FILL IN FINAL TITLE

Judith Tonhauser° and Judith Degen•

°University of Stuttgart, Department of Linguistics, Stuttgart, Germany,
judith.tonhauser@ling.uni-stuttgart.de (corresponding author)
•Stanford University, Department of Linguistics, Stanford, USA, jdegen@stanford.edu

April 12, 2025

Abstract

FILL IN ABSTRACT AT THE END

Keywords: Projection inferences, presuppositions, naturalness ratings in explicit ignorance and support contexts.

Acknowledgments:
For helpful comments on this project we thank Craige Roberts, Gregory Scontras, Mandy Simons, the reviewers for L&P, and the audience at the syntax/semantics discussion group at the University of Stuttgart.

Funding: The experiment was funded by Judith Tonhauser's research funds.

Competing interests: The authors have no competing interests to declare.

Ethics approval and informed consent: The experiment was approved by the ethics review committee of the University of Stuttgart. Informed consent was obtained from the participants.

Availability of materials, data, and code: See footnote 6.

Author contributions: Both authors contributed to the conception and design of the experiment. Material preparation, data collection and analysis were performed by both authors. The first draft of the first and second versions of the manuscript were written by Judith Tonhauser. Both authors commented on previous versions of the manuscript. Both authors read and approved the final manuscript.

I don't know if naturalness ratings distinguish presuppositions from nonpresuppositions. Did Mandelkern et al. 2020 discover that they do?

Abstract

Presuppositions have long been taken to be a well-defined class of content that is characterized by typically projecting and being conventionally associated with particular lexical items or constructions (e.g., Heim 1983; van der Sandt 1992; Mandelkern et al. 2020; Beaver et al. 2024). While the strength of the projection inference is often taken to distinguish presuppositions from nonpresuppositions, experimental investigations that used inference tasks suggested that the strength of the projection inference does not distinguish the two (e.g., de Marneffe et al. 2019; Tonhauser et al. 2018; Degen and Tonhauser 2022). Mandelkern et al. 2020 proposed that an alternative measure, namely comparison of naturalness ratings in explicit ignorance and support contexts, distinguishes presuppositions from nonpresuppositions. This paper presents the results of an experiment designed to investigate Mandelkern et al.'s 2020 proposal for the contents of the complements of factive and nonfactive clause-embedding predicates. The results do not support Mandelkern et al.'s 2020 proposal, which means that the question of how presuppositional projection inferences can be distinguished from nonpresuppositional ones is still open.

1 Introduction

Projection inferences are inferences to content that is contributed by an expression in the scope of an entailment-canceling operator (e.g., Potts 2005; Tonhauser et al. 2013, 2018; Degen and Tonhauser 2022). For instance, if an interpreter infers from an utterance of any of the sentences in (1) that Julian dances salsa, this would be considered a projection inference, as the content that Julian dances salsa is contributed by an expression in a polar question:

(1) a. Is my friend Julian, who dances salsa, invited to your party?
b. Does Cole know that Julian dances salsa?
c. Did Cole acknowledge that Julian dances salsa?
d. Does Cole think that Julian dances salsa?

The projection inferences arising from (1a-b) have received specialized analyses: The content of the nonrestrictive relative clause in (1a) is typically analyzed as a conventional implicature, and the content of the clausal complement of *know* in (1b) as a presupposition (see, e.g., Chierchia and McConnell-Ginet 1990; Potts 2005; Tonhauser et al. 2013). Projection inferences arising from utterances of (1c-d), on the other hand, have traditionally not been considered presuppositions. A pressing question for formal analyses of presuppositions is how to distinguish supposed presuppositions, like the content of the clausal complement of (1b), from supposed nonpresuppositions, like (1c-d).

A long-standing assumption is that presuppositions differ from nonpresuppositions in the strength of the projection inference, which is assumed to be stronger with presuppositions than nonpresuppositions. This assumption challenged, however, by experimental research that found that the strength of the projection inference for some supposed nonpresuppositions is as strong or even stronger than that of some supposed presuppositions (e.g., de Marneffe et al. 2019; Degen and Tonhauser 2022). A sample illustration of this result is in Fig. 1, which is based on Degen and Tonhauser's 2022 Exp. 1a, which shows the mean projection ratings of five supposed presuppositions contributed by the clausal complements of factive predicates in orange and five and 15 supposed nonpresuppositions contributed by the clausal complements

of nonfactive predicates in black.[1] As shown, the strength of the projection inference to supposed nonpresupposition contributed by *acknowledge, hear* and *inform* are at least as strong, if not stronger, than the strength of the projection inference to supposed presuppositions. Furthermore, Fig. 1 does not suggest a clear cut-off in projection strength between supposed presuppositions and nonpresuppositions. In sum, the strength of the projection inference does not appear to distinguish presuppositional projection inferences from nonpresuppositional ones.

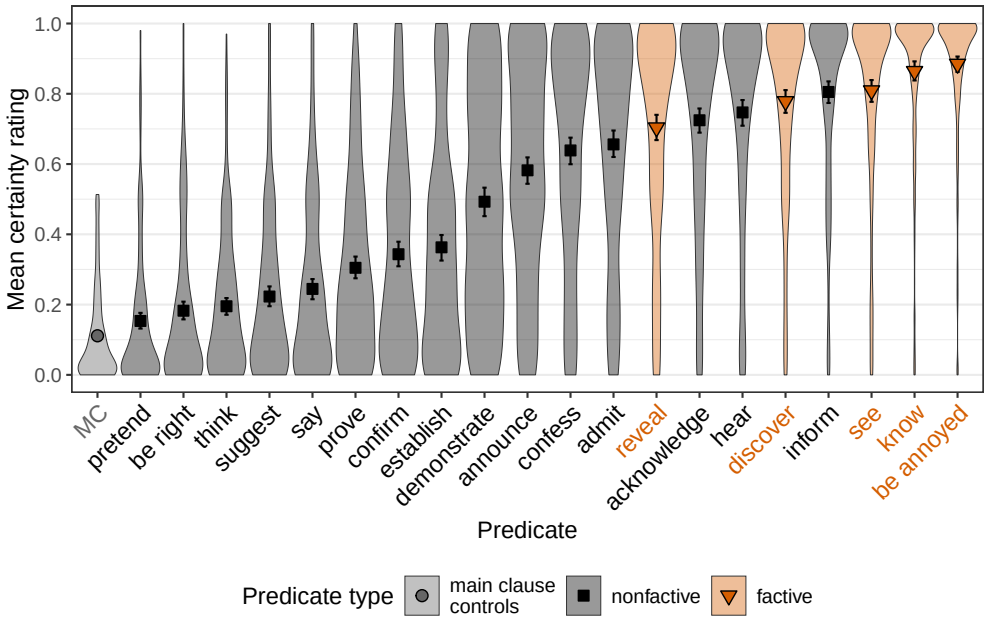


Figure 1: Mean certainty rating (measuring projection) of main clause ('MC') content and the contents of the complements of the 20 factive and nonfactive predicates investigated in Exp. 1a of Degen and Tonhauser 2022. Error bars indicate 95% bootstrapped confidence intervals. Violin plots indicate the kernel probability density of the individual participants' ratings.

An alternative proposal for how to distinguish presuppositional projection inferences from nonpresuppositional ones was made in Mandelkern et al. 2020.[2] Specifically, they proposed (p.497) that "comparing contexts which support the inference to contexts in which it has been made clear that the speaker is ignorant about the inference provides a way to distinguish a broad class of natural and invited pragmatic inferences from those that are really encoded as presuppositions". For example, on this proposal, the presupposition contributed by *stop* in (2c) is expected to receive relatively higher naturalness ratings in the support context in (2b) than in the explicit ignorance context in (2a). On the other hand, the naturalness ratings for the nonpresupposition contributed by *now frowns on* in (2c) are not expected to vary by context. What this measure is taken to distinguish, thus, are projection inferences that are merely "natural and invited pragmatic inferences" and projection inferences that are "encoded as presuppositions, and thus have no choice but to project" (Mandelkern et al. 2020:497)

(2) Mandelkern et al. 2020:490f.

[1]The data from Degen and Tonhauser's 2022 Exp. 1a were accessed here: `https://github.com/judith-tonhauser/projective-probability/tree/master/results/5-projectivity-no-fact`. The script that was used to create Fig. 1 can be found in the GitHub repository linked in footnote 6.

[2]How to distinguish presuppositions from nonpresuppositions is not the main research question of Mandelkern et al. 2020. Rather, this paper is primarily concerned with the question of whether presupposition filtering is symmetric.

a. Explicit ignorance context:
   Mary always was involved in a lot of sports, but I don't know whether she ever did any yoga.
b. Support context:
   Mary always was involved in a lot of sports, and she used to do yoga, too.
c. Sentence with presupposition / nonpresupposition:
   If Mary {has stopped / now frowns on} doing yoga, then Matthew will interview her for his story.

