Making the Question under Discussion explicit shifts counterfactual interpretation

Ebru Evcen (eevcen@ucsd.edu)
Department of Linguistics, 9500 Gilman Drive #0108
San Diego, CA 92093 USA

Eva Wittenberg (wittenberge@ceu.edu)
Department of Cognitive Science, Quellenstraße 51, A-1100
Vienna, Austria

Abstract

The comprehension of counterfactual statements ('If there had been zebras, there would have been lions') has been subject to much research, but two key questions remain: Can comprehenders interpret counterfactuals without relying on causal inferences? And can comprehenders reach the actual state interpretation relying only on grammatical cues, or is this interpretation triggered by communicative goals? We answer these questions by relying on non-causal counterfactuals, and by manipulating the Question under Discussion between experiments: In Exp. 1, we replicate Orenes et al. (2019), using a web-based eye-tracking paradigm. In Exp. 2, we make the QuD explicit by asking about the actual state of affairs. The results reveal that making a contextually relevant alternative explicit via the QuD shifts counterfactual interpretation, but in general, the suppositional state interpretation is preferred in non-causal counterfactuals. These results imply that the driving forces behind counterfactual processing are pragmatic, not syntactic.

Keywords: counterfactual interpretation; visual-world eye tracking; consideration of alternatives; QuD

Introduction

One often-cited uniquely human ability is the capacity to think and talk about things that we know are not real: fantastical creatures, events in the future, or counterfactual states. In this paper, we reverse-engineer the driving forces behind the interpretation of counterfactuals: How do we know that an utterance is not pointing to the real world, but to the one we know to be false? And how much of this interpretation is triggered by the grammatical structure itself, as opposed to causal inference?

It has been argued that language itself, and specifically the logical properties of its compositional syntax, is the driving force behind the capacity: When a speaker assembles a sentence such as 'If cats were vegetarians, families could feed them with a bowl of carrots' (Ferguson & Sanford, 2008), the combined powers of the subjunctive ('were'), the conjunction ('if'), and the clause structure (antecedent and consequent) trigger a composition process in the hearer that leads to an interpretation that, in fact, cats are not vegetarians and do not eat carrots.

Under this view, the process of comprehending counterfactuals resulting in an *implied actual state* interpretation (non-vegetarian cats) is driven by the aforementioned bundle of grammatical means (subjunctive, conjunction, clause structure). This account has been formalized as Mental Model Theory: The meaning of a counterfactual conditional '*if p then q*' is selected from its logically possible alternatives, which are the *suppositional state p & q* (cats are vegetarians and eat carrots), the logically possible *not-p & q* (cats are not vegetarians and eat carrots), and the implied actual state *not-p & not-q* (cats are not vegetarians and don't eat carrots; Johnson-Laird & Byrne, 2002; Byrne, 2005). The Mental Model theory predicts that listeners simultaneously draw upon both the suppositional state and implied actual state while having to resolve the conflict for successful communication and that they arrive swiftly and reliably at the implied actual state interpretation.

An alternative account of the interpretation of counterfactual statements is Suppositional Theory (Evans & Over, 2004; Evans, 2007), which ascribes more importance to pragmatic factors. This account argues that listeners only arrive at the implied actual state interpretation when it is pragmatically preferred. That is, people consider one single interpretation at a time, based on the most relevant (or possible) interpretation in a given context. Then, this interpretation is evaluated with respect to communicative goals of the context and accepted or modified as a result (Evans, 2006). In counterfactual conditionals, just like in indicatives, a single interpretation is evoked (e.g., vegetarian cats), and using the pragmatic cues available, people decide whether the consequent would follow (e.g., cats eating carrots), based on the probability of *p* given *q*. However, if the link between the two is not strong enough (i.e., not sufficiently supposed by the context), another possibility (e.g., cats are not vegetarians and don't eat carrots) is generated. In short, the Suppositional Theory argues that the sentential pragmatics of a counterfactual can sway the consideration of the implied actual state over the suppositional state.

