I don't know if projection is categorical.
Did Mandelkern et al. 2020 discover that it is?

Judith Tonhauser° and Judith Degen•

°University of Stuttgart
•Stanford University

November 27, 2023

Abstract

Presuppositions are taken to typically project out of entailment-canceling environments like the scope of negation (e.g., Chierchia and McConnell-Ginet 1990). While projection is often characterized as categorical (that is, content either typically projects or not), there is mounting empirical evidence that projection is gradient, with content being more or less projective (e.g., Karttunen 1971, Xue and Onea 2011, de Marneffe et al. 2019, Tonhauser et al. 2018, Degen and Tonhauser 2022). Mandelkern et al. 2020 critically evaluated the inference rating measures on which this empirical evidence is based and claimed that a different measure, namely naturalness ratings in explicit ignorance contexts, provides support for categorical projection and categorically distinguishes presuppositions from nonpresuppositions. This paper presents the results of an experiment designed to investigate Mandelkern et al.'s 2020 claim for factive predicates (presumed presupposition triggers) and nonfactive ones (nontriggers). The results do not support Mandelkern et al.'s 2020 claim and rather align with Degen and Tonhauser's 2022 result that there is no evidence for a categorical distinction between factive and nonfactive predicates.

1   Introduction

Presuppositions are often characterized as content that typically projects out of entailment-canceling environments, like the scope of negation or a polar interrogative (e.g., Chierchia and McConnell-Ginet 1990).[1] Thus, a content like the content of the clausal complement (CC) of *know* in (1a) is taken to be a presupposition because a speaker who utters (1a) is typically taken to believe the CC, that Julian dances salsa, even though the CC occurs in a polar interrogative. By contrast, the CC of *think* in (1b) is not assumed to be a presupposition because it does not typically project. Accordingly, *know* is analyzed as a presupposition trigger, but *think* is not.

(1)   a.   Does Cole know that Julian dances salsa?
       b.   Does Cole think that Julian dances salsa?

Many theoretical works assume that projection is a categorical property of utterance content, that is, content either typically projects (in which case it is a presupposition) or it does not (in which case it is not a presupposition). There is, however, mounting empirical evidence that projection is not a categorical property of content but rather a gradient one, with content being more or less projective. The measure used to establish this evidence are inference rating measures: Such measures involve presenting participants with an utterance (like (1a)) and contents (like that Julian dances salsa), and asking for ratings that are taken to reveal the existence or strength of the inference to the content. Karttunen 1971, for instance, suggested that the CCs of *regret* and *discover*, which he took to be presuppositions, exhibit projection variability in examples like (2). Specifically, he suggested that a speaker who utters (2a) believes that the addressee has not told the truth, whereas a speaker who utters (2b) "is not sure about the truth of the complement" (p.63).

(2)   a.   Did you regret that you had not told the truth?
       b.   Did you discover that you had not told the truth?   (Karttunen 1971:63)

More recently, experimental investigations have also observed projection variability among presuppositions (e.g., Xue and Onea 2011, de Marneffe, Simons, and Tonhauser 2019, Tonhauser, Beaver, and Degen 2018, Degen and Tonhauser 2022). Tonhauser et al. 2018, for instance, observed in their Exps. 1 that the CC of *know* is more projective than that of *discover*, and the CC of *discover* is more projective than that of *reveal*. These results led Tonhauser et al. 2018:498 to propose that "projectivity [. . . is] a gradient property of content, rather than a binary, categorical one". Further support for this proposal comes from investigations of both factive and nonfactive predicates in Degen and Tonhauser 2022. In a series of experiments and analyses of existing datasets, this work not only found projection variability between the CCs of factive predicates, but also no empirical support for a categorical distinction in the projection of the CCs of factive and nonfactive predicates. This is because the CCs of both factive and nonfactive predicates were projective to varying degrees compared to nonprojective main clause content. A sample illustration of these results is given in Fig. 1 from their Exp. 1a, which shows the mean projection ratings of 5 factive predicates (in orange) and 15 nonfactive predicates (in green).

Recently, Mandelkern, Zehr, Romoli, and Schwarz 2020:§6.2 challenged the hypothesis that projection is gradient by arguing that inference rating measures are not suitable to detect categorical projection. Such measures are not suitable, they argued, because they do not distinguish between, on the one hand, contents that project because they are presuppositions and, on the other hand, contents that project because they are "a natural conclusion to draw for any of a variety of pragmatic reasons short of entailment or presupposition" (p.497). Inference rating measures, they suggested, might invite participants to draw projection inferences even when not lexically required because "the content in question is right there before the [participants']
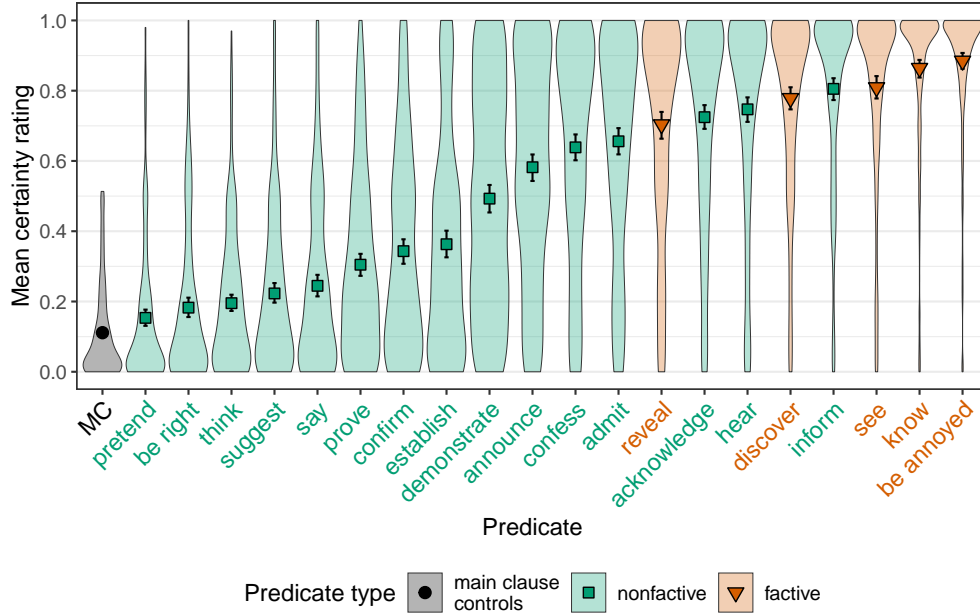