The data reported on in Mandelkern et al. 2020:Exp. 3 supports their proposal, as shown in panel (a) of Fig. 2.[3] This figure shows the mean naturalness ratings, by context, of the four supposed presuppositions and four supposed nonpresuppositions featured in their Exp. 3 in sentences like those in (2c). As shown, the ratings for *continue, be happy, stop* and *be aware* (in orange) are higher in the support context than in the explicit ignorance context, whereas the ratings for *now frown on, be sure, be hoping* and *enjoy* (in black) do not exhibit by-context variation. For Mandelkern et al. 2020, these results support the assumption that the projection inferences associated with *continue, be happy, stop* and *be aware* are presuppositions, while those of *now frown on, be sure, be hoping* and *enjoy* are not.

However, data reported on in Kalomoiros and Schwarz 2024 does not support Mandelkern et al.'s 2020 proposal as clearly.[4] Panels (b) and (c) of Fig. 2 show the mean naturalness ratings, by context, of the expressions featured in their Exps. 1 and 2, respectively.[5] (These experiments did not include nonpresuppositional projection inferences.) As shown in panel (b), the supposed presuppositions of *again, continue* and *find out* appear to be sensitive to the context manipulation, but those of *be aware, stop* and *be happy* do not, contrary to expectation. As shown in panel (c), the relevant data from their Exp. 2 do support Mandelkern et al.'s 2020 proposal for the supposed presuppositions of *continue, again* and *stop*. The experiments in Mandelkern et al. 2020 and Kalomoiros and Schwarz 2024 differed on a number of dimensions, including whether nonpresuppositional projection inferences were included, the number and type of presuppositional projection inferences investigated, and whether the target stimuli investigated presupposition filtering in conjunctions and disjunctions, or only in conjunctions. It is an open question which factors are responsible for the differences in the results for sentences like (2c). It is therefore also an open question whether comparison of naturalness ratings in explicit ignorance and support contexts distinguishes presuppositional projection inferences from nonpresuppositional ones, as proposed in Mandelkern et al. 2020.

This paper presents the results of an experiment designed to investigate Mandelkern et al.'s 2020 proposal that comparison of naturalness ratings in explicit ignorance and support contexts distinguishes presuppositional from nonpresuppositional projection inferences.[6] To allow for comparison with the results of Degen and Tonhauser 2022, the contents of the complements of the same 20 (non)factive clause-embedding predicates (see Fig. 1) were investigated. Our experiment slightly deviated from design of Mandelkern et al.'s 2020 Exp. 3 to also allow for comparison with the result of Degen and Tonhauser 2021 that projection is sensitive to prior probability. Specifically, what Degen and Tonhauser 2021 found was that the higher the prior probability of a content, the stronger the projection inference. For instance, a speaker's commitment to the CC of (3c), that Julian dances salsa, was stronger in the context in (3b), where the CC has a higher prior probability, than in the context in (3a), where the CC has a lower prior probability.

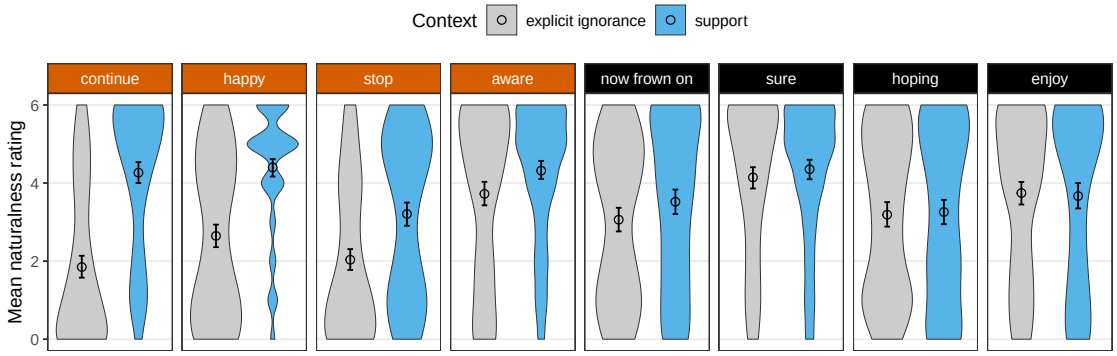
(3) a. Julian is German. [lower prior probability]

[3]The data from Mandelkern et al. 2020:Exp. 3 are available at `https://osf.io/2b9m7/`. The R script that generates the figures in panel (a) of Fig. 2 and panel (a) of Fig.3 is in `NEEDTOFILLINLINKATTHEEND`.

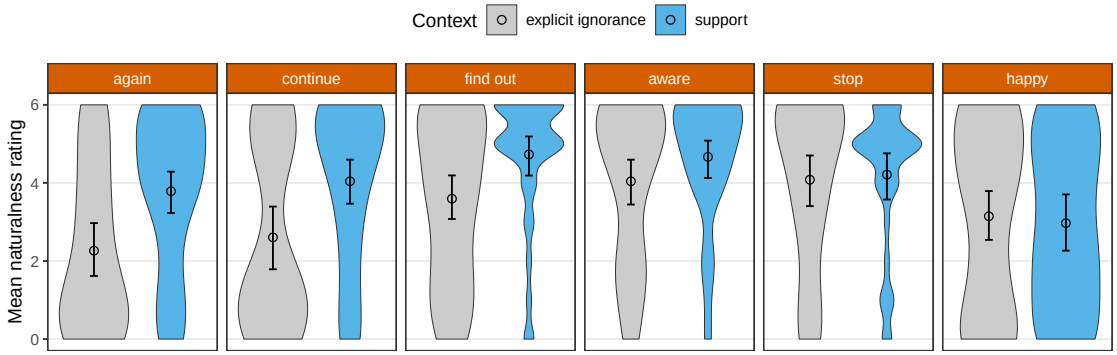
[4]As with Mandelkern et al. 2020, the distinction between presuppositions and nonpresuppositions was not the main research question of Kalomoiros and Schwarz 2024, but rather the question of whether presupposition filtering is symmetric.

[5]The data from Kalomoiros and Schwarz 2024 are available at `https://osf.io/3p68r/`. The R script that generates the figure in panels (b) and (c) of Fig. 2 and panel (b) and (c) of Fig.3 is in `NEEDTOFILLINLINKATTHEEND`.

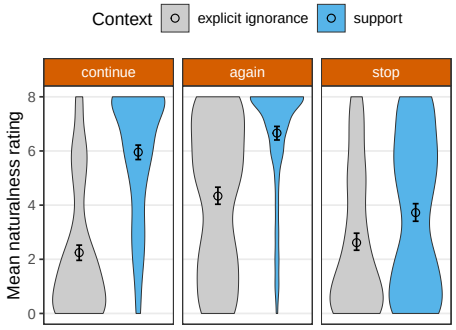
[6]The experiment was approved by the ethics review committee of [university name suppressed]. The experiment, materials, data, and analysis scripts can be accessed here: `https://anonymous.4open.science/r/naturalness-in-EIC-5847`.



(a) Mandelkern et al. 2020:Exp. 3



(b) Kalomoiros and Schwarz 2024:Exp. 1



(c) Kalomoiros and Schwarz 2024:Exp. 2

Figure 2: Mean naturalness rating in explicit ignorance and support contexts for the (a) eight presumed presuppositions and nonpresuppositions investigated in Mandelkern et al. 2020:Exp. 3, (b) six presumed presuppositions investigated in Kalomoiros and Schwarz 2024:Exp. 1, and (c) three presumed presuppositions investigated in Kalomoiros and Schwarz 2024:Exp. 2. Presumed presuppositions given in orange, presumed nonpresuppositions in black. Expressions associated with the (non)presuppositions ordered from left to right by decreasing difference between naturalness rating means in the two contexts. Error bars indicate 95% bootstrapped confidence intervals. Violin plots indicate the kernel probability density of the individual participants' ratings.

b. Julian is Cuban. [higher prior probability]

c. Did Cole discover that Julian dances salsa?

Our experiment replaced Mandelkern et al.'s 2020 support context (which entailed the relevant content) with two support contexts that differed in the prior probability of the CC, namely a lower and a higher prior probability context.[7]

Our experiment also included nontarget stimuli with *stop, continue, again, too, also* and an *it*-cleft, which were presented only in explicit ignorance contexts. The inclusion of these stimuli was motivated by the fact that some research on presuppositions assumes that presuppositions exhibit variation in their acceptability in explicit ignorance contexts. Specifically, Simons 2001 and Abusch 2010 assumed that acceptability in explicit ignorance contexts distinguishes nondefeasible (or: hard) presuppositions from defeasible (or: soft) ones. For instance, the presuppositions associated with *stop*, *discover* and *win* were assumed to be defeasible because they were judged to be acceptable in explicit ignorance contexts, as shown in (4a-c), respectively. By contrast, the presuppositions associated with *again*, the *it*-cleft and *too* were assumed to be nondefeasible because they were judged to be unacceptable in explicit ignorance contexts, as shown in (5a-c).