Here, we study whether the comprehension of counterfactuals is also guided by broader pragmatic factors, that is, the communicative context itself. A listener could make one of two assumptions about counterfactual, and the intentions behind its utterance: On the one hand, she may think the suppositional world is relevant, otherwise, it would not have been mentioned; on the other hand, the communicative goal of the speaker could be to establish the implied actual state as "Question Under Discussion" (QuD; Roberts, 1996; 2004). Understanding what other people 'mean' relies on the ability to make inferences about people's intentions (Sperber & Wilson, 2002). These inferences are based on cooperative principles of communication (Grice,

1975) and generated automatically and effortlessly to avoid delays in comprehension. For instance, although the counterfactual 'If cats were vegetarians, families could feed them with a bowl of carrots' implies that cats are not vegetarians and don't eat carrots (Ferguson & Sanford, 2008), the sentence itself need not always give rise to this inference. Instead, this inference is dependent on the question that is being discussed, which may be explicit or implicit (e.g., Skordos & Barner, 2019). If the QuD is about a cat's hypothetical eating habits, then the suppositional alternative may be preferred (vegetarian cats eating carrots). If, on the other hand, the QuD is about the real-world/actual state of affairs, then the implied factual alternative (non-vegetarian cats) may be preferred. This leads to different predictions for the two theories under consideration: For the Mental Model Theory, the communicative context is not relevant as consideration of implied actual state is automatic; only the Suppositional Theory would predict that with the QuD explicitly referring to the actual world (along with some other factors), consideration of the actual implied state would increase. One difficulty, however, is that counterfactual comprehension is tightly linked to causal coherence. Below, we will argue why in our studies, we rely on non-causal counterfactuals.

Present studies: Non-causal counterfactuals

One crucial property of counterfactual statements is that they allow inferences about reasoning processes. In fact, the most prominent way to study causal reasoning is to use counterfactual scenarios (Gerstenberg et al., 2017, 2020; Kominsky et al., 2021), precisely because the causal relationship between antecedent and consequent is a core component of most counterfactuals we encounter.

People automatically try to draw (causal) inferences between propositions (Kehler, 2002):

(1) *Joe bites his nails. Mary left him.*

(2) *Joe bites his nails. He has a brother.*

In (1) one cannot help but infer that Mary's leaving was caused by Joe's nailbiting (or vice versa), whereas (2) does not result in causal inference; not coincidentally, while the counterfactual (1') is coherent, the counterfactual (2') is hard to make sense of:

(1') *If Joe didn't bite his nails, Mary wouldn't have left him.*

(2') *If Joe didn't bite his nails, he wouldn't have a brother.*

Thus, the presence of causal coherence plays a role in comprehending counterfactuals, but how instrumental this role is for comprehension remains unclear; in fact, one open question is how well people understand counterfactual statements in the absence of causal coherence between clauses – in fact, the existing literature has studied counterfactual comprehension only in causal scenarios, from nonrealistic, real-world inconsistent scenarios (e.g., 'If dogs had gills, Dobermans would breathe underwater'; Nieuwland, 2013) to everyday events (e.g., 'If David had been wearing his glasses, he would have read the poster easily'; Ferguson and Cane, 2015; see also Romoli et al., 2019; 2022).

Even within the empirical literature on causally structured counterfactuals, the evidence is mixed. While some results indicate that listeners only ever interpret counterfactuals referring to the *suppositional state p & q* (vegetarian cats eating carrots) (Ferguson, Scheepers & Sanford, 2009; Nieuwland & Martin, 2012), others highlight individual differences, showing that some people interpret the sentence referring to the *implied actual state not-p & not-q* (cats are not vegetarians and don't eat carrots; de Vega, Urrutia, & Riffo, 2007; de Vega & Urrutia, 2012; Ferguson, 2012; Stewart, Haigh, & Kidd; 2009), or consider both alternative interpretations simultaneously (Quelhas, Rasga, & Johnson-Laird, 2018; Santamaria, Espino, & Byrne, 2005; Thompson & Byrne, 2002).

The mixed empirical picture could be due to multiple factors, from task demands to the semantic content of the stimuli: Between studies, one finds a varying degree of plausibility in the scenarios, from causal sequences of connected events, which, in turn, might lead participants to follow the implied actual state very closely (e.g., de Vega & Urrutia, 2007; Ferguson & Cane, 2015), to scenarios inconsistent with real-world knowledge, which might have led participants to consider the supposed alternatives on the basis of the relevance assumption: The mentioned facts must be relevant, otherwise they would not be mentioned in the first place (e.g., Nieuwland & Martin, 2012). It has also been argued that in nonrealistic scenarios, the factual-counterfactual distinction is more grounded and can be more easily retrieved (Black, Williams, & Ferguson, 2018; Dai, Kaan, & Xu, 2021).

In short, the nature, strength, and plausibility of the causal relationships within counterfactual scenarios introduce noise that can obscure the mechanisms behind counterfactual comprehension itself. To really test the predictions of the Mental Model Theory, one needs counterfactuals that are built without the reliance on pragmatic context; likewise, to test whether QuD affects counterfactual comprehension, causal inferences should be avoided as to not interfere with the communicative context. One study in which counterfactuality is established only by linguistic means is by Orenes et al. (2019). They examined the online processing of counterfactual and indicative conditionals (3-4) in a visual world eye-tracking study (Fig. 1). Crucially, the relationship between antecedent and consequent was not based on causal linkage; instead, besides being drawn from semantically close neighborhoods, the propositions in antecedent and consequent had no obvious connection:

(3) If there had been zebras, there would have been lions.

