences like, “Every box contains A or B” (Crnic et al., 2015). More generally, different scales are known to differ in how robustly they generate implicatures, and a growing literature has documented substantial scalar diversity (van Tiel et al., 2016; Pankratz & van Tiel, 2021; Gotzner et al., 2018; Hu et al., 2022; Meyer & Feiman, 2021).

In addition to synthesizing findings across different domains of conversational implicature, these findings may also provide insight into why young children often interpret conditionals in non-adult-like ways. In particular, previous studies report that school-aged children (7- to 12-year-olds) struggle to derive the literal meanings of conditionals (e.g., Barrouillet & Lecas, 2002; Klaczynski et al., 2004; Gauffroy & Barrouillet, 2009; De Neys & Everaerts, 2008; DiBacco, 2008; Barrouillet et al., 2008; Markovits, 2017, but see Rumain et al., 1983). Given our finding that perfection is impacted by reasoning about alternatives, one possible explanation of children’s difficulties with perfection is that they struggle to access alternatives. In particular, in order to arrive at a literal, non-perfected, meaning of a statement like, “If you mow the lawn, you will get $5”, children may need to access alternative antecedents (e.g., “If you wash the dishes”) such that mowing the lawn is not the only condition under which the consequent is true (see also Evcen & Barner, 2025b, for evidence that children can compute literal interpretations when alternatives are made explicit in context). Consistent with the idea that they may struggle with this, a growing literature suggests that children’s difficulty computing scalar implicatures arises in part from their inability to access and negate relevant alternatives (Barner & Bachrach, 2010; Bale & Barner, 2013; Barner et al., 2011; Foppolo et al., 2012; Gotzner et al., 2020; Skordos & Papafragou, 2016; Stiller et al., 2015; Tieu et al., 2015).

Further, there is also evidence that children are more likely to make alternative-based inferences when contextual cues explicitly highlight the speaker's intended QUD (Skordos & Papafragou, 2016; Skordos et al., 2022). For instance, when children are tested with alternative expressions that include the quantifier *all* prior to being tested on critical trials containing *some*, children compute implicatures at near adult levels (Skordos & Papafragou, 2016). These findings align with broader evidence that children's sensitivity to alternatives improves when the experimental context clearly signals the speaker's intended QUD, particularly when quantity variations are salient (e.g., Skordos et al., 2020; 2022).

Several interesting questions remained unresolved by the current study. First, although our study shows that listeners reason about speaker states to compute inferences, it does not address the time course for these epistemic inferences, which may be relevant to differentiating between different accounts of implicature. For example, on some past accounts of implicature, listeners assume that speakers are knowledgeable by default, and do not need to actively query the epistemic states of speakers in order to support implicatures (e.g., Rooij & Schulz, 2004). Consistent with this, when participants are placed under cognitive load, they frequently compute *some/all* implicatures even when contextual cues indicate that the speaker is clearly ignorant (Bale et al., 2024). Relatedly, adolescents with autism spectrum disorders compute *some/all* implicatures at a high rate even when the speaker is clearly ignorant (Hochstein et al., 2018). While no previous studies have investigated the time course of epistemic inference in the case of conditionals, past studies, as already discussed above, have argued that CP is computed without significant cognitive cost (Blochowiak et al., 2022; Evcen & Barner, 2025a; Marcus & Rips, 1979; van Tiel & Schaeken, 2016), a result that is compatible with the idea that speaker knowledge is assumed, by default, and only revised on the basis of specific contextual evidence of ignorance. However, future studies should directly test this question, and whether listeners actively reason about speaker states as a precondition for perfection, or simply assume speaker knowledge, and assume a perfected meaning by default.

Another question left open by this study is why some previous reports fail to find the effects of the QUD. While our study does not directly address this question, multiple differences between our methods and those of previous studies might be relevant. First, a key feature of the button choice method used here, and in some previous studies of implicature (Hochstein et al., 2018; Kampa & Papafragou, 2020), is that it asks listeners to make an inference about the world that depends upon strengthening an utterance – as one might in naturalistic conversation – rather than asking them to make metalinguistic judgments about the truth, felicity, or appropriateness of an utterance itself. As one example, in the study of perfection by Cariani and Rips (2023), participants were given a conditional statement “if p then q” under different QUDs and then asked to make an entailment judgment (e.g., does “if p then q” and “q” (together) imply p?). This judgment requires holding in working memory a QUD, and a conditional statement, and then making a meta-linguistic judgment of this conditional with respect to a set of hypothetical truth conditions, all of which may demand more cognitive resources than viewing a scene and evaluating a yes/no question. While our study cannot directly resolve this question, future work should examine the role that different experimental designs play in eliciting perfection.