(4) a. I have no idea whether Jane ever smoked, but she hasn't stopped smoking. (Simons 2001:443)

b. Context: "two people [...] know that Henry is searching for Jane, but who don't themselves know where Jane is:"
If Henry discovers that Jane is in New York, there'll be trouble. (Simons 2001:434)

c. I have no idea whether John ended up participating in the Road Race yesterday. But if he won it, then he has more victories than anyone else in history. (Abusch 2010:39)

(5) a. #I don't know if Jane ever rented "Manhattan" before, but perhaps she's renting it again. (Simons 2001:443)

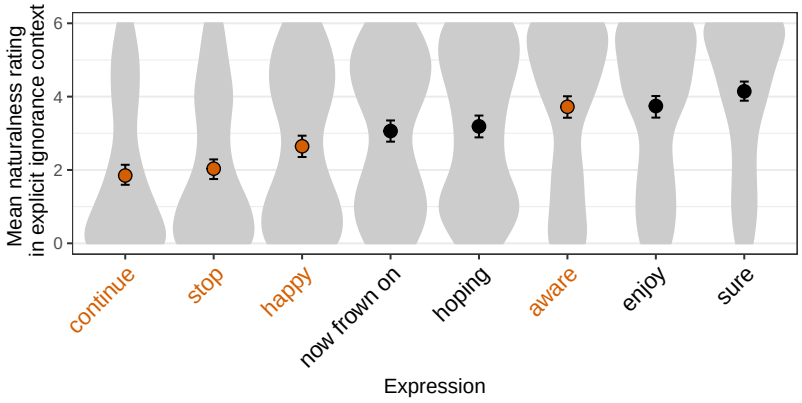
b. ??I have no idea whether anyone read that letter. But if it is John who read it, let's ask him to be discreet about the content. (Abusch 2010:40)

c. ??I have no idea whether John read that proposal. But if Bill read it too, let's ask them to confer and simply give us a yes-no response. (Abusch 2010:40)

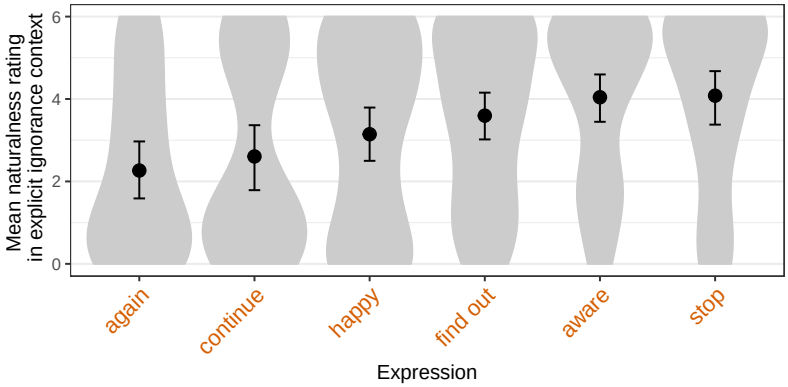
Panels (a) to (c) of Fig. 3 plots the mean naturalness ratings of the (non)presuppositions investigated in Mandelkern et al. 2020:Exp. 3, and Exps. 1 and 2 of Kalomoiros and Schwarz 2024. As shown, the results of these experiments support the assumption of Simons 2001 and Abusch 2010 that presuppositions vary in their acceptability in explicit ignorance contexts. Furthermore, some of the results agree with the empirical claims of Simons 2001 and Abusch 2010. For instance, the mean naturalness rating of *again* is quite low in Kalomoiros and Schwarz's 2024 Exp. 1 in panel (b), while that of *stop* is quite high, in line with Abusch's 2010 claim that the presupposition of *again* is nondefeasible and Simons's 2001 claim that the presupposition of *stop* is defeasible. The results of Kalomoiros and Schwarz's 2024 Exp. 2 in panel (c), however, show the opposite pattern for *stop* and *again*. It therefore remains an open question whether naturalness ratings of supposed presuppositions in explicit ignorance contexts distinguish nondefeasible and defeasible presuppositions.[8]

[7] Another change was that participants gave their naturalness ratings on a slider with endpoints labeled 'totally unnatural' and 'totally natural' rather than on a 7-point Likert scale with endpoints labeled 'completely unnatural' and 'completely natural' (as in Mandelkern et al. 2020:Exp.3 and Kalomoiros and Schwarz 2024:Exp. 1) or on a 9-point Likert scale with the aforementioned endpoints (Kalomoiros and Schwarz 2024:Exp. 2).

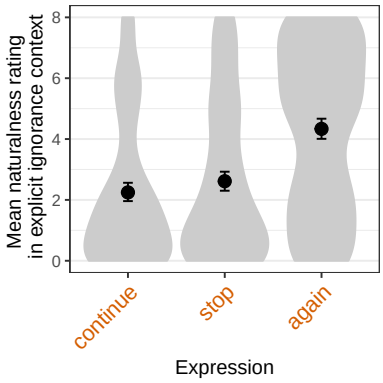
[8] Roberts and Simons 2024:734ff. suggested that *continue*, unlike *stop*, contributes an anaphoric contextual requirement. This hypothesis goes well with the observation that the naturalness ratings for *continue* are very low across all of the data shown in Fig. 3.



(a) Mandelkern et al. 2020:Exp. 3



(b) Kalomoiros and Schwarz 2024:Exp. 1



(c) Kalomoiros and Schwarz 2024:Exp. 2

Figure 3: Mean naturalness rating in explicit ignorance for the (a) eight presumed presuppositions and non-presuppositions investigated in Mandelkern et al. 2020:Exp. 3, (b) six presumed presuppositions investigated in Kalomoiros and Schwarz 2024:Exp. 1, and (c) three presumed presuppositions investigated in Kalomoiros and Schwarz 2024:Exp.2. Presumed presuppositions given in orange, nonpresuppositions in black. Error bars indicate 95% bootstrapped confidence intervals. Expressions ordered from left to right by increasing mean naturalness rating. Violin plots indicate the kernel probability density of the individual participants' ratings.

2 Experiment

To investigate Mandelkern et al.'s 2020 proposal that comparison of naturalness ratings in explicit ignorance and support contexts distinguishes presuppositional projection inferences from nonpresuppositional ones, participants read two-sentence discourses consisting of a declarative (which provided the context) and an interrogative (which realized a factive or nonfactive predicate), and rated the naturalness of the interrogative in the context of the declarative.

2.1 Methods

2.1.1 Participants

We recruited 425 participants on Prolific. Due to a programming error, the data of only 398 participants was recorded (ages: 19-73, mean: 40.8; 187 women, 201 men, 8 non-categorical, 2 preferred to not disclose).

The recruited participants were required to live in the USA, to speak English as their first language, to have completed at least 100 tasks, and to have an approval rating of at least 99%. The median time spent on the task was 6:24 minutes. Participants were paid $1.78, corresponding to an hourly pay of $16.6.

2.1.2 Materials

Participants read two-sentence discourses consisting of a declarative followed by an interrogative, as shown in (6). In the target stimuli, the interrogatives combined the 20 (non)factive predicates of Degen and Tonhauser 2022 (see Fig. 1) with the 20 complement clauses of Degen and Tonhauser 2022, for a total of 400 interrogatives. The preceding declaratives implemented a three-level context condition: In the 'explicit ignorance' context (6a), the declarative sentence conveyed the speaker's ignorance about the CC. In the 'lower prior probability' context (6b), the CC had a comparatively lower prior probability, whereas it had a comparatively higher prior probability in the 'higher prior probability' context (6c). The two contexts for each of the 20 complement clauses were normed in Degen and Tonhauser 2021. See Supplement A for the full set of 20 complement clauses and the two contexts for each complement clause.

(6) a. Explicit ignorance context:
I have no idea if Julian dances salsa. Did Cole discover that Julian dances salsa?

b. Lower prior probability context:
Julian is German. Did Cole discover that Julian dances salsa?

c. Higher prior probability context:
Julian is Cuban. Did Cole discover that Julian dances salsa?'

In addition to the 1,200 target stimuli (400 interrogatives × 3 contexts), the materials also included the six nontarget stimuli in (7). The interrogatives of these stimuli featured six expressions typically analyzed as presupposition triggers, namely *stop, continue, again, too, also*, and an *it*-cleft. The preceding declaratives conveyed the speaker's explicit ignorance about presuppositions associated with these expressions, namely the pre-state contents of *stop* and *continue*, the prejacent of *again*, the content that there is an alternative true proposition of *too* and *also*, and the existential content of the *it*-cleft.[9] As noted in §1, Simons 2001 took *stop* to be a soft trigger and *again* a hard trigger; Abusch 2010 assumed that *too* and *it*-clefts are hard triggers. The nontarget stimuli were included for comparison to the target stimuli, to the claims of Simons 2001 and Abusch 2010, and to the results of Mandelkern et al. 2020:Exp. 3 and Kalomoiros and Schwarz 2024:Exps. 1-2, where *continue, stop* and *again* were target expressions.