(4) If there are zebras, then there are lions.

In three experiments, Orenes et al. (2019) asked at what point in time in comprehension of counterfactuals people consider the suppositional state [+ZEBRA, +LION], the implied actual state [-ZEBRA, -LION], or both, by analyzing participants' gaze pattern to each alternative state from the onset of the antecedent NP 'zebras' until the end of the utterance (Fig. 1).The results showed that participants fell into one of three groups: one group only considered the suppositional state [+ZEBRA, +LION], the second group only considered the implied actual state [-ZEBRA, -LION], and the third group considered both alternatives simultaneously. Orenes et al. (2019) attributed the behavior of the first group to the differences in their working memory capacity (Ferguson & Cane, 2015), shallow processing (Ferreira, Bailey & Ferraro, 2002), or they simply might not have understood the task demands. Although only half of the participants represented the implied actual state (or both), Orenes et al. (2019) concluded that the data overall reject the hypothesis that people only ever represent the suppositional alternative (cf., Evans & Over, 2004).

However, in this set of experiments, it is unclear how participants actually interpreted the sentences, and specifically, whether they commit to the actual state: the interpretation was not measured. The 'explicit' instruction provided in Exp 1 was 'to look at the image that corresponds to the meaning of the sentence'. However, the 'meaning' of a counterfactual could be mapped either to the suppositional state (e.g., Nieuwland & Martin, 2012) or to the implied factual state (e.g., Ferguson & Cane, 2015), possibly to both. Furthermore, the task at the end of each trial was to answer a yes/no comprehension question about the initial context of the story, not about the critical utterance.

Here, we aim to understand the role of explicitly shifting the QuD to the implied actual state, by first replicating Orenes et al. (2019) repeating their ambiguous instruction to look at the image corresponding to the 'meaning' of the sentence, that is, setting no QuD. Second, we will shift the QuD explicitly to the implied actual state, by asking participants what the real world would look like. Our key prediction is that if people map incoming utterances to what they consider pragmatically relevant (i.e., Suppositional Theory), then explicit QuD is expected to evoke more looks to the implied actual state. If, on the other hand, the implied actual state alternative is always strongly evoked as part of the logical meaning of counterfactuals (e.g., mental model theory), then the presence of an explicit QuD will not yield a different pattern.

Exp. 1: No explicit QuD

Participants: We recruited 82 self-declared native speakers of English via Amazon Mechanical Turk. We used CloudResearch (Litman, Robinson, & Abberbock, 2017) services, restricting the participant pool to users with an IP address in the United States, with a completed task acceptance rate of 80% or higher, and with at least 100 tasks completed. Following Morgan et al. (2020), we calculated track loss for the eye-tracking duration (around 8000ms) and excluded those participants (N=28) who failed to provide at least one trial's worth of data (with 25% of track loss as threshold) in each of the 4 cells of the experiment.

Materials and Procedure: Participants were presented with simultaneous auditory and visual input. We replicated the design by Orenes et al. (2019) as truthfully as possible, translating stimuli from Spanish to English. The experiment used a 2x2 within-subject design: Conditional (counterfactual and indicative) and Conjunction (affirmative and negative).

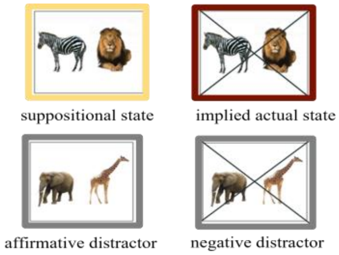

Figure 1: Example display from Orenes et al. (2019): The suppositional alternative [top left, +ZEBRA, +LION]; implied actual state for counterfactuals [top right; -ZEBRA, -LION]; bottom images are distractors.

Table 1: Example stimuli

Condition	Example utterance
Opening utterance	*Jack went to the zoo to visit the animals with his parents. While there, they heard some people say*
Critical utterance	*If there are/had been zebras, then there are/would have been lions.*
Follow-up utterance	*Jack realized that there were (no) zebras and there were (no) lions.*
Closing utterance	*Finally, Jack and his family went to a restaurant to eat.*

In each trial, participants listened to a pre-recorded opening scenario (Table 1). Each critical sentence was paired with a visual scene containing four images (Fig.1): two target images (e.g., a zebra and a lion, and the same image crossed out) as well as two distractor images (e.g., another pair of wild animals such as an elephant and a giraffe, and the same image crossed out).