Finally, future studies should continue to explore the precise ways in which the QUD impacts CP. In our Experiment 2, we found that participants often perfected conditionals in response to an overly informative answer, addressing a dispute among previous accounts regarding the nature of the QUD more generally (e.g., Groenendijk & Stokhof, 1985; Matsumoto, 1995; Benz et al., Benz, 2006; Lewis, 1979). However, our results leave open various ways in which perfection might arise from overly informative answers. One possibility, for example, is that some participants accommodated a more specific QUD, so that an overly informative answer was interpreted as an optimally informative answer to an implicit question. For instance, while the explicit QUD was, “Which shape, triangles or squares, will play a dog sound?”, some listeners may have construed it as, “Which colored square will play a dog barking?” or “Which colored shape will play a dog barking?” In that case, the statement, “If you press the blue square, it will play a dog barking” would be evaluated against more specific alternatives (e.g., “If you press the red square”), yielding the inference that only the blue square produces the sound, with no implications for the other shapes. Future work should explore this possibility, and whether listeners interpret answers relative to the explicit QUD or accommodate a more specific, contextually refined one.

In summary, findings from three experiments indicate that access to alternatives impacts conditional reasoning and the availability of perfected meanings. Experiments 1 and 2 together found that likelihood of conditional perfection depends on how the contextually defined QUD structures the set of relevant alternatives: listeners were more likely to perfect when the question targeted alternative antecedents and did so even when the speaker’s answer referred to a specific subset of that set, suggesting that both optimally and overly answers can serve as viable alternatives for strengthening. Experiment 3 further found that a speaker’s knowledge state also plays a crucial role and that listeners were more likely to perfect conditionals when the speaker was knowledgeable about all relevant alternatives than when they were ignorant. Future research should examine these effects in young children to understand how reasoning about alternatives develops over time. Additionally, investigating the time course of these processes could provide insight into how alternative-based reasoning emerges during language acquisition and is processed in real-time comprehension.

References

Bale, A., & Barner, D. (2013). Grammatical alternatives and pragmatic development. In A. Fălăuş (Eds.) *Alternatives in Semantics. Palgrave Studies in Pragmatics, Language and Cognition* (pp. 238-266). Palgrave Macmillan, London. https://doi.org/10.1057/9781137317247_9

Bale, A., Noguchi, H., Rolland, M. & Barner, D. (2025). Competence by default: Do listeners assume that speakers are knowledgeable when computing conversational inferences? *Journal of Semantics, ffae016,* https://doi.org/10.1093/jos/ffae016

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology, 60*, 40-62.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84-93. https://doi.org/10.1016/j.cognition.2010.09.011

Barrouillet, P., & Lecas, J. F. (1998). How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition*, *67*(3), 209-253. https://doi.org/10.1016/S0010-0277(98)00035-8

Barrouillet, P., Gauffroy, C., & Lecas, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, *115*(3), 760. https://doi.org/10.1037/0033-295X.115.3.760

Beck, S. & Rullmann, H. (1999). *A flexible approach to exhaustivity in questions*. *Natural Language Semantics, 7*(3), 249–298. https://doi.org/10.1023/A:1008373224343

Benz, A. (2006). Utility and relevance of answers. In A. Benz, G. Jäger, & R. van Rooij (Eds.), *Game theory and pragmatics* (pp. 195–224). Palgrave Macmillan.

Benz, A., Jäger, G., & van Rooij, R. (2006). An introduction to game theory for linguists. In A. Benz, G. Jäger, & R. van Rooij (Eds.), *Game theory and pragmatics* (pp. 1–82). Palgrave Macmillan.

Bergen, L. & Grodner. D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*(5), 1450-1460.

Blochowiak, J., Castelain, T., Rodriguez-Villagra, O. A., & Musolino, J. (2022). If and only if people were logical! The effect of pragmatic enrichment on reasoning with abstract and realistic materials. *Journal of Pragmatics*, *197*, 137-158. https://doi.org/10.1016/j.pragma.2022.05.013

Breheny, R., H. J. Ferguson, & N. Katsos. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition, 126*(3), 423-440.

Cariani, F., & Rips, L. J. (2023). Experimenting with (Conditional) Perfection*:* Tests of the exhaustivity theory. *In* S. Kaufmann, D. Over & G. Sharma (Eds.). *Conditionals: Logic, Semantics, Psychology (*pp. 235-274).