(7) a. I don't know if Stephen was ever in the habit of vaping. Has Stephen recently stopped vaping?

b. I don't know if John was ever reading "Dune". Has John recently continued reading "Dune"?

c. I don't know if William was ever interested in history. Is William interested in history again?"

d. I don't know if Ann plays any instrument. Does Ann play the flute, too?

e. I don't know if Svenja plays any sport. Does Svenja also play soccer?

f. I don't know if anyone was playing outside with the kids. Was it Jack who was playing outside with the kids?

The experiment also included 4 filler stimuli that were used to exclude the data of participants not attending to the task (see Supplement B). The filler stimuli, which also consisted of a declarative followed by an interrogative, were expected to receive high naturalness ratings.

[9]The contents investigated for *too* and *also* in (7d) and (7e) are that Ann plays an instrument other than the flute and that Svenja plays a sport other than soccer. This interpretation arises if *too* and *also* associate with *the flute* and *soccer*, respectively. While prosody was not controlled for, this focus association is made plausible by the preceding explicit ignorance statements, which evoke the question of whether Ann plays any instrument and whether Svenja plays any sport, respectively.

A random set of 30 stimuli was created for each participant. Each set contained 20 target stimuli in which each of the 20 complement clauses was paired with a unique clause-embedding predicate. Twelve of the target stimuli were presented in the explicit ignorance context, and the other eight in a low or a higher prior probability context (four each). Each participants' set also contained the same six nontarget stimuli and the same four filler stimuli. Each of the 30 stimuli were presented as utterances by a unique named speaker. Trial order was randomized for each participant.

2.1.3 Procedure

Participants were instructed to rate how natural the question sounds in the context of the statement. As shown in Figure 4, they were asked give their rating on a slider from 'totally unnatural' (coded as 0) to 'totally natural' (coded as 1). The experiment began with four practice trials to familiarize participants with the task (see Supplement B for details). After rating the 30 trials, participants filled out a short optional demographic survey. To encourage truthful responses, participants were told that they would be paid no matter what answers they gave in the survey.

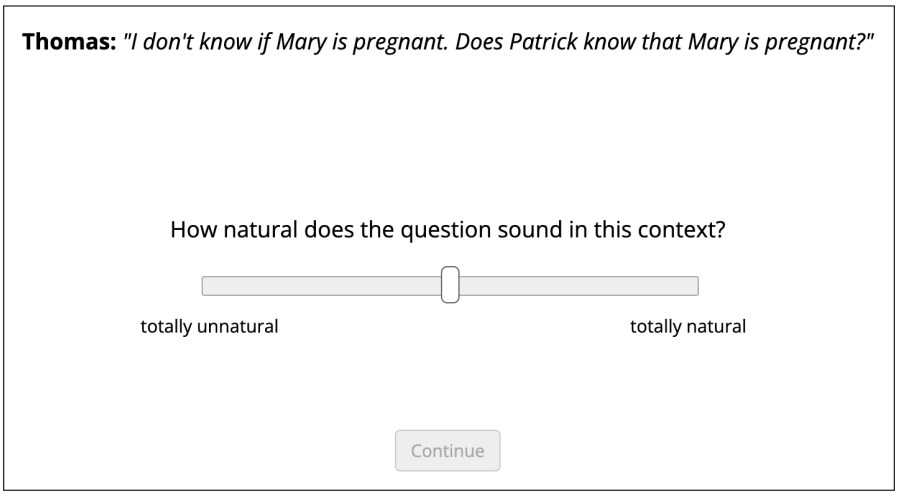


Figure 4: Sample trial with explicit ignorance context.

2.1.4 Data exclusion

We excluded the data of 5 participants who did not self-identify as native speakers of American English and that of 23 participants whose mean response to the fillers (expected to be natural) was more than 2 sd below the group mean.[10] The data of 370 participants entered into the analysis (ages: 19-80, mean: 40.7; 175 women, 185 men, 8 nonbinary, 2 did not disclose). The 20 predicates each received at least 200 ratings in the explicit ignorance context (mean: 222 ratings), at least 59 ratings in the lower prior probability context (mean: 74), and at least 61 ratings in the higher prior probability context (mean: 74). The six nontargets received 370 ratings each.

2.2 Results

We first investigate Mandelkern et al.'s 2020 proposal that comparison of naturalness ratings in explicit ignorance and support contexts distinguishes presuppositional projection inferences from nonpresuppositional

[10]Contrary to what was planned, one of the four filler stimuli was not used to exclude participants' data because it had a mean rating of only .5. See Supplement B for details.

ones (§2.2.1). We then address the question of how projection inferences behave in explicit ignorance contexts and, specifically, whether naturalness ratings in such contexts distinguish nondefeasible and defeasible presuppositions, as proposed in Simons 2001 and Abusch 2010 (§2.2.2).

2.2.1 Comparison of naturalness ratings in explicit ignorance and support contexts

Does comparison of naturalness ratings of in explicit ignorance and support contexts distinguish between presuppositional projection inferences and nonpresuppositional ones, as proposed in Mandelkern et al. 2020? Fig. 5 shows mean naturalness ratings for the CCs of the 20 factive and nonfactive predicates in the three contexts featured in our experiment with predicates ordered left-to-right (and top-to-bottom) by decreasing difference in the ratings in the higher prior probability and explicit ignorance contexts. The figure also includes the results of the statistical analysis, namely posthoc pairwise comparisons of the estimated means for each context using the 'emmeans' package (Lenth 2023); a separate comparison was conducted for each of the 20 predicates. The input to the pairwise comparisons were 20 Bayesian mixed-effects beta regression models with weakly informative priors that were fit using the 'brms' package (Bürkner 2017). The models predicted naturalness ratings[11] from a fixed effect of context (with treatment coding and 'explicit ignorance' as reference level) and included a random by-item intercept (where an item is a complement clause) and a random by-item slope for context.[12] The output of the pairwise comparison for each expression was the 95% HDI of estimated marginal mean differences between each of the three contexts. We assume that two contexts differ if their HDI does not include 0. In Fig. 5, a solid line spanning two contexts indicates that the ratings in the two contexts differ.[13]

As shown, the effect of context is not uniform for the projection inferences of the purported factive predicates. For *be annoyed, know* and *discover*, the mean naturalness rating is lower in the explicit ignorance context than in the higher prior probability context, but this is not the case for *see* and *reveal*, which pattern like several nonfactive predicates in this regard. Thus, on Mandelkern et al.'s 2020 proposal, only inferences to the contents of the complements of *be annoyed, know* and *discover* would be considered presuppositional projection inferences; these three predicates, which include an emotive predicate, a cognitive stative predicate, and a cognitive change-of-state predicate, do not necessarily constitute a natural class (see, e.g., Karttunen 2016; Djärv 2019). Inferences to the contents of the complements of *see* and *reveal* would be considered nonpresuppositional projection inferences, like those to the contents of the complements of communication predicates like *inform* and *announce* (for discussion, see Anand and Hacquard 2014).[14] Alternatively, the results suggest that comparison of naturalness ratings in explicit ignorance and support contexts does not reliably distinguish presuppositional projection inferences from nonpresuppositional ones. Importantly, the difference between the higher prior probability and explicit ignorance contexts for *be annoyed* and *know* are driven primarily by the low naturalness ratings in the explicit ignorance contexts, a point we return to in §2.2.3.

We also observe that the prior probability of content modulates naturalness ratings with all predicates, except *confess, demonstrate, pretend* and *confirm*, such that content is judged as more natural in a context in

[11] To model the ratings using a beta regression, the ratings were first transformed from the interval [0,1] to the interval (0,1) using the method proposed in Smithson and Verkuilen 2006.

[12] By-participant random effects were not included because each participant saw each predicate only once and in only one context condition.

[13] The full model output is available here: `https://anonymous.4open.science/r/naturalness-in-EIC-5847/results/main/models/analysis2/fullModelOutput/fullModelOutput.pdf`.

[14] Our five purportedly factive predicates received ratings in the higher prior probability support context that are roughly comparable to the ratings that the factive predicates of Mandelkern et al. 2020 received in the support context. Specifically, the mean naturalness ratings in the higher prior probability support context ranged from .64 (*reveal*) to .78 (*know*) on a scale from 0 to 1; the mean naturalness ratings in the support context of Mandelkern et al. 2020 were 4.32 for *be aware* and 4.4 for *be happy* on a scale from 1 to 7.

which it has a higher prior probability than in a context in which it has a lower prior probability. This result dovetails with the result of Degen and Tonhauser 2021 that projection is modulated by prior probability. Thus, a methodological implication of our experiment is that future investigations of projection inferences using naturalness ratings will need to consider the effect of the prior probability of the inference.