Figure 1: Mean certainty rating (measuring projection) of main clause ('MC') content and the CCs of the 20 factive and nonfactive predicates investigated in Exp. 1a of Degen and Tonhauser 2022 (adapted by collapsing the nonfactive predicates into a single category). Error bars indicate 95% bootstrapped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

eyes" (p.498). They concluded that "we should think twice before embracing a notion of presupposition projection that is gradient based on results from inference tasks alone" (p.497).

Mandelkern et al. 2020:§6.2 suggested that naturalness ratings of utterances in explicit ignorance contexts are better suited to distinguish presuppositions from nonpresuppositional projection inferences.[2] On this measure, participants are presented with utterances in which a speaker states that they are ignorant about a content, as in (3a), and then follows up with an utterance in which the content occurs embedded in an entailment-canceling environment, such as the antecedent of a conditional in (3c). Mandelkern et al. 2020 assumed that if the content is a presupposition (such as the pre-state content of *stop* in (3c)), the utterance is rated as unnatural in the explicit ignorance context because presuppositions are contents that the speaker believes to be true. That is, in such constellations, participants "will not have an alternative to interpreting the presupposition at the utterance level, and in turn seeing the utterance as infelicitous and the speaker as incoherent" (p.497). If, on the other hand, the content is not a presupposition (such as the pre-state content of *now frown on* in (3c)), the utterance is assumed to be natural: Even if participants might, in a different context, draw a projection inference "for any of a variety of pragmatic reasons short of entailment or presupposition, [...] they will tend to relinquish that inclination when there is pragmatic pressure to do so" as in the explicit ignorance contexts (p.497). Both presuppositions and nonpresuppositions were expected to be judged as natural in the support contexts, which entail the relevant content, illustrated in (3b).

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

    a.   Explicit ignorance context:
          Mary always was involved in a lot of sports, but I don't know whether she ever did any yoga.

---

[2]Mandelkern et al. 2020 refer to the measure as an acceptability rating measure, but since participants were asked to rate the naturalness of utterances, we refer to it as a naturalness rating measure.

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.

In Mandelkern et al.'s 2020 Exp. 3, which used the proposed measure, items like (3) were control items. Mandelkern and colleagues found that the naturalness ratings for nontriggers were relatively high in both contexts (with means just below 5, where 1 meant "completely unnatural" and 7 meant "completely natural" on their 7-point Likert scale). The presupposition triggers, on the other hand, had a lower mean naturalness rating in the explicit ignorance context (mean around 3) than in the support context (mean at about 5.3). Mandelkern and colleagues took this to "establish[] the effectiveness of the methodology" (p.492). They proposed (p.497):

> [...] 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, and thus have no choice but to project. [...] In methodological terms, we strongly recommend at least a two-pronged approach, with careful attention paid to results stemming from the evaluation of acceptabilty [sic] in different contexts.

However, the empirical evidence in Mandelkern et al. 2020 in support of the claim projection is categorical is quite weak. This is primarily because the goal of their work was not to investigate this claim, but rather whether presupposition projection in conjunctions is symmetric. It is also weak, however, because the data provided in Fig. 8 in their Appendix 3 do not appear to support a categorical distinction between the four presupposition triggers (*be aware, be happy, stop, continue*) and the four nontriggers (*now frown on, enjoy, hope, be sure*) that were investigated. First, considering just the mean naturalness ratings in the explicit ignorance context, it appears that the ratings for the presuppositions and nonpresuppositions overlap: The ratings for the presuppositions ranged from 1 (for the pre-state of *continue*) to about 3.5 (for the CC of *be aware*), and the ratings for the nonpresuppositions range from just below 3 (for the CC of *hope*) to 4.5 (for the CC of *be aware*). It is therefore not clear that naturalness ratings in explicit ignorance contexts provide evidence for a categorical distinction between the four presuppositions and the four nonpresuppositions investigated in Mandelkern et al. 2020. Second, there is also no consistent effect of context. Whereas the difference between the mean naturalness ratings in the explicit ignorance and support contexts was quite large for the (presuppositional) pre-state of *continue*, the difference was much less pronounced for the (presuppositional) CC of *be aware*, where it was roughly comparable to that of the (nonpresuppositional) pre-state of *frown on*. Thus, it is not clear that comparisons of ratings in support and explicit ignorance contexts provides empirical support for a categorical projection either.