Participants were randomly assigned to one of 8 counterbalanced, Latin-squared lists containing 36 critical trials each (9 per each Conditional x Conjunction). The order of each item and the position each image appeared on the screen were randomized per participant. There was also a practice block of 4 trials before the experimental trials.

The experiment was hosted online on PennController IBEX (Zehr & Schwarz, 2018), which uses the JavaScript library Webgazer.js (Papautsaki, Laskey, & Huang, 2017), and participants completed the study remotely via their own webcam. Each trial began with a central fixation cross while participants listened to the opening scene. Then, the visual display appeared for 200ms followed by the critical utterance and remained on the screen until the end of the story. Each trial was followed by a simple yes/no comprehension check question (e.g., 'Did Jack and his family go to the zoo?') and participants answered by clicking either a 'yes' or 'no' button. The experiment took 25-30 minutes to complete.

Analyses and Results: We followed Orenes et al. (2019) and ran *t*-tests against baseline; additionally, we conducted

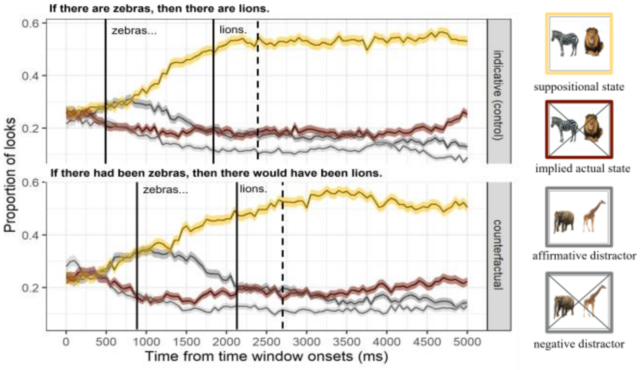

Figure 2: Probabilities of fixations for indicatives (top) and counterfactuals (bottom) in Exp 1. Standard errors are represented by transparent ribbons.

growth curve analyses to capture non-linear changes in fixation proportions over the time course (hereafter GCA, Mirman, Dixson & Magnuson, 2008; Mirman, 2017). We only report GCA results here due to space constraints, but the *t*-tests comparisons revealed similar statistical results, which can be found on OSF.[1]

For growth curve analyses, we aggregated looks to the images corresponding to the suppositional state and the actual state across participants and items and calculated empirical logit transformation of fixation probabilities to the images for correction (Barr, 2007). The empirical logit transformed proportions to the critical images served as our dependent variable. Like in the original study, the critical time window lasted from 300ms (i.e., earliest word-driven fixations) to 2050ms (i.e., the onset of the second NP in the conditional, 'lions'). We conducted separate growth curve analyses for each critical image (i.e., zebras & lions or crossed-out zebras & lions) with the deviation-coded fixed effects of Conditional (indicative, -0.5; counterfactual, 0.5) and its interaction with each of the time terms (linear, quadratic, and cubic). The model included random intercepts by participant and item, as well as random slopes of conditional by participant and by item. We report the intercept (total fixations), the linear term (how fast the curve increases), the quadratic term (U-shaped pattern of fixation ratios), and the cubic term (the sharpness of rise and fall).

Figure 2 shows participants' eye-gaze patterns during the course of each condition (top: indicative, bottom: counterfactual), revealing how the proportions of fixations to each image on the visual scene changed through the course of the utterances. In both types of conditionals, participants increased their fixation on the suppositional state [+ZEBRA, +LION] and the affirmative distractor [+ELEPHANT, +GIRAFFE] from very early on, and as soon as the referent ambiguity is resolved, i.e., after 200ms of the mention of 'zebras', their looks on the affirmative distractor rapidly decreased. Fixations on the suppositional state remained stable through the time measured both for indicative and counterfactual conditionals.

For looks to the suppositional state (Fig. 3A), we found a significant interaction between Conditional and the linear term (β=-1.65, SE=0.43, t=-3.82, p<.001) and between Conditional and cubic term (β=1.67, SE=0.43, t=3.88, p<.001), reflecting the faster and steeper fixations on the suppositional state in indicative conditionals than in counterfactual conditionals. In looks to the actual state (Fig. 3B), there was a significant interaction between Conditional and the linear term (β=-0.82, SE=0.35, t=-2.31, p<.05) and between Conditional and the quadratic term (β=1.1, SE=0.35, t=3.09, p<.01), reflecting the linear decrease in looks to the actual state in indicative conditionals, and a U-shape trend (i.e., a decrease followed by an increase) in counterfactual conditionals. Numerically, the coefficient on the main effect of Conditional was negative, indicating more looks to the actual state in indicative conditionals than in counterfactuals, but this pattern did not reach significance (Fig. 3).