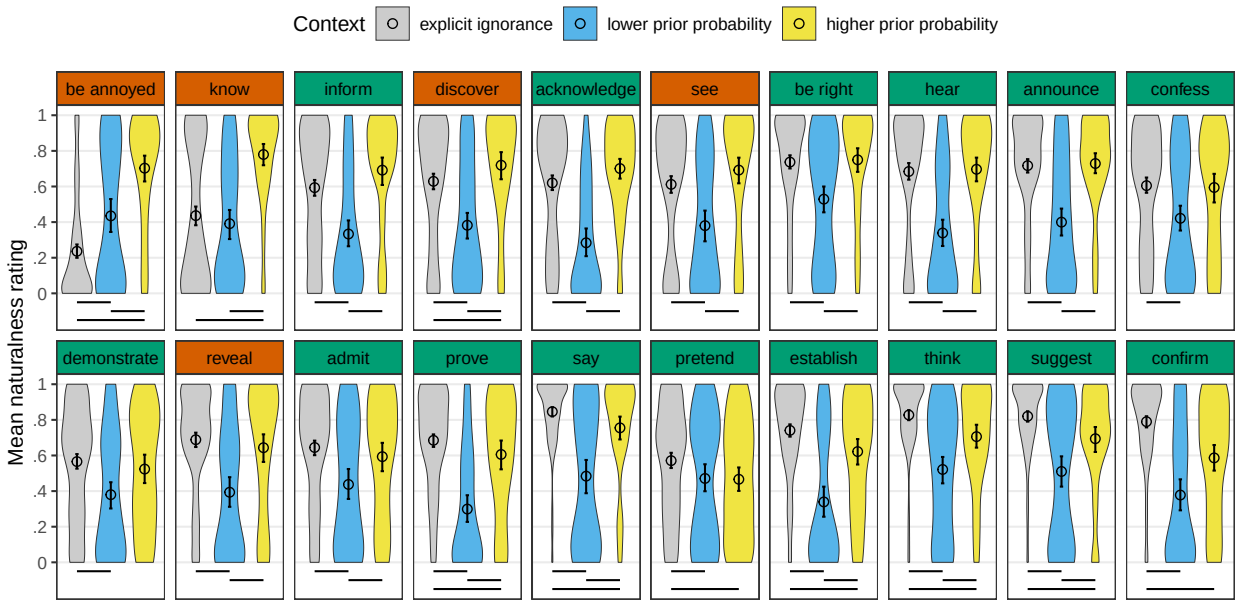


Figure 5: Mean naturalness rating by predicate (factive, nonfactive) and context, with predicates ordered from left to right (and top to bottom) by decreasing difference between the mean rating in the higher prior probability support context and the explicit ignorance context. Error bars indicate 95% bootstrapped CIs. Violin plots indicate kernel probability density of participants' ratings. Below each facet, a solid line spanning two contexts indicates that the 95% HDI of estimated marginal mean differences between the two contexts does not include 0, that is, that the ratings in the two contexts differ.

2.2.2 Naturalness ratings of projection inferences in explicit ignorance contexts

Recall that Simons 2001 and Abusch 2010 assumed that presuppositions vary in their acceptability in explicit ignorance contexts and that, specifically, nondefeasible presuppositions are unacceptable in such contexts and defeasible ones acceptable. Figure 6 shows mean naturalness ratings in the explicit ignorance context by expression (in distinct colors: factive predicates, nonfactive predicates, nontarget). Consistent with Simons's 2001 and Abusch's 2010 assumption, there is variation in the mean naturalness ratings of the projection inferences investigated. Looking first at the six nontarget stimuli (for which there was only one item each), the naturalness means for *continue, too, also* and *again* are at floor, which would suggest, given Simons's 2001 and Abusch's 2010 assumption, that the projection inferences associated with these expressions are not defeasible, that is, hard (see also Roberts and Simons 2024:734ff. for *continue*). This result is in line with Simons's 2001 assumption for *again* and Abusch's 2010 assumption for *too*. Recall, however, from §1 that in Kalomoiros and Schwarz 2024:Exp. 2 the mean naturalness rating for *again* was much higher than that those of *continue* and *stop*.

The results for the two items with *stop* and the *it*-cleft are less clear: The mean naturalness ratings are higher than those of *continue, too, also* and *again*, but not as high as for some of the other contents investigated. These results do not support the assumption that naturalness ratings in explicit ignorance

contexts categorically distinguish between defeasible (soft) and nondefeasible (hard) projection inferences.

The results of our experiment further complicate the picture for *stop*, whose presupposition was taken to be defeasible in Simons 2001, but whose mean rating is lower than the mean rating for the *it*-cleft, whose presupposition was taken to be nondefeasible in Abusch 2010.[15] Recall from §1 that the pre-state content of *stop* received quite low naturalness ratings in Mandelkern et al. 2020:Exp. 3 and Kalomoiros and Schwarz 2024:Exp. 2, but not in Kalomoiros and Schwarz 2024:Exp. 1. These observations suggest that it is still an open question of how defeasible the pre-state content of *stop* is (see also Abrusán 2016:192f. for considerations along these lines).

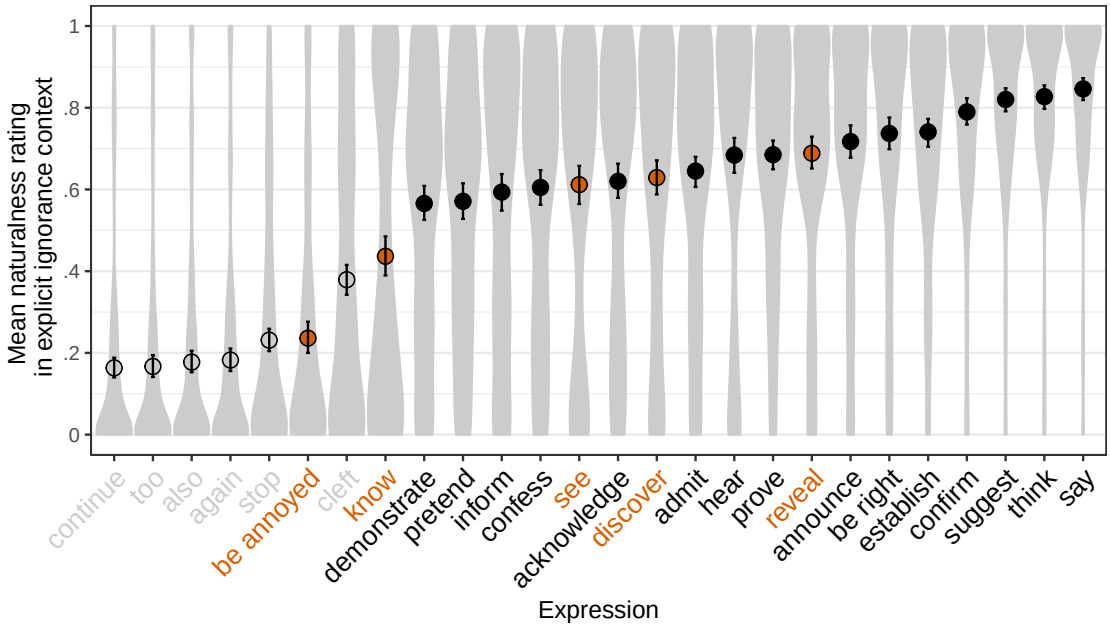


Figure 6: Mean naturalness rating in explicit ignorance context by expression (factive, nonfactive, nontarget). Error bars indicate 95% bootstrapped confidence intervals. Overlaid violin plots indicate the kernel probability density of the participants' ratings.

The naturalness ratings for the five purportedly factive predicates in explicit ignorance contexts exhibit a heterogeneous pattern, with the mean for *be annoyed* as low as that of *stop*, that of *know* higher but not as high as that of *see, discover* and *reveal*. These means suggest that the projection inferences associated with *see, discover* and *reveal* are defeasible, in line with Simons's 2001 assumption about *discover*. As already noted in §2.2.1, the means for *be annoyed* and *know* are quite low, a point we return to in §2.2.3.

These observations were confirmed by a posthoc pairwise comparison of the estimated means for each target and nontarget expression in the explicit ignorance context using the 'emmeans' package (Lenth 2023) in R (R Core Team 2016). The input to the pairwise comparison was a mixed-effects beta regression model with weakly informative priors that was fit using the 'brms' package (Bürkner 2017). The model predicted naturalness ratings[16] from a fixed effect of expression (with treatment coding and 'continue' as reference level) and included random by-participant and by-item intercepts (where an item is a complement clause). The output of the pairwise comparison were 95% highest density intervals (HDIs) of estimated marginal mean differences between each of the expressions. We assume that two expressions differ in their natural-

[15] Smith and Hall 2011 also observed that presupposition of the *it*-cleft did not pattern like a nondefeasible inference. Jayez et al. 2015 found that French *aussi* 'too', *regretter* 'regret' and clefts are not entirely unnatural in explicit ignorance contexts.