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and Beyond, 3,* 39-103.

Chierchia, G., Fox, D. & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. Semantics: An International Handbook of Natural Language Meaning 3, 2297-2332.

Cornulier, B. (1983). ‘If’ and the Presumption of Exhaustivity. *Journal of Pragmatics, 7*(3), 247-249. https://doi.org/10.1016/0378-2166(83)90012-7

Crnič, L., Chemla, E., & Fox, D. (2015). Scalar implicatures of embedded disjunction. *Natural Language Semantics*, *23*(4), 271-305.

De Neys, W., & Everaerts, D. (2008). Developmental trends in everyday conditional reasoning: The retrieval and inhibition interplay. *Journal of Experimental Child Psychology, 100*(4), 252-263. https://doi.org/10.1016/j.jecp.2008.01.007

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, 54(2), 128–133.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667-710.

Di Bacco, F. (2018). Ambiguous questions and perfectible conditionals: the perspective from language acquisition [Doctoral dissertation]. Ulster University.

Dieuleveut, A., Chemla, E., & Spector, B. (2019). Distinctions between primary and secondary scalar implicatures. *Journal of Memory and Language*, *106*, 150-171.

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *Quarterly Journal of Experimental Psychology*, 64(12), 2352–2367.

Evcen, E. & Barner, D. (2025a). Already perfect: Language users access the pragmatic meanings of conditionals first. *Open Mind, 9,* 1098–1120. https://doi.org/10.1162/opmi.a.17

Evcen, E., & Barner, D. (2025b). Preschoolers compute literal and pragmatic meanings of conditionals with contextual support. *Proceedings of the Annual Meeting of the Cognitive*

Science Society, 47.

Farr, M.C. (2019). Focus Influences the Presence of Conditional Perfection: Experimental Evidence. *Proceedings of Sinn Und Bedeutung*, *15*, 225–240. Retrieved from https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/374

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development, 8*(4), 365-394. https://doi.org/10.1080/15475441.2011.626386

Fox, D. (2007). Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics* (pp. 71-120). London: Palgrave Macmillan UK.

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review, 29*(4), 249-282. https://doi.org/10.1016/j.dr.2009.09.002

Gazdar, G. (1979). Pragmatics: implicature, presupposition, and logical form. New York: Academic Press.

Geurts, B. (2010). Quantity Implicatures. Cambridge: Cambridge University Press.

Ginzburg, J. (1995a). *Resolving Questions, I. Linguistics and Philosophy, 18*(5), 459–527. https://doi.org/10.1007/BF00985365

Ginzburg, J. (1995b). *Resolving Questions, II. Linguistics and Philosophy, 18*(6), 567–609. https://doi.org/10.1007/BF00983299

Goodman, N.D. and A. Stuhlmüller. (2013), Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics Cognitive Science, 5,* 173-184.

Gotzner, N., Barner, D., & Crain, S. (2020). Disjunction triggers exhaustivity implicatures in 4- to 5-year-olds: Investigating the role of access to alternatives. *Journal of Semantics, 37*(2), 219-245. https://doi.org/10.1093/jos/ffz021

Gotzner, N., Solt, S., & Benz, A. (2018). Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in Psychology*, *9*, 1659.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41–58). Academic Press.

Groenendijk, J & Stokhof, M. (1984). On the Semantics of Questions and the Pragmatics of Answers. In F. Landman & F. Veltman (Eds.), *Varieties of Formal Semantics: Proceedings of the Fourth Amsterdam Colloquium* (pp. 143-170). Foris.

Grusdt, B., Liu, M., & Franke, M. (2023). Testing the influence of QUDs on the occurrence of Conditional Perfection. Proceedings of *Experiments in Linguistic Meaning*, *2*, 104-116.https://doi.org/10.3765/elm.2.5413

Herburger, E. (2015). Conditional perfection: the truth and the whole truth. In *Semantics and Linguistic Theory* (pp. 615-635).

Hirschberg, J. B. (1985). A theory of scalar implicature (natural languages, pragmatics, inference) [Doctoral dissertation]. University of Pennsylvania.

Hochstein, L., Bale, A. & Barner. D. (2018). Scalar implicature in absence of epistemic reasoning? The case of Autism Spectrum Disorders. *Language Learning and Development, 14*(3), 224-240.

Horn, L. R. (1972). On the Semantic Properties of Logical Operators in English. [Doctoral dissertation] University of California Los Angeles

Horn, L. R. (2000). From if to iff: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics, 32*(3), 289-326.