Discussion: Our results did not replicate Orenes et al. (2019). In Orenes et al. (2019), for the overall group data, fixations to the suppositional state were more and quicker in indicatives than in counterfactuals, and the fixations to the implied actual state were more and quicker in counterfactuals than in indicatives. However, in our study, participants' overall looks to the images did not differ across conditions: they increasingly directed their gaze towards the image corresponding to the suppositional state [+ZEBRA, +LION] in both indicative and counterfactual conditionals (Fig. 2).

. With these data as a baseline, we can trace the effect of making the QuD explicit on the consideration of alternatives in counterfactual comprehension.

Exp. 2: Shifting the QuD to the implied actual state

Participants: 56 self-declared native speakers of English participated in this experiment, recruited as before. The exclusion criteria were the same as Exp 1.; in consequence, two participants were excluded from the analysis.

Materials and Procedure: The aim of Exp 2 was to measure the effect of explicit QuD to counterfactual interpretation. We added a task that aimed to shift the attention of the participants to the actual state of the world. We asked the participants what the fictional listener in the story should expect. The question first appeared in the instructions and repeated after each trial. The experiment used Conditional (counterfactual and indicative) as a within-subject factor. We excluded the follow-up sentences (since we were interested only in the conditional utterance) and so we modified the opening and closing utterance accordingly

In each trial, participants listened to an opening scenario (e.g., 'While Jack was at the zoo visiting the animals, he said to his friend') followed by the critical utterance, either an indicative conditional (e.g., 'If there are zebras, then there are lions') or a counterfactual conditional (e.g., 'If there had been zebras, then there would have been lions').

[1] Data and analysis code are publicly available at https://osf.io/3zy6v/.

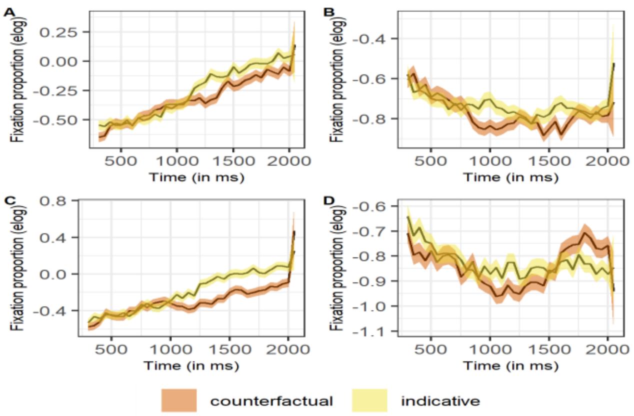

Figure 3: Fixation probabilities on the suppositional state (A)/(C) and the implied actual state (B)/(D) in Exp. 1 (top) and Exp. 2 (bottom) by Conditional from 300ms until the onset of NP in the consequent.

Each trial was followed by the question ‘What should the friend expect?’, and participants clicked on the corresponding picture. The question was formulated in such a way that in the indicative conditional, any of the pictures could be a potential answer; but crucially, in counterfactual conditional, only the implied actual state [-ZEBRA, -LION] would be. Each list included 36 critical trials (18 per each Conditional). Participants were randomly assigned to one of the 4 lists. The order of each item and the position each image appeared on the screen in were randomized per participant.

Results: Following our analysis in Exp 1, we fitted separate linear models for each critical image to predict log-odds from fixed effects of Conditional and its interaction with all the time terms. For looks to the suppositional state (Fig. 3C), there was a main effect of Conditional (*β*=-0.11, SE=0.04, *t*=-2.56 *p*<.05), revealing more fixations in indicative conditionals than in counterfactuals. A significant interaction between the linear term and Conditional (*β*=-3.45 SE=0.45, *t*=-7.64, *p*<.001) also supported this effect, indicating a sharper/faster increase in the looks in indicative conditionals. There was also a significant interaction between the Conditional and the cubic term (β=1.92, SE=0.45, *t*=4.26, *p*<.001), which reflected the fluctuations in counterfactual conditionals (i.e., an increase followed by a decrease and then flatness). In looks to the implied actual state (Fig. 3D), there was a significant interaction between the Conditional and the linear term (*β*=1.84, SE=0.35, *t*=5.25, *p*<.001) and the Conditional and the quadratic term (*β*=1.16, SE=0.34, *t*=3.33, *p*<.001). In counterfactual conditionals, looks to the implied actual state followed a U-shaped pattern: A decrease in the fixations was followed by a steeper/faster linear increase in counterfactuals whereas the fixations continuously decreased and became flat in indicative conditionals.