[16] As in the model reported in the previous section, the ratings were transformed to the interval (0,1); see footnote 11.

ness in the explicit ignorance context if their HDI does not include 0. As shown in Table 1,[17] the pairwise comparison suggests that the naturalness ratings are lowest for *continue, too, also* and *again*, slightly higher for *stop* and *be annoyed*, slightly higher again for the *it*-cleft and *know*, and higher yet again for *see, discover* and *reveal*. Overall, the results suggest that projection inferences vary in how defeasible they are in explicit ignorance contexts, but the observed variation does not suggest a binary, categorical distinction in defeasibility.

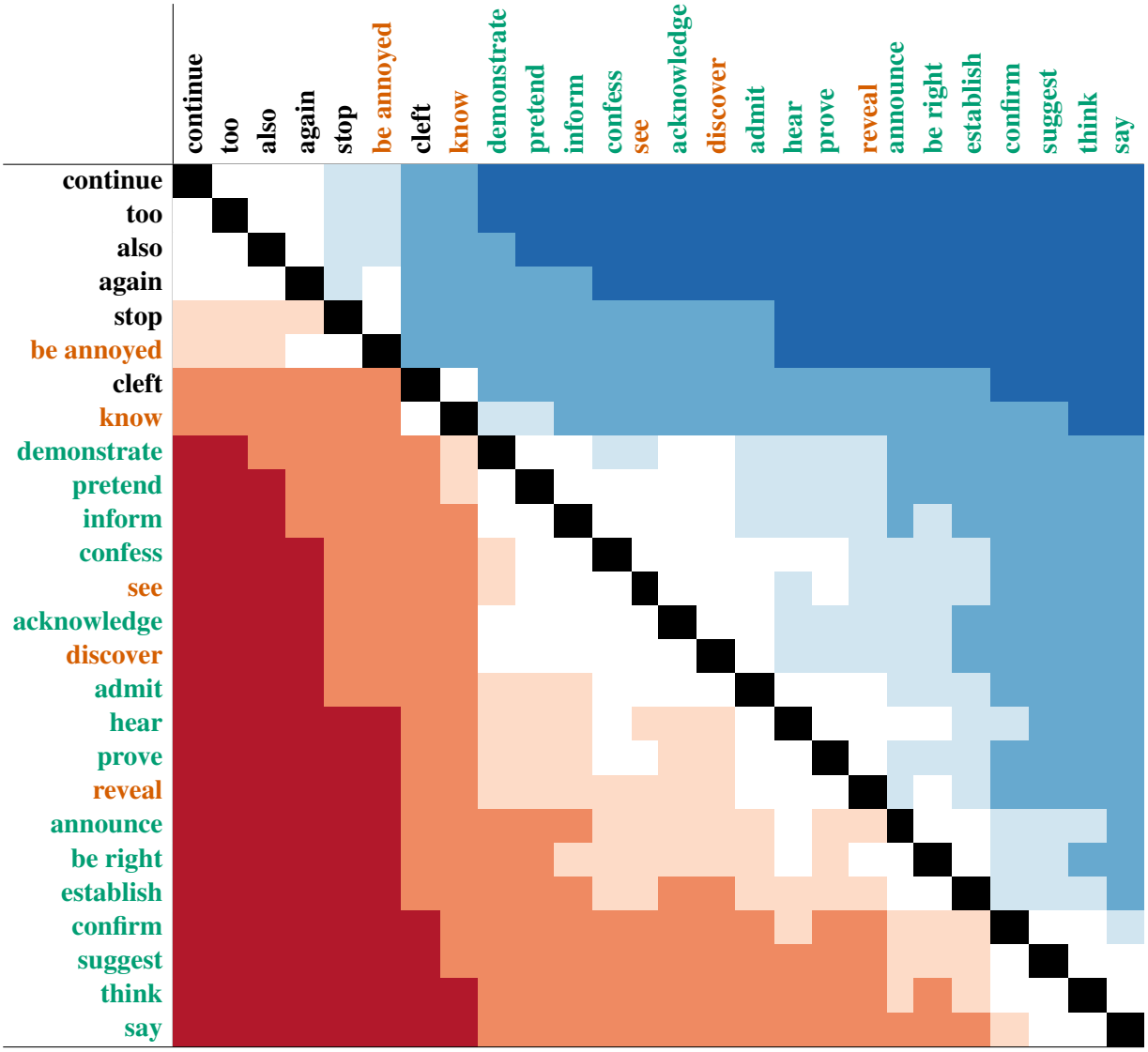


Table 1: Pairwise differences between expressions in the explicit ignorance context. The color coding indicates whether the difference between the means of row expressions and column expressions was positive or negative, as well as the size of the difference. A white cell means that the 95% HDI included 0.
Positive difference: >= 1.5, (1.5, 0.5], (0.5, 0]
Negative difference: >= −1.5, (−1.5, −0.5], (−0.5, 0]

[17]While the naturalness ratings collected from the participants range from 0 to 1, the greater than |1| estimated marginal mean differences are the result of the beta regression using a logit link function for the mean parameter. The full model output is available here: `https://anonymous.4open.science/r/naturalness-in-EIC-5847/results/main/models/analysis1/fullModelOutput/fullModelOutput.pdf`.

2.2.3 Discussion

The analysis in §2.2.1 investigated Mandelkern et al.'s 2020 proposal that comparison of naturalness ratings in explicit ignorance and support contexts distinguishes presuppositional projection inferences from nonpresuppositional ones. What we found was that only the contents of the complements of *be annoyed, know* and *discover* were identified as presuppositional projection inferences, not those of *see* and *reveal*, which patterned like nonpresuppositional projection inferences. This result could be taken to mean that the contents of the clausal complements of *see* and *reveal* are not presuppositions or that the measure proposed by Mandelkern et al. 2020 does not reliably distinguish presuppositional projection inferences from nonpresuppositional ones. This latter conclusion is further supported by the observation in §1 that the projection inferences of *be aware, stop* and *be happy* were also not identified as presuppositional projection inferences in Kalomoiros and Schwarz's 2024 Exp. 1. An alternative position, which we engage with in the General Discussion in §3, is that presuppositions are not a well-defined subset of projection inferences.

As pointed out in §2.2.1, the difference in the naturalness mean in the explicit ignorance and higher prior probability support context for *be annoyed* and *know* was primarily due to the mean rating being low in the explicit ignorance context. As shown in §2.2.2, the low mean naturalness ratings of these two predicates stand out from among those of the other 18 predicates. In the following, we consider the question of why these two predicates might have received such low ratings in explicit ignorance contexts.

For *be annoyed*, it is worth emphasizing that the projection inference investigated was the inference to the truth of the content of the complement, just like for the other clause-embedding predicates. In the explicit ignorance condition of the experiment, participants rated whether a question like that in (8) sounds natural in the context of the explicit ignorance statement.

(8) Sandy: I don't know if Julian dances salsa. Is Sam annoyed that Julian dances salsa?

As shown by the violin plot in Fig. 6, participants' ratings were quite low overall, suggesting that the question was judged to be quite unnatural. This is puzzling because emotive predicates like *be annoyed* do not require that the speaker (here, Sandy) is committed to the truth of the content of the complement; rather, emotive predicates merely require that the attitude holder (here, Sam) is committed to the truth of the content of the complement (e.g., Heim 1992; Karttunen 2016; Djärv 2019). (9), for instance, is acceptable even though the speaker is not committed to the truth of Sue being able to stay in bed.

(9) Mary, who was under the illusion that it was Sunday, was glad that she could stay in bed.
(Klein 1975, as cited in Gazdar 1979:122)

Karttunen 2016 suggested that the inference to speaker commitment is a generalized conversational implicature that comes about in contexts that support the "default assumption" that the commitments of the speaker and the attitude holder are aligned (p.712). This analysis would, however, lead us to expect that *be annoyed* is rated as natural in contexts in explicit ignorance contexts because the inference that the attitude holder is committed to the truth of the content of the complement does not conflict with the explicitly stated speaker ignorance about the truth of the content of the complement. Why, then, did *be annoyed* receive such a low mean naturalness rating in explicit ignorance contexts?

One hypothesis is that naturalness ratings in explicit ignorance contexts are sensitive not just to whether the projection inference is defeasibile but also to its at-issueness. On this hypothesis, the explicit ignorance statement (e.g., *I don't know whether Julian dances salsa*) does not merely convey the speaker's ignorance about the content c to be investigated, but also identifies that the speaker takes the question under discussion (QUD) to be whether c is true (?c e.g., *Does Julian dances salsa?*). The speaker's immediately following interrogative utterance is interpreted with respect to the QUD ?c. The naturalness of the interrogative is then modulated by whether the QUD ?c and the interpretation of the interrogative constitute a felicitous strategy of inquiry (Roberts 2012:32f.), that is, by whether the interrogative can be interpreted as inquiring

about *c*. As shown in Tonhauser et al. 2018, projection inferences differ in at-issueness. The inference to the content of the complements of emotive predicates, including *be annoyed*, were particularly resistant to being interpreted as at-issue. Thus, an interrogative with *be annoyed* (e.g., *Is Cole annoyed that Julian dances salsa?*) is unlikely to be interpreted as congruent with a QUD about the content of the complement (e.g., *Does Julian dance salsa?*). This incongruence may be why *be annoyed* received quite low naturalness ratings in the explicit ignorance context. This hypothesis is in line with the results of Mandelkern et al. 2020:Exp. 3 and Kalomoiros and Schwarz 2024:Exp. 1, shown in Fig. 2, where the mean naturalness ratings for the emotive predicate *be happy* were (at least numerically) lower than for the cognitive predicate *be aware*. However, if this hypothesis is on the right track, it also means that the low naturalness mean of *be annoyed* is not necessarily connected to the defeasibility of the projection inference, but rather to its not-at-issueness. This, in turn, may call into question the assumptions behind Mandelkern et al.'s 2020 proposal that comparison of naturalness ratings in explicit ignorance and support contexts distinguish "encoded" presuppositions that "have no choice but to project" from "natural and invited pragmatic infereces" (p.497).