Another issue with Mandelkern et al.'s 2020 claim is that there is disagreement in the literature about the naturalness of presuppositions in explicit ignorance contexts. Whereas Mandelkern and colleagues assumed that presuppositions are unnatural in explicit ignorance contexts, Simons 2001 and Abusch 2010 assumed that only 'hard' (nondefeasible) presuppositions are, but that 'soft' (defeasible) ones are natural. For instance, the examples with *stop*, *discover*, and *win* in (4) were assumed to be natural in explicit ignorance contexts, and the examples with *again*, the *it*-cleft and *too* in (5) unnatural.[3]

(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)

---
[3] Abusch 2010 marked (5b) and (5c) with '??' instead of '#'.

4

    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)

This paper presents the results of an experiment designed to investigate Mandelkern et al.'s 2020 claim that the measure they proposed provides evidence that projection is categorical and distinguishes presuppositions from nonpresuppositions.[4] To allow for comparison with the results of Degen and Tonhauser 2022, the CCs of the same 20 (non)factive clause-embedding predicates (see Fig. 1) were investigated. The experiment also included control stimuli with *stop, continue, again, too, also*, and an *it*-cleft to allow for comparison with the results of Mandelkern et al. 2020 and the assumptions made in Simons 2001 and Abusch 2010.

We slightly changed the design of Mandelkern et al.'s 2020 Exp. 3 to allow for comparison with the result of Degen and Tonhauser 2021 that the projection of content is sensitive to the content's 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 (6c), that Julian dances salsa, was stronger in the context in (6b), where the CC has a higher prior probability, than in the context in (6a), where the CC has a lower prior probability.

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

    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 that entailed the relevant context with two support contexts that differed in the prior probability of the CC, namely a lower and a higher prior probability context.[5] Under analyses of presuppositions like Heim 1983 or van der Sandt 1992, presuppositions are by default globally accommodated in contexts that are compatible with the presupposition, like the lower and higher prior probability contexts. Accordingly, such analyses predict that presuppositions are rated as natural in both lower and higher probability contexts, just as in Mandelkern et al.'s 2020 entailing support context.

2 Experiment

To investigate Mandelkern et al.'s 2020 claim that their naturalness rating measure provides support for a categorical distinction between presuppositions and nonpresuppositions, participants read two-sentence discourses consisting of a declarative (which provided the context) and an interrogative (which realized a (non)factive 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 server 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).

---

[4]Link to experiment, materials, and analysis scripts: **xxx DO THIS AFTER FINALIZING SCRIPTS**

[5]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'.

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 (7). 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 (7a), the declarative sentence conveyed the speaker's ignorance about the CC. In the 'lower prior probability' context (7b), the CC had a comparatively lower prior probability, whereas it had a comparatively higher prior probability in the 'higher prior probability' context (7c). The two contexts for each of the 20 complement clauses that were combined with the 20 predicates 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.

(7) a. Explicit ignorance context:
         I have no idea if Julian dances salsa. Does 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 control stimuli shown in (8). The interrogatives of the control 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 the presuppositions associated with these expressions.[6] Two of these expressions, namely *stop* and *continue*, were target expressions in Mandelkern et al.'s 2020 Exp. 3, where they were assumed to be unnatural in explicit ignorance contexts. This is in contrast to Simons 2001, who took *stop* to be a soft trigger that is natural in explicit ignorance contexts. Three expressions were assumed to be unnatural in explicit ignorance contexts in Simons 2001 (*again*) and Abusch 2010 (*too, it*-cleft). The control 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.

(8) 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?

---

[6]The contents investigated for *too* and *also* in (8d) and (8e) are that Ann plays an instrument other than the flute and that Svenja plays a sport other than soccer. This interpretation arises if *Ann* and *Svenja* associate with *too* and *also*, 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 an instrument and whether Svenja plays any sport, respectively.

The experiment also included 4 filler stimuli that were used to exclude participants not attending to the task (see Supplement B). The filler stimuli were expected to receive high naturalness ratings.

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 of 30 stimuli contained the same six control 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 2, 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 2: Sample trial with explicit ignorance context.

2.1.4 Data exclusion

We excluded the data of five participants who did not self-identify as native speakers of American English as well as the data of 23 participants whose mean response to the filler stimuli (expected to be natural) was more than 2 sd below the group mean.[7] The data of 370 participants entered into the analysis (ages: 19-80, mean: 40.7; 175 women, 185 male, 8 non-categorical, 2 preferred to not disclose). Each of the 20 predicates 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 control stimuli received 370 ratings each.

---

[7]Contrary to what was preregistered, one of the four filler stimuli was not used to exclude participants because it received a mean rating of only .5. See Supplement B for details.

2.2 Results and discussion

We first address the question of whether naturalness ratings in explicit ignorance contexts provide empirical evidence for a categorical difference between presuppositions and nonpresuppositions (§2.2.1). We then consider whether a comparison of naturalness ratings in the three contexts provides such evidence (§2.2.2).