Analysis by clicks: We split the data by clicks: There were two groups of people: those who clicked on the suppositional state (N=34) and those who clicked on the implied actual state (N=20) after seeing the question of ‘What should the friend expect?’. We filtered out the selections of distractors as they could be an indication of loss of attention or such, which resulted in 1.49% of data exclusion for indicative conditionals and 1.13% of data exclusion for counterfactual conditionals. Figure 4 shows the proportion of fixations to each image on the visual scene by selection for counterfactual conditionals only. One group of participants behaved similar to Exp.1 and only considered the suppositional state, and the other group considered all pictures equally and then clicked on the implied actual state picture only after hearing the first NP.

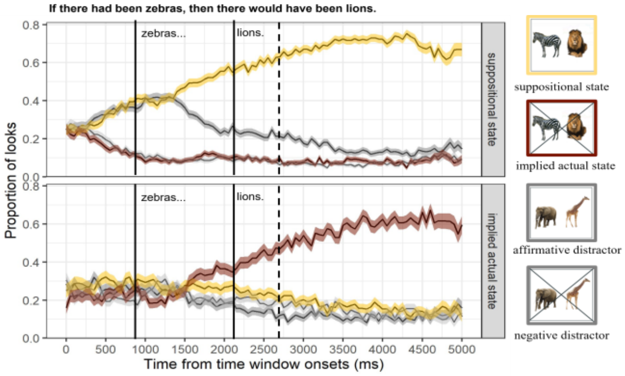

Figure 4: Probabilities of fixations for counterfactuals by subgroups selecting the suppositional state as answer (top), or the implied actual state (bottom). Indicative results not shown.

To compare the effect of explicit QuD on the consideration of alternatives, we calculated the target preference score and used as our dependent variable: *ln*(P(implied actual state)/P (suppositional state), where *ln* refers to the natural algorithm. We fitted a model with fixed effects of Selection (deviation-coded, suppositional state: -0.5; implied actual state: 0.5) Time terms and their interaction. The model also included random intercepts for Participant and Item and random slope of Selection by Participant and Item. We found a main effect of Selection (*β*=0.25, SE=0.06, *t*=3.75, *p*<.001) and an interaction between Selection and linear term (*β*=7.14, SE=0.51, *t*=13.88, *p*<.001), which revealed that there were more looks to the target and the looks to the target gradually increased for those who clicked on the implied actual state. while there was no change in the looks to the target for those who clicked on the suppositional state.

Comparison of Exp 1 vs Exp 2

The pattern for the indicative conditionals in both experiments was as predicted, and similar to the pattern in Orenes et al. (2019): Participants consistently looked at the suppositional state alternative [+ZEBRA, +LION]. However, for the counterfactual conditionals, Exp 2 revealed a different pattern from Exp 1: There were fewer looks to the suppositional state than in indicative conditionals, and the looks to the implied actual state alternative [-ZEBRA, -LION] increased significantly faster in counterfactuals than in indicatives. To predict target preference score, we fitted a model with fixed effects of Time terms, deviation-coded Experiment (Exp 1: -0.5, Exp 2: 0.5), and their interaction. The model also included random intercepts for Participant and Item and random slope of Experiment by Participant and

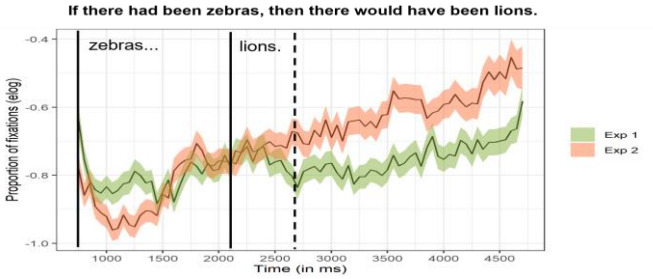

Figure 5: Probabilities of fixations on the implied actual state in Exp 1 (green) vs Exp 2 (orange), showing that an explicit QuD triggered a shift to the implied-state interpretation of counterfactuals.

Item. The critical time window lasted from the onset of the first NP (i.e., ‘zebras’) to 2000ms after the end of the conditional to capture the time until the selection. The model revealed a significant interaction between Experiment and the linear term (β=5.19, SE=0.35, *t*=14.48, *p*<.001), Experiment and the quadratic term (β=-1.43, SE=0.35, *t*=-4.01, *p*<.001), and Experiment and the cubic term (β=-0.82, SE=0.35, *t*=-2.31, *p*<.05). Although not significant, the coefficient of the main effect of Experiment was positive, indicating more looks to the implied actual state in Exp 2 than in Exp 1. The interaction between Experiment and the time terms means the following: (i) The increase in looks to the implied actual state was faster/steeper in Exp 2 than in Exp 1, and (ii) there were more fluctuations in the looks to the actual state in Exp 1 than in Exp 2, as shown in Figure 5.