For *know*, the violin plot in Fig. 6 shows that participants' naturalness ratings exhibit a bimodal pattern, with about half of the participants rating *know* as relatively unnatural in the explicit ignorance context and about half as relatively natural. For the first group of participants, one could assume either that they reacted to the contradiction between the speaker's explicit ignorance statement and the projection inference. Under this hypothesis, the projection inference is not defeasible for this group of participants. Alternatively, one might assume that for these participants the projection inference is not at-issue, and therefore in conflict with the implicit QUD, as suggested for *be annoyed* above.

For the second group of participants, those who judged speakers' utterances like (10) as fairly natural, one hypothesis is that they read the utterances with a (possibly implicit) prosody on which (10) sound natural. Two possible prosodies are indicated by the ToBI transcriptions of in (10); in both, the interrogative is realized with the so-called rise-fall-rise contour (see, e.g., Ward and Hirschberg 1985; Büring 1997, 2003; Wagner et al. 2013).[18]

(10) a. I don't **know** if Julian dances **salsa**. Does **Cole** know that Julian dances **salsa**?
L* H- L* L-H% L*+H L-H%

b. **I** don't **know** if Julian dances **salsa**. Does **Cole** know that Julian dances **salsa**?
L*+H L* H- L* H-H% L*+H L-H%

If there are prosodies on which sentences like (10) sound natural, then naturalness ratings in explicit ignorance contexts do not merely reflect the defeasibility of a projection inference but rather are also sensitive to (implicit) prosody. This sensitivity must be taken into consideration in the interpretation of the results.

In sum, these considerations about the low naturalness ratings of *be annoyed* and *know* in explicit ignorance contexts suggest that naturalness ratings in such contexts may not just be sensitive to whether the associated projection inference is defeasible, but also to its at-issueness and its information structural status.

3 General discussion

Presuppositions have long been taken to be a well-defined subset of inferences, namely ones that are typically backgrounded, typically projective and associated with particular expressions or constructions (see, e.g., Potts 2005; Beaver et al. 2024). On a type of analysis that is still widely assumed, they are analyzed as conventionally specified constraints on the common ground that must be satisfied prior to interpretation

[18]ToBI refers to the Tones and Break Indices annotation system of Beckman and Ayers 1997. In this system, L*+H identifies a pitch accent whose low target is aligned with the stressed syllable and which is followed by a trailing high target. A L* pitch accent may also be followed by a low (L-) or high (H-) intermediate phrase tone. An intonational phrases may end with a low intermediate phrase tone and a high intonational phrase tone (L-H%) or a high intermediate and intonational phrase tone (H-H%).

(e.g., Heim 1983; van der Sandt 1992). As noted in §1, a pressing question for such analyses of presuppositions is to identify which projection inferences are presuppositions and which ones are not. As also noted in §1, the results of experimental investigations suggest that the strength of the projection inference does not distinguish presuppositional projection inferences from nonpresuppositional ones (see, e.g., de Marneffe et al. 2019; Degen and Tonhauser 2022). And the results of the experiment reported on in §2 (as well as the results of Kalomoiros and Schwarz 2024:Exp. 1) suggest that comparison of naturalness ratings in explicit ignorance and support contexts also does not reliably distinguish presuppositional projection inferences from nonpresuppositional ones. Where do we go from here?

One possibility, also suggested in Degen and Tonhauser 2022:§4, is to ask whether there are more appropriate measures to identify presupposed projection inferences than inference rating tasks (used in, e.g., Degen and Tonhauser 2022) or comparison of naturalness ratings across contexts (used in the experiment in §2). Alternatively, one could give up on the assumption that there is a well-defined subset of projection inferences that we could call 'presuppositions'. This latter position is supported by the fact that research on projection inferences has long identified that the set of contents that have been called presuppositions is heterogeneous along several dimensions. One dimension of variation that was already mentioned in §1 is projection. Experimental investigations suggest that contents traditionally called presuppositions exhibit variation in the strength of the inference to the content (e.g., Xue and Onea 2011 on German, Tonhauser 2020 on Paraguayan Guaraní, and de Marneffe et al. 2019; Tonhauser et al. 2018; Degen and Tonhauser 2022 on English). In English, for instance, the CC of *discover* is less projective than the CC of *know*. These results challenge the traditional characterization of presuppositions as contents that "typically project".

A second dimension of variation is whether the content is associated with what Tonhauser, Beaver, Roberts, and Simons 2013 referred to as a 'strong contextual felicity' constraint. Some presumed presuppositions are not associated with such a constraint, which means that they are judged to be acceptable in a context that is neutral with respect to the content. This is the case for the CC of *know*, illustrated in (11a), which is informative in such a context. Other presumed presuppositions are associated with such a constraint, that is, they are judged to be unacceptable in a context that does not entail or satisfy the content. This is the case for the existence requirement of *too*, illustrated in (11b).

(11) Tonhauser et al. 2013:78, 80

a. A girl backs out of a driveway and hits Susi's car. A woman comes running out of the house, apologizes that her daughter hit Susi's car, and says:
She knows that she has to use her glasses to drive.

b. Malena is eating her lunch, a hamburger, on the bus going into town. A woman who she doesn't know sits down next to her and says:
#Our bus driver is eating empanadas, too.

This variation suggests that analyses of presuppositions as conventionally-coded felicity requirements (e.g., Heim 1983; van der Sandt 1992) generalize an analysis that is empirically justified for contents associated with a strong contextual felicity constraint to contents that can be informative.

A third dimension of variation is at-issueness, that is, whether the content can address the question under discussion. For instance, as shown in Tonhauser et al. 2018, the pre-state content of *stop* and the CC of *discover* are more at-issue than the CCs of *know* and *be annoyed*. This variability challenges a unified characterization of presuppositions as backgrounded content. Finally, there is the dimension investigated in our experiment, referred to as defeasibility in Simons 2001 and Abusch 2010. The results of our experiment suggest that contents traditionally called presuppositions exhibit variation in how natural (or: defeasible) they are in explicit ignorance contexts. For instance, the pre-state content of *continue* was rated as less natural than the pre-state content of *stop*, which was rated as less natural than the existential content of the *it*-cleft, which was rated as less natural than the CC of *see*, which was rated as less natural than the CC of

reveal. No unified characterization of 'presupposition' is forthcoming from this measure either.

These dimensions of variation strongly suggest that the set of contents traditionally referred to as presuppositions does not constitute a natural class. It is in response to such variation that researchers have developed analyses of subsets of these contents that capture projection variability and do not rely on a conventional specification of felicity conditions (e.g., Abrusán 2011, 2013, 2016; Abusch 2002, 2010; Romoli 2015). A fruitful path forward for contemporary research on projective content is to engage with this variation rather than to keep taking for granted a characterization of 'presupposition' based on projection or an analysis as conventionally-specified felicity conditions.

4 Conclusions

For many decades the term 'presupposition' has served researchers well to refer to content with properties distinct from, for instance, entailments, conventional implicatures and conversational implicatures. However, over the past two decades, both empirical and theoretical research has pointed out that the set of contents typically referred to as presuppositions is not as homogenous as perhaps initially assumed, resulting in theoretical analyses of particular subsets and an understanding of multiple dimensions of empirical variation among this set of contents. Against this background, the proposal by Mandelkern et al. 2020 that there is an empirical measure that distinguishes presuppositions from nonpresuppositions sparked hope that there might be, after all, a way of preserving the term 'presupposition' as a well-defined notion in the way it has been traditionally used. Unfortunately, the results of our experiment did not support this claim. We suggest retiring the term 'presupposition' or, at least, confining its use to a particular subtype of projective content, such as content associated with a strong contextual felicity constraint.

References

Abrusán, Márta. 2011. Predicting the presuppositions of soft triggers. *Linguistics & Philosophy* 34:491–535.

Abrusán, Márta. 2013. A note on quasi-presuppositions and focus. *Journal of Semantics* 30:257–265.