2.2.1 Naturalness ratings in explicit ignorance contexts

Results. Recall that Simons 2001 and Abusch 2010 assumed that only hard presuppositions are unnatural in explicit ignorance contexts, whereas Mandelkern et al. 2020 assumed that all presuppositions are unnatural in such contexts. Figure 3 shows mean naturalness ratings in the explicit ignorance context by expression (in distinct colors: factive predicates, nonfactive predicates, controls). As shown, the purported presupposition triggers, that is, the controls and the factive predicates, do not exhibit a unified pattern: The means of four controls (*continue, too, also, again*) are at floor, those of *stop* and *be annoyed* are slightly higher, those of the *it*-cleft and *know* are higher again, but lower than those of *see* and *discover*, and, finally, those of *reveal* are highest. We also see that the naturalness ratings of *see, discover* and *reveal* are just as high as those of some nonfactive predicates. These results do not provide empirical support for a categorical distinction between presuppositions and nonpresuppositions, contrary to what Mandelkern et al. 2020 claimed. These results also do not provide support for a class of factive predicates that is categorically distinct from nonfactive ones, in line with the result of Degen and Tonhauser 2022 based on inference rating measures.
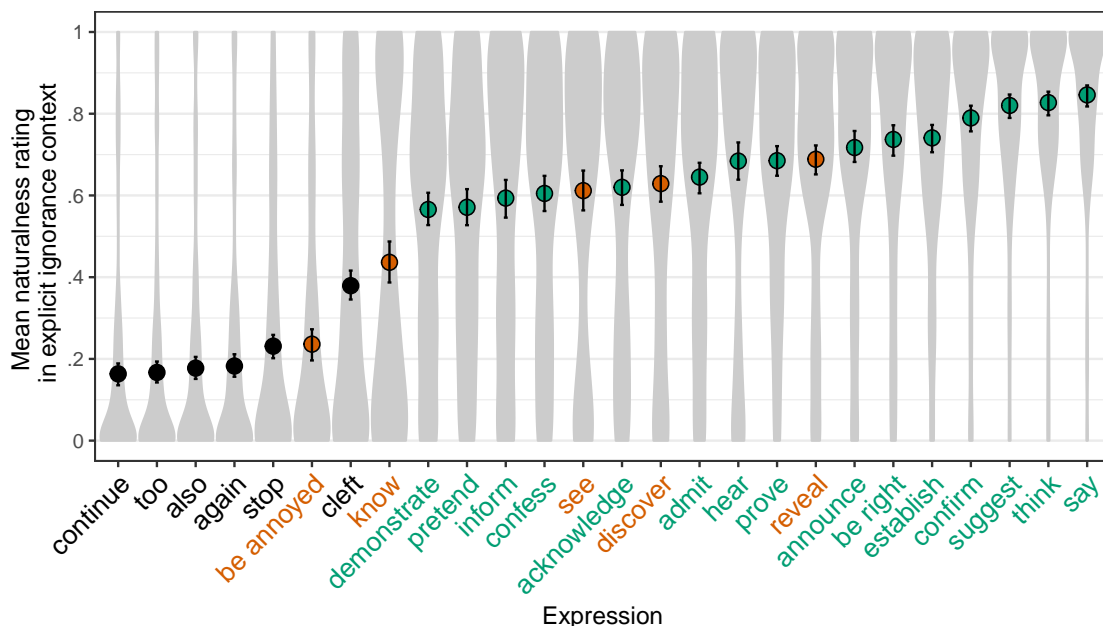

Figure 3: Mean naturalness rating in explicit ignorance context by expression (factive, nonfactive, filler). Error bars indicate 95% bootstrapped confidence intervals. Overlaid violin plots indicate the kernel probability density of the participants' ratings.

These observations were confirmed by a posthoc pairwise comparison of the naturalness ratings of the contents associated with the six controls and 20 target expressions 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 Bayesian mixed-effects beta regression model with weakly informative priors that was fitted using the 'brms' package (Bürkner 2017). The model predicted naturalness ratings 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 are 95% highest density intervals of estimated marginal mean differences between each of the 26 contents under investigation. We assume that two contents differ in their naturalness in the explicit ignorance context if the 95% highest density interval does not include 0.

As shown in Table 1,[8] the pairwise comparison suggests differences between the naturalness ratings of the CCs of the five factive predicates: The ratings of *be annoyed* are lower than those of *know*, which in turn are lower than those of *see* and *discover*, which in turn are lower than those of *reveal*. There is also variation between the factive predicates and the presuppositional controls: The CCs of each of the five purportedly factive predicates are more natural in explicit ignorance contexts than the presuppositions contributed by *too* and *also*. Furthermore, the naturalness ratings for the CCs of the five factive predicates are indistinguishable from different sets of expressions: The naturalness ratings for *be annoyed* do not differ from those of *stop* (but are higher than those the prejacents of the controls *continue, too*, and *also*), the naturalness ratings for *know* do not differ from those of the *it*-cleft, the naturalness ratings for *see* and *discover* do not differ from those of the nonfactives *confess, inform*, and *acknowledge*, and those of *reveal* do not differ from those of the nonfactives *admit, hear*, and *prove*. These results suggest that naturalness ratings in explicit ignorance contexts do not provide empirical support for a categorical distinction between factive and nonfactive predicates, or between presuppositions and nonpresuppositions.