Hu, J., Levy, R., & Schuster, S. (2022). Predicting scalar diversity with context-driven uncertainty over alternatives. In E. Chersoni, N. Hollenstein, C. Jacobs, Y. Oseki, L. Prévot, & E. Santus (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 68–74). Association for Computational Linguistics.

Kampa, A., & Papafragou, A. (2020). Four-year-olds incorporate speaker knowledge into pragmatic inferences. *Developmental Science*, *23*(3), e12920.

Katsos, N., Gonzales de Linares, B., Ostashchenko, E., & Wilson, E. (2023). Perspective-taking in deriving implicatures: The listener's perspective is important too. *Cognition, 241*, 105582. https://doi.org/10.1016/j.cognition.2023.105582

Klaczynski, P. A., Schuneman, M. J., & Daniel, D. B. (2004). Theories of conditional reasoning: a developmental examination of competing hypotheses. *Developmental Psychology*, *40*(4), 559.

Kursat, L., & Degen, J. (2020). Probability and processing speed of scalar inferences is context-dependent. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 42.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Lewis, D. (1973). Counterfactuals. Oxford: Basil Blackwell.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339-359.

Marcus, S. L., & Rips, L. J. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior, 18*(2), 199-223. https://doi.org/10.1016/S0022-5371(79)90127-0

Markovits, H. (2017). In the beginning stages: Conditional reasoning with category based and causal premises in 8-to 10-year olds. *Cognitive Development, 41*, 1-9. https://doi.org/10.1016/j.cogdev.2016.10.002

Marty, P. P., & Chemla, E. (2013). Scalar implicatures: Working memory and a comparison with only. *Frontiers in Psychology*, 4, 403.

Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, 133, 152–163.

Matsumoto, Y. (1995). The conversational condition on Horn scales. *Linguistics and Philosophy*, 21-60.

Meyer, M. C., & Feiman, R. (2021). Priming reveals similarities and differences between three purported cases of implicature: Some, number and free choice disjunctions. *Journal of Memory and Language*, *120*, 104206.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165-188.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203-210.

Pankratz, E., & Van Tiel, B. (2021). The role of relevance for scalar diversity: a usage-based approach. *Language and Cognition, 13*(4), 562-594.

Roberts, C. (1996). Information structure in discourse: Toward a unified theory of formal pragmatics. *Ohio State University Working Papers in Linguistics, 49*, 91-136.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics, 5*, 1-69. https://doi.org/10.3765/sp.5.6

Ronai, E., & Xiang, M. (2021). Pragmatic inferences are QUD-sensitive: an experimental study. *Journal of Linguistics*, *57*(4), 841–870.

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. *Developmental Psychology, 19*(4), 471. https://doi.org/10.1037/0012-1649.19.4.471

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition, 153*, 6-18. https://doi.org/10.1016/j.cognition.2016.04.006

Skordos, D., Feiman, R., Bale, A., & Barner, D. (2020). Do children interpret 'or' conjunctively?. *Journal of Semantics, 37*(2), 247-267. https://doi.org/10.1093/jos/ffz022

Skordos, D., Myers, A., & Barner, D. (2022). Quantifier spreading and the question under discussion. *Cognition, 226*, 105059. https://doi.org/10.1016/j.cognition.2022.105059

Stalnaker, R. C. (1968). A theory of conditionals. In *Ifs: Conditionals, belief, decision, chance and time* (pp. 41-55). Dordrecht: Springer Netherlands.

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development, 11*(2), 176-190. https://doi.org/10.1080/15475441.2014.927328

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2016). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, *33*(2), 269-298.

van der Auwera, J. (1997). Pragmatics in the last quarter century: The case of conditional perfection. *Journal of Pragmatics, 27*(3), 261-274. https://doi.org/10.1016/S0378-2166(96)00058-6 van Rooij, R. (2004). Utility of Mention-Some Questions. *Research on Language and Computation*, 2(3), 401–416. https://doi.org/10.1007/s11168-004-1975-0

van Rooij, R. & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information, 13*, 491–519.

van Tiel, B., & Schaeken, W. (2017). Processing conversational implicatures: Alternatives and counterfactual reasoning. *Cognitive Science*, 41, 1119–1154.

van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, *33*(1), 137-175.

von Fintel, K. (2001). Conditional strengthening. Unpublished manuscript.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).

Zondervan, A., Meroni, L., & Gualmini, A. (2008). Experiments on the role of the question under discussion for ambiguity resolution and implicature computation in adults. *Proceedings of Semantics and Linguistic Theory*, *18*, 765-777.