Abrusán, Márta. 2016. Presupposition cancellation: Explaining the 'soft-hard' trigger distinction. *Natural Language Semantics* 24:165–202.

Abusch, Dorit. 2002. Lexical alternatives as a source of pragmatic presupposition. *Semantics and Linguistic Theory* 12:1–19.

Abusch, Dorit. 2010. Presupposition triggering from alternatives. *Journal of Semantics* 27:37–80.

Anand, Pranav and Valentine Hacquard. 2014. Factivity, belief and discourse. In L. Crnič and U. Sauerland, eds., *The Art and Craft of Semantics: A Festschrift for Irene Heim*, pages 69–90. MIT Working Papers in Linguistics.

Beaver, David I., Bart Geurts, and Kristie Denlinger. 2024. Presupposition. In E. Zalta and U. Nodelman, eds., *The Stanford Encyclopedia of Philosophy*. Stanford University.

Beckman, Mary E. and Gayle M. Ayers. 1997. *Guidelines for ToBI labelling, version 3.0*. The Ohio State University.

Büring, Daniel. 1997. *The Meaning of Topic and Focus — The 59th Street Bridge Accent*. London: Routhledge.

Büring, Daniel. 2003. On D-trees, beans and B-accents. *Linguistics & Philosophy* 26:511–545.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1):1–28.

Chierchia, Gennaro and Sally McConnell-Ginet. 1990. *Meaning and Grammar*. Cambridge, MA: MIT Press.

de Marneffe, Marie, Mandy Simons, and Judith Tonhauser. 2019. The CommitmentBank: Investigating projection in naturally occurring discourse. *Sinn und Bedeutung* 23:107–124.

Degen, Judith and Judith Tonhauser. 2021. Prior beliefs modulate projection. *Open Mind* 5:59–70.

Degen, Judith and Judith Tonhauser. 2022. Are there factive predicates? An empirical investigation. *Language* 98:552–591.

Djärv, Kajsa. 2019. *Factive and assertive attitude reports*. Ph.D. thesis, University of Pennsylvania.

Gazdar, Gerald. 1979. *Pragmatics: Implicature, Presuppositions, and Logical Form*. New York: Academic Press.

Heim, Irene. 1983. On the projection problem for presuppositions. *West Coast Conference on Formal Linguistics* 2:114–125.

Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9:183–221.

Jayez, Jacques, Valeria Mongelli, Anne Reboul, and Jean-Baptiste van der Hengst. 2015. Weak and strong triggers. In F. Schwarz, ed., *Experimental Perspectives on Presuppositions*, pages 173–193. Heidelberg: Springer.

Kalomoiros, Alexandros and Florian Schwarz. 2024. Presupposition projection from 'and' vs. 'or': Experimental data and theoretical implications. *Journal of Semantics* 41:331–372.

Karttunen, Lauri. 2016. Presupposition: What went wrong? In *Proceedings of Semantics and Linguistic Theory (SALT) XXVI*, pages 705–731. Ithaca, NY: CLC Publications.

Klein, Ewan. 1975. Two sorts of factive predicates. Pragmatics Microfiche it 1.1. frames B5C14.

Lenth, Russell V. 2023. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.8.8.

Mandelkern, Matthew, Jérémy Zehr, Jacopo Romoli, and Florian Schwarz. 2020. We've discovered that projection across conjunction is asymmetric (and it is!). *Linguistics & Philosophy* 43:473–514.

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5:1–69. Reprint of 1998 revision of 1996 OSU Working Papers publication.

Roberts, Craige and Mandy Simons. 2024. Preconditions and projection: Explaining non-anaphoric presupposition. *Linguistics & Philosophy* 47:703–748.

Romoli, Jacopo. 2015. The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of Semantics* 32:173–291.

Simons, Mandy. 2001. On the conversational basis of some presuppositions. *Semantics and Linguistics Theory* 11:431–448.

Smith, E. Allyn and Kathleen Currie Hall. 2011. Projection diversity: Experimental evidence. *2011 ESSLLI Workshop on Projective Content.* pages 156–170.

Smithson, Michael and Jay Verkuilen. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11:54–71.

Tonhauser, Judith. 2020. Projection variability in Paraguayan Guaraní. *Natural Language and Linguistic Theory* 38:1263–1302.

Tonhauser, Judith, David Beaver, and Judith Degen. 2018. How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics* 35:495–542.

Tonhauser, Judith, David Beaver, Craige Roberts, and Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89:66–109.

van der Sandt, Rob. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9:333–377.

Wagner, Michael, Elise McClay, and Lauren Mak. 2013. Incomplete answers and the rise-fall-rise contour. In *Proceedings of the 17th Workshop on the Semantics and Pragmatics of Dialogue*, pages 140–149.

Ward, Gregory and Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61:746–776.

Xue, Jingyang and Edgar Onea. 2011. Correlation between projective meaning and at-issueness: An empirical study. *2011 ESSLLI Workshop on Projective Content.* pages 171–184.

Supplemental materials

A Experiment stimuli

The twenty clause-embedding predicates used in the experiment are the same as in Degen and Tonhauser 2021, 2022:

(12) a. Factive: be annoyed, discover, know, reveal, see
b. Non-factive: acknowledge, admit, announce, be right, confess, confirm, establish, hear, inform, pretend, prove, say, suggest, think

Eventive predicates, like *discover* and *hear*, were realized in the past tense and stative predicates, like *know* and *be annoyed*, were realized in the present tense. The direct object of *inform* was realized by the proper name *Sam*. Each clause-embedding predicate was paired with a unique subject proper name. The speaker of the target stimuli was realized by a randomly sampled unique proper name.

The following list shows the 20 clauses that realized the complements of the predicates in the target stimuli, together with their lower and higher probability facts, respectively, that realized the preceding declarative sentences in the two support contexts.

1. Mary is pregnant. Facts: Mary is a middle school student / Mary is taking a prenatal yoga class
2. Josie went on vacation to France. Facts: Josie doesn't have a passport / Josie loves France
3. Emma studied on Saturday morning. Facts: Emma is in first grade / Emma is in law school
4. Olivia sleeps until noon. Facts: Olivia has two small children / Olivia works the third shift
5. Sophia got a tattoo. Facts: Sophia is a high end fashion model / Sophia is a hipster
6. Mia drank 2 cocktails last night. Facts: Mia is a nun / Mia is a college student
7. Isabella ate a steak on Sunday. Facts: Isabella is a vegetarian / Isabella is from Argentina
8. Emily bought a car yesterday. Facts: Emily never has any money / Emily has been saving for a year
9. Grace visited her sister. Facts: Grace hates her sister / Grace loves her sister
10. Zoe calculated the tip. Facts: Zoe is 5 years old / Zoe is a math major
11. Danny ate the last cupcake. Facts: Danny is a diabetic / Danny loves cake
12. Frank got a cat. Facts: Frank is allergic to cats / Frank has always wanted a pet
13. Jackson ran 10 miles. Facts: Jackson is obese / Jackson is training for a marathon
14. Jayden rented a car. Facts: Jayden doesn't have a driver's license / Jayden's car is in the shop
15. Tony had a drink last night. Facts: Tony has been sober for 20 years / Tony really likes to party with his friends
16. Josh learned to ride a bike yesterday. Facts: Josh is a 75-year old man / Josh is a 5-year old boy
17. Owen shoveled snow last winter. Facts: Owen lives in New Orleans / Owen lives in Chicago
18. Julian dances salsa. Facts: Julian is German / Julian is Cuban
19. Jon walks to work. Facts: Jon lives 10 miles away from work / Jon lives 2 blocks away from work
20. Charley speaks Spanish. Facts: Charley lives in Korea / Charley lives in Mexico

B Filler and practice stimuli

The following list shows the four filler stimuli where the interrogative was expected to receive high naturalness ratings in the context of the preceding declarative sentence. The values in parentheses indicate the mean naturalness rating that each of the four filler stimuli received. As shown, the third filler stimulus did not receive the expected high naturalness mean ratings, probably because participants were unwilling to accommodate that Hendrick has a car in a context in which Hendrick was looking to buy a car. This filler stimulus was therefore not used to exclude participants' data.

1. I don't know if Samantha has a new hat. Does Samantha have a new hat? (.89)
2. I don't know if this pizza has mushrooms on it. Does this pizza have mushrooms on it? (.87)
3. Hendrick was looking to buy a car. Was Hendrick's car expensive? (.5)
4. Mary visited her aunt yesterday. Is Mary's aunt sick? (.91)

The following list shows the four practice stimuli in the order in which they were presented to the participants. Participants were able to advance to the experiment only if they gave a naturalness rating higher than .6 for the first and third stimulus, and a naturalness rating lower than .4 for the second and fourth stimulus.

1. I have no idea where Natalie is from. Is Natalie from the USA?
2. I don't have any sisters. Have you met my sister yet?
3. I am going on vacation to Ireland. Does Fritz realize that Joe is going with me?
4. I have no idea if Anna has any dogs. Is Samuel glad that Anna fed her dogs?