Discussion. Contrary to what is assumed in Mandelkern et al. 2020 and Kalomoiros and Schwarz 2021, to appear, presuppositions are not invariably unnatural in explicit ignorance contexts. Rather, in line with what was assumed in Simons 2001 and Abusch 2010, some presuppositions seem to be rated as more unnatural in such contexts than others. On the assumption that naturalness ratings in explicit ignorance contexts differentiate between undefeasible/hard presuppositions and defeasible/soft presuppositions, one might take the results of our experiment to suggest that the pre-state contents of *continue* and *again*, and the existence requirements of *too* and *also* are not defeasible. The CCs of *see, discover*, and *reveal*, on the other hand, might be taken to be defeasible. Contrary to what Abusch 2010 assumed, the existence requirement of the *it*-cleft is not as clearly undefeasible as that of other projective contents.[9] Finally, the results of our experiment align with our observation about possible by-content variation in Mandelkern et al.'s 2020 Exp. 3. Recall that in their Exp. 3, the pre-state of *continue* was rated as (at least numerically) less natural in explicit ignorance contexts than the CC of the cognitive predicate *be aware*. This is in line with the result of our experiment that the pre-state of *continue* is rated less natural than the CC of the cognitive predicate *know*.

In our experiment, the CC of *be annoyed* received lower naturalness ratings than the CC of *know*. This result, if taken at face value, leads to the undesirable conclusion that the CC of *be annoyed* is more plausibly analyzed as a presupposition (perhaps even a hard one) than the CC of *know*. This conclusion is undesirable because the presupposition of emotive predicates like *be annoyed* is that the attitude holder believes the CC to be true (e.g., Heim 1992, Karttunen 2016), not that the speaker believes the CC to be true, as with *know*. In (9), for instance, the speaker does not believe that Sally did not leave a tip. Karttunen 2016 suggested that the inference to speaker belief is a generalized conversational implicature that comes about in contexts that support the "default assumption" that the beliefs of the speaker and the attitude holder are aligned (p.712).

(9)   Sally misremembered not having left a tip and regretted it.   (Karttunen 2016:712)

Given this status of the CC of *be annoyed*, it is surprising that this content received lower mean naturalness ratings in the explicit ignorance context than any of the other purportedly factive predicates. After all,

---

[8]The full model output is available here: `LINKTOTABLEINREPO` **ADD THIS AFTER FINALIZING SCRIPTS**.

[9]Smith and Hall 2011 also observed that the *it*-cleft did not pattern like a hard trigger in their experiment. Jayez et al. 2015 found that French *aussi* 'too', *regretter* 'regret' and clefts are not entirely unnatural in explicit ignorance contexts.
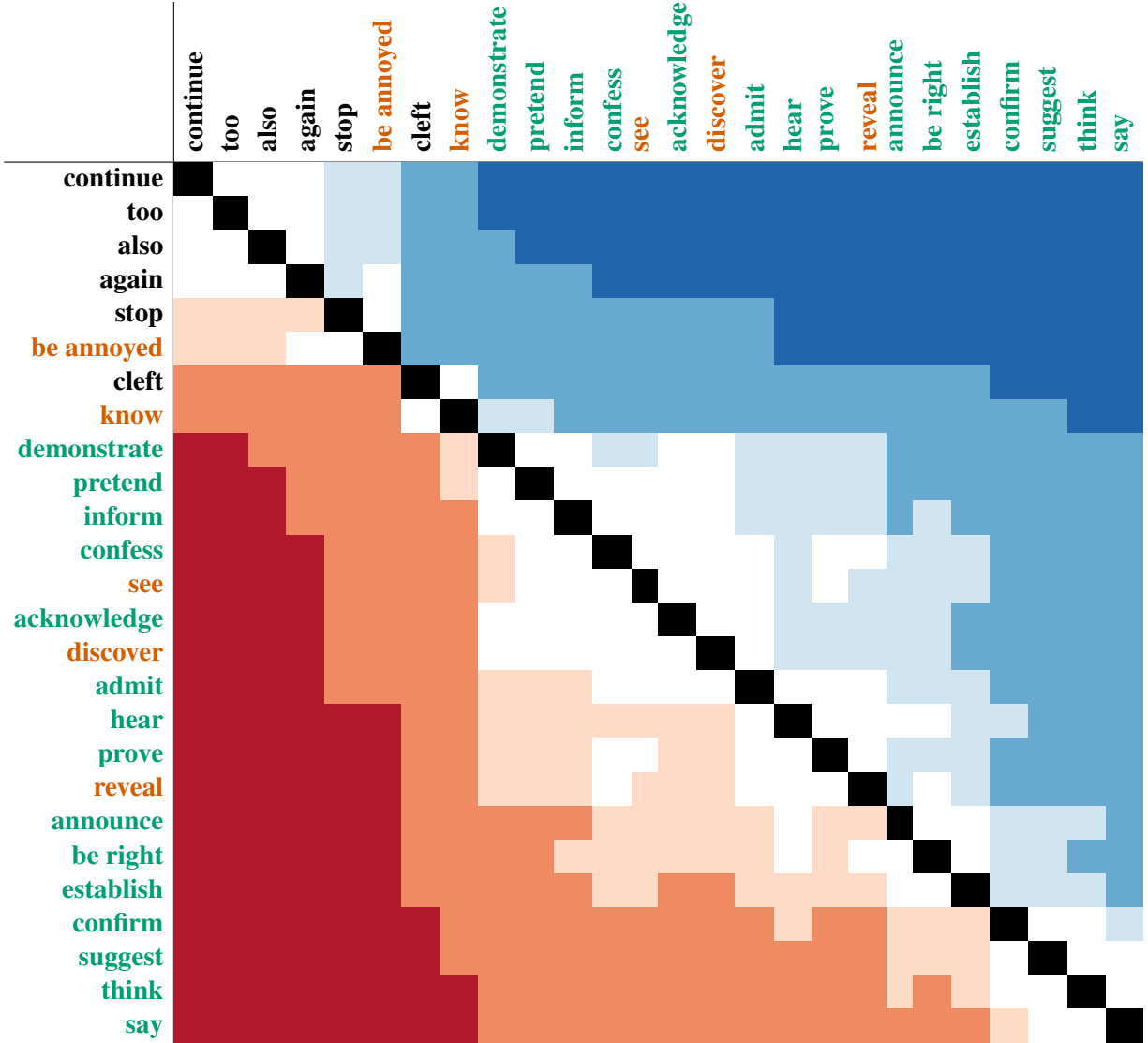
Table 1: Posterior distributions of estimated marginal means of pairwise differences in the explicit ignorance context, with expressions ordered by mean naturalness rating in the same context. Color coding indicates whether the difference between the naturalness ratings for contents of the expressions in the first column and those of the expressions in the top row were positive or negative, and the size of the difference; a white cell means that there was no difference:

Positive difference: [>= 1.5] , [(1.5, 0.5]] , [(0.5, 0]]
Negative difference: [>= −1.5] , [(−1.5, −0.5]], [(−0.5, 0]]

in a context in which the speaker is explicitly ignorant of the CC, the default assumption that the speaker's and the attitude holder's beliefs are aligned does not arise, and so the interrogative with *be annoyed* should be rated as natural in the explicit ignorance context, as it only presupposes that the attitude holder, not the speaker, believes the CC.

One hypothesis for why the CC of *be annoyed* received such low ratings is that naturalness ratings in explicit ignorance context are sensitive not just to whether the presupposition is defeasible (which is not the case for *too, also* and *again*) but also to 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 to be investigated, but also identifies that the speaker takes the question under discussion (QUD) to be whether the CC is true (?CC; e.g., *Does Julian dances salsa?*). The speaker's immediately following interrogative utterance is interpreted with respect to the QUD ?CC. If so, the naturalness of the interrogative is modulated by whether the QUD ?CC 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 the CC. As shown in Tonhauser et al. 2018, projective content differs in whether it can be at-issue. The CCs of emotive predicates, including *be annoyed*, are 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 the QUD ?CC (e.g., *Does Julian dance salsa?*). This incongruence may be why the CC of *be annoyed* received comparatively low naturalness ratings in the explicit ignorance context.

For the CC of *know*, Fig. 3 shows a bimodal pattern for the naturalness ratings, with about half of the participants rating it as fairly unnatural in the explicit ignorance context and about half as fairly 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 inference arising from the interrogative that the speaker believes the CC to be true, or one could assume an analysis along the same lines as for *be annoyed* above as the CC of *know* is also highly not-at-issue (Tonhauser et al. 2018). 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 prosody that supports an interpretation on which the CC does not project out of the interrogative utterance. Two possible prosodies are indicated by the ToBI transcriptions of in (10);[10] in both, the interrogative is realized with the so-called rise-fall-rise contour.

(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%

Analyses of the rise-fall-rise contour generally agree that it conveys that the speaker is uncertain about something and that their utterance does not fully answer the QUD (though the analyses differ in other regards; see, e.g., Ward and Hirschberg 1985, Büring 1997, 2003, Wagner et al. 2013). In the context of the utterance of the declarative, which conveys that the speaker takes the QUD to be whether Julian dances salsa, the rise-fall-rise realization of the interrogative conveys that this utterance does not fully answer the QUD. This interpretation of the interrogative is possible only if the speaker is uncertain about the CC, in which case the interrogative does not contradict the speaker's preceding explicit ignorance statement.

In sum, naturalness ratings in explicit ignorance contexts do not support Mandelkern et al.'s 2020 claim of a categorical distinction between presuppositions and nonpresupposition. Instead, the results suggest that naturalness ratings in explicit ignorance contexts may be sensitive to whether the content is defeasible, to at-issueness, and to prosodically contributed meanings.

2.2.2 Comparison of naturalness ratings across contexts

Results.   We next consider whether presuppositions and nonpresuppositions differ with respect to the ratings in explicit ignorance and support contexts. Recall that Mandelkern et al. 2020 claimed that presuppositions are sensitive to the context manipulation, but not nonpresuppositions. Fig. 4 shows mean naturalness

---

[10]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%).

ratings for the CCs of the 20 factive and nonfactive predicates in the three contexts featured in our experiment (with predicates ordered by their mean rating in the explicit ignorance context, as in Fig. 3).

As shown, the effect of context is not uniform for the factive predicates. For *be annoyed* and *know* the mean naturalness rating is quite a bit lower in the explicit ignorance context than in the higher prior probability context, whereas the difference is much less pronounced for *see* and *discover*, and not observed for *reveal*. This last pattern is also observed for several nonfactive predicates, including *confess, hear, announce*, and *be right*. Thus, naturalness ratings in explicit ignorance vs. higher prior probability contexts do not provide support for a categorical distinction between the CCs of factive and nonfactive predicates. A comparison between the explicit ignorance and lower prior probability contexts also suggests differences between the factive predicates: Whereas for *be annoyed* the mean naturalness rating is lower in the explicit ignorance context than in the lower prior probability context, the two means are equally high for *know*, but the rating in the lower prior probability context is lower than that in the explicit ignorance context for *see, discover*, and *reveal*. This last pattern is observed for all of the nonfactive predicates. Thus, naturalness ratings in explicit ignorance vs. lower prior probability contexts also do not provide support for a categorical distinction between the CCs of factive and nonfactive predicates.