General Discussion

Our objective was to reverse-engineer the components of counterfactual comprehension, starting with the role of explicit QuD. We replicated and extended a paradigm that, in contrast to most literature, did not rely on causal links between the antecedent and consequent to minimize the influences of causal reasoning.

Our study yielded two main findings. First, unlike Orenes et al. (2019), we found that people did not ever consider the implied actual state (Exp 1), and second, almost half of the participants shifted their interpretation to the implied actual state when they were explicitly asked about it (Exp 2). In other words, making implied actual state interpretation a contextually relevant alternative shifted counterfactual interpretation in even non-causal counterfactuals. Comprehenders, in general, considered suppositional state alternatives when they did not rely on causal inferences, but with the explicit QuD referring to the actual world, consideration of the implied actual state increased. These results are hard to explain for Mental Model Theory (Johnson-Laird & Byrne, 2002): The implied actual state interpretation did not seem to be automatically evoked in the absence of causal link as well as an explicit QuD. However, the results provide support the Suppositional Theory indirectly (Evans & Over, 2004): People suppose a situation where antecedent *p* is true and mentally simulate the situation in which the consequent *q* follows – unless the communicative context requires otherwise. However, it is worth noting that we do not suggest that people do not access ever the implied actual state in interpreting counterfactuals. Instead, we suggest that people make use of linguistic devices (e.g., subjunctive mood) as well as the QuD to determine the specific interpretation of a counterfactual utterance.

The facilitative effect of the availability of the QuD has also been shown in children’s ability to compute scalar implicatures. Kids failed in making adult-like inferences at classical truth-value judgment tasks where the relevant alternatives, i.e., QuD were not made clear (e.g., Katsos & Bishop, 2011): Given ‘I ate some of the cookies’, one might intend to convey ‘some, but not all the cookies’, ‘cookies, but not cake’, or ‘I ate the cookies, but not Jane’. However, when relevant alternatives and the QuD are controlled to be contextually salient, children were able to use it to make inferences in an adult-like manner (e.g., Skordos & Papafragou, 2016). Arguably, we might draw parallels between scalar items and counterfactual conditionals since consideration of alternatives is facilitated by the QuD in both cases.

Two open questions remain. First, in Orenes et al. (2019), almost half of the participants considered the suppositional alternative whereas the other half considered the implied actual state, or both regardless of the absence of an explicit QuD. Since there was no task at the end of the trials related to the critical utterance, it might be the case that participants settled on an alternative without any particular reason. Such groups did not emerge in our replication study. We do not have a clear idea as to why it happened, but it might be attributed to the experimenter effect (Wijenayake, Berkel & Goncalves, 2020). Note that our experiment used a webcam-based eye-tracking due to the pandemic and there was no experimenter to answer potential questions, clarify instructions, or unintentionally motivate participants to behave in line with the experimenters’ preferred study outcome (e.g., Rosenthal, Freidman & Kurland, 1966; Strickland & Suben, 2012). Second, although Exp 2 led to more looks to the implied actual state overall, and almost half of the participants considered only the implied actual state in counterfactual condition, the other half behaved similarly to the group in Exp 1. One reason behind this could be the visual world itself. If we take a closer look at the pattern in Exp 2 by Clicks (Fig. 4), we see that those who clicked on the suppositional state did not ever consider the crossed-out alternatives from very early on. Even before the first NP in the antecedent is heard, they fixated their gaze to non-crossed-out images only (e.g., zebra and lion & elephant and giraffe). However, those who clicked on the implied actual state considered all four alternatives equally first, then settled on the implied actual state, unlike those who disregarded crossed-out alternatives from the beginning. This difference in looks might raise some question marks as to how participants interpreted the crossed-out images, which are unnatural (the negation of [+ZEBRA,+LION] would be a blank page or set of other related wild animals). We aim to address these issues in future experiments.

Acknowledgments

We thank Isabel Orenes for kindly sharing their stimuli, Lea Zaric and Miguel Mejia for helping with stimuli creation, and David Barner and anonymous reviewers for helpful discussion and advice.

References

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Black, J., Williams, D., & Ferguson, H. J. (2018). Imagining counterfactual worlds in autism spectrum disorder. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1444.

Byrne, R. M. (2005). *The rational imagination*. Cambridge, MA: MIT Press.