Finally, we observe that the prior probability of content modulates its naturalness ratings with all predicates, except *demonstrate* and *pretend*: Content is judged as more natural in a context in which it has a higher prior probability than in a context in which it has a lower prior probability. This result is in line with the result of Degen and Tonhauser 2021 that projection modulated by prior probability.
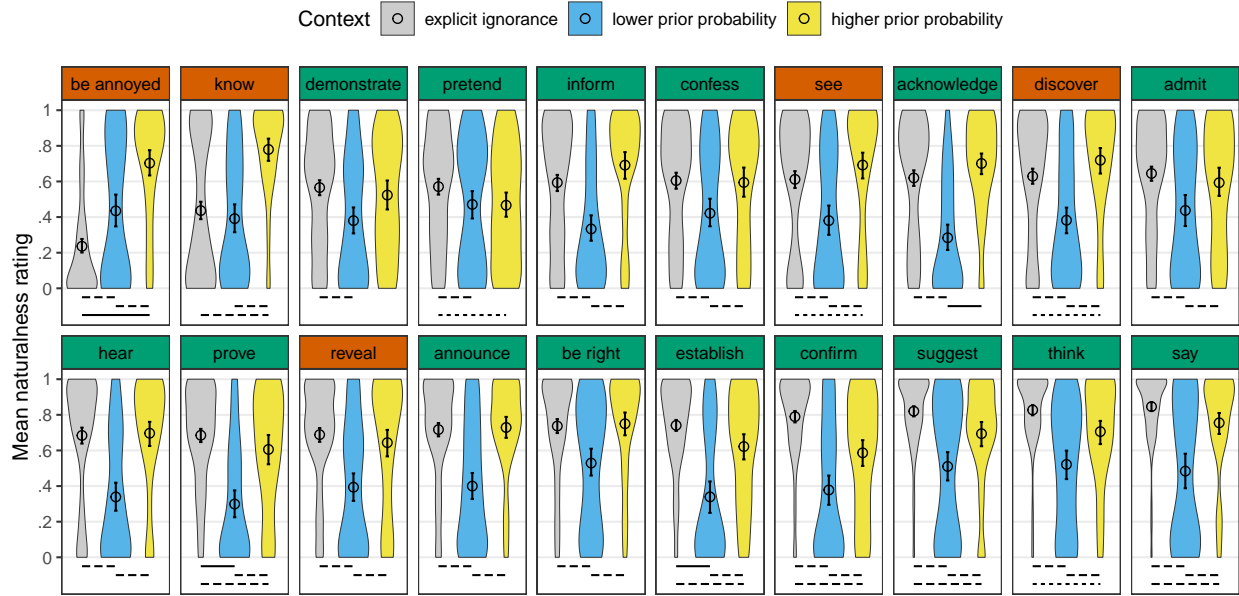

Figure 4: Mean naturalness rating by predicate (factive , nonfactive) and context, with predicates ordered as in Fig. 3. Error bars indicate 95% bootstrapped CIs. Violin plots indicate kernel probability density of participants' ratings. At the bottom of each facet, a line spanning two contexts indicates that the ratings in the two contexts differ according to the pairwise comparison. The linetype indicates whether the difference is larger than 1.5 (solid line: —), below 1.5 but above 0.5 (dashed line: – –), or below 0.5 and above 0 (dotted line: ⋯).

These observations were confirmed by posthoc pairwise comparisons of the naturalness ratings in the three contexts for each of the 20 predicates, again using the 'emmeans' package (Lenth 2023). The input to the pairwise comparisons were 20 Bayesian mixed-effects beta regression models with weakly informative

priors that were fitted using the 'brms' package (Bürkner 2017). The models predicted, for each predicate, the naturalness ratings from a fixed effect of context (with treatment coding and 'explicit ignorance' as reference level) and included random by-item intercepts. The output of the pairwise comparison are 95% highest density intervals of estimated marginal mean differences between each of the three contexts for each of the 20 contents. We assume that the naturalness ratings between two contexts differ if the 95% highest density interval does not include 0.[11] As shown in Fig. 4, naturalness ratings in explicit ignorance vs. lower or higher prior probability contexts do not provide support for a categorical distinction between presuppositions and nonpresuppositions.

Discussion. In contrast to Mandelkern et al.'s 2020 Exp. 3, our experiment did not collect naturalness ratings in a support context that entails the relevant content but in higher and lower prior probability contexts that are compatible with the CC. As mentioned above, contemporary presupposition analyses (e.g., Heim 1983, van der Sandt 1992) assume that presuppositions are by default globally accommodated when they are not entailed by or satisfied in the context, and they can be locally accommodated (e.g., under the polar question operator) to avoid contradiction, uninformativity or problems with binding. Such analyses lead us to expect that presuppositions are rated as natural in higher and lower prior probability contexts, as these contexts do not contradict the CCs, make the CCs uninformative, or lead to problems with binding. Contrary to this expectation, the CCs of the five purportedly factive predicates received lower ratings in the lower prior probability context than in the higher prior probability context.[12] This result suggests that naturalness ratings are sensitive to the prior probability of content, such that an interrogative with a clause-embedding predicate is judged as more natural in a context in which the CC has a higher prior probability than in a context in which it has a lower prior probability.