Dai, H., Kaan, E., & Xu, X. (2021). Understanding counterfactuals in transparent and nontransparent context: An event-related potential investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

de Vega, M., & Urrutia, M. (2012). Discourse updating after reading a counterfactual event. *Psicologica: International Journal of Methodology and Experimental Psychology*, *33*(2), 157–173.

de Vega, M., Urrutia, M., & Riffo, B. (2007). Canceling updating in the comprehension of counterfactuals embedded in narratives. *Memory & Cognition*, *35*(6), 1410–1421.

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395.

Evans, J. S. B. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Psychology Press.

Evans, J. S. B., Over, D. E., & others. (2004). *If: Supposition, pragmatics, and dual processes*. Oxford University Press, USA.

Ferguson, H. J. (2012). Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. *Quarterly Journal of Experimental Psychology*, *65*(5), 939–961.

Ferguson, H. J., Breheny, R., Scheepers, C., & Sanford, A. J. (2009). *Reading the minds of others: Disentangling the gender-specific mechanisms*.

Ferguson, H. J., & Cane, J. E. (2015). Examining the cognitive costs of counterfactual language comprehension: Evidence from ERPs. *Brain Research, 1622*, 252–269.

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, *58*(3), 609–626.

Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008a). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research*, *1236*, 113–125.

Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*, *25*(3), 297–346.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological Science*, *28*(12), 1731-1744.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review 128*(5), 936–975.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*(4), 646.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI publications Stanford, CA.

Kominsky, J. F., Gerstenberg, T., Pelz, M., Sheskin, M., Singmann, H., Schulz, L., & Keil, F. C. (2021). The trajectory of counterfactual simulation in development. *Developmental Psychology*, *57*(2), 253.

Kulakova, E., & Nieuwland, M. S. (2016a). Pragmatic skills predict online counterfactual comprehension: Evidence from the N400. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(5), 814–824.

Kulakova, E., & Nieuwland, M. S. (2016b). Understanding counterfactuality: A review of experimental evidence for the dual meaning of counterfactuals. *Language and Linguistics Compass*, *10*(2), 49–65.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. Com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Mirman, D. (2017). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Nieuwland, M. S. (2013). "If a lion could speak…": Online sensitivity to propositional truth-value of unrealistic counterfactual sentences. *Journal of Memory and Language*, *68*(1), 54–67.

Nieuwland, M. S., & Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, *122*(1), 102–109.

Orenes, I., Garcia-Madruga, J. A., Gomez-Veiga, I., Espino, O., & Byrne, R. M. (2019). The comprehension of counterfactual conditionals: Evidence from eye-tracking in the visual world paradigm. *Frontiers in Psychology*, *10*, 1172.

Papoutsaki, A., Laskey, J., & Huang, J. (2017). Searchgazer: Webcam eye tracking for remote studies of web search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 17-26.

Quelhas, A. C., Rasga, C., & Johnson-Laird, P. N. (2018). The relation between factual and counterfactual conditionals. *Cognitive Science*, *42*(7), 2205–2228.

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. (2004). Context in dynamic interpretation. *The Handbook of Pragmatics*, *197*, 220.

Romoli, J., Santorio, P., & Wittenberg, E. (2019). Fixing De Morgan's laws in counterfactual antecedents. *Proceedings of the 2019 Amsterdam Colloquium*, 347-356.

Romoli, J., Santorio, P., & Wittenberg, E. (2022). Alternatives in counterfactuals: What Is Right and What Is Not, *Journal of Semantics*, ffab023, https://doi.org/10.1093/jos/ffab023

Rosenthal, R., Friedman, N., & Kurland, D. (1966). Instruction-reading behavior of the experimenter as an unintended determinant of experimental results. *Journal of Experimental Research in Personality*.

Santamaría, C., Espino, O., & Byrne, R. M. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1149.

Skordos, D., & Barner, D. (2019). Language comprehension, inference, and alternatives. In *The Oxford Handbook of Experimental Semantics and Pragmatics*.

Skordos, D., & Papafragou, A. (2016). Children’s derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1–2), 3–23.

Stewart, A. J., Haigh, M., & Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *Quarterly Journal of Experimental Psychology*, *62*(11), 2113–2125.

Strickland, B., & Suben, A. (2012). Experimenter philosophy: The problem of experimenter bias in experimental philosophy. *Review of Philosophy and Psychology*, *3*(3), 457–467.

Thompson, V. A., & Byrne, R. M. (2002). Reasoning counterfactually: Making inferences about things that didn’t happen. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1154.

Wijenayake, S., van Berkel, N., Kostakos, V., & Goncalves, J. (2020). Impact of contextual and personal determinants on online social conformity. *Computers in Human Behavior*, *108*, 106302.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832