This result has several methodological and theoretical implications. First, as the result aligns with the result of Degen and Tonhauser 2021 based on an inference rating measure, it suggests that inference rating and naturalness rating measures are sensitive to some of the same factors. Second, naturalness ratings in explicit ignorance and other contexts are sensitive to a whole slew of factors, including the (implicit) prosody of the utterance and the defeasibility, at-issueness, and the prior probability of content. Finally, for analyses of presuppositions that local accommodation, the result means that local accommodation is sensitive to the prior probability of content, not just to whether the content is contradictory or uninformative in the context.

3 General discussion

The experiment was designed to investigate Mandelkern et al.'s 2020 claim that the measure used in their Exp. 3 provides support for a categorical distinction between presuppositions and nonpresuppositions. Neither the comparison of naturalness ratings in explicit ignorance contexts in §2.2.1 nor the comparison of naturalness ratings in explicit ignorance and support contexts in §2.2.2 provided empirical support for a categorical distinction between the CCs of factive and nonfactive predicates, or, more generally, between presuppositions and nonpresuppositions. Rather, the results align with that of Degen and Tonhauser 2022, namely that factive predicates are not a natural class that is categorically different from nonfactive predicates. In this section, we discuss theoretical implications of the results.

---

[11] The full model output is available here: LINKTOTABLEINREPO.**ADD AFTER FINALIZING SCRIPTS**

[12] In the higher prior probability context, our five purportedly factive predicates received ratings roughly comparable to the ratings of the factive predicates investigated in Mandelkern et al. 2020 in their support context. Specifically, in our higher prior probability context, the mean naturalness ratings for the CCs of the five factive predicates ranged from .64 (*reveal*) to .78 (*know*) on a scale from 0 to 1. In Mandelkern et al.'s 2020 support context, the CCs of *be aware* and *be happy* received a mean naturalness rating of about 4.4 and 4.7 out of 7 (based on Fig. 8 in their Appendix).

3.1 What are presuppositions?

Contemporary literature like Mandelkern et al. 2020 still frequently treats presuppositions as if they were a homogenous class, by characterizing them as backgrounded content that typically projects out of entailment-canceling environments or by analyzing them as conventionally-coded felicity requirements. However, research on projective content has long recognized that the set of contents that have been called presuppositions is heterogeneous. 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 et al. 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. Context: 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. Context: 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 characterizations is forthcoming from this measure either.

These dimensions of variation show 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 "presuppositions" based on projection or an analysis as conventionally-specified felicity conditions.

3.2 Are there factive predicates?

As discussed in Degen and Tonhauser 2022, factive predicates are variably defined as predicates whose CC is presupposed (e.g., Kiparsky and Kiparsky 1970) or as predicates whose CC is presupposed and entailed (e.g., Gazdar 1979, Schlenker 2010, Abrusán 2011). Degen and Tonhauser 2022 did not find empirical support for a natural class of factive predicates that is categorically distinct from nonfactive ones: First, based on the results of two experiments designed to investigate the projection of the CCs of (non)factive predicates as well as the analyses of three existing projection datasets with 45, 78, and 517 (non)factive predicates, respectively,[13] this work suggested that projection does not categorically distinguish factive and nonfactive predicates. Second, based on four experiments designed to investigate the entailment of the CCs of (non)factive predicates as well as analyses of entailment in two of the aforementioned datasets, this work suggested that considering entailment in addition to projection also does not provide support for a natural class of factive predicates that is categorically distinct from nonfactive ones.

Degen and Tonhauser 2022 used a variety of inference rating measures to investigate projection: In some measures, participants read constructed examples, in others naturally occurring ones; in some, examples were presented with a context, in others without a context; in some, the examples had diverse content, in others, they had minimal lexical content. Across the measures, the clause-embedding predicates were embedded under different entailment-canceling operators, participants were asked to respond to different questions, and participants gave their response on different scales and sliders. Despite this variation, the results invariably suggested that factive predicates are not categorically distinct from nonfactive ones.

The experiment presented in this paper presents supports Degen and Tonhauser's 2022 result based on a different measure, namely naturalness ratings in explicit ignorance contexts. The CCs of the five purportedly factive predicates exhibited variation in how natural they were judged to be when the speaker was explicitly ignorant of the CC, and there were no differences between the CCs of *see, discover* and *reveal* and several nonfactive predicates. Thus, neither data based on inference rating measures nor data based on naturalness ratings in explicit ignorance contexts provide support for a natural class of factive predicates. As discussed in Degen and Tonhauser 2022, "the search for a class of factive predicates might continue" (p.586), but we believe that it is more fruitful to investigate the more fine-grained distinctions between clause-embedding predicates that can lead to a better understanding of their use in discourse.

3.3 Is projection a categorical or a gradient property of content?

Mandelkern et al. 2020:497 urged us to "think twice before embracing a notion of [...] projection that is gradient based on results from inference tasks alone", advocating instead for a view of projection that is categorical. In this section, we turn to the question of whether projection is a categorical or a gradient property of content. Before addressing the question, we first consider what it is (not) about.

What the question is not about is the projection of *utterance content*. What we mean by that is the strength of the projection inference that a particular interpreter draws from a particular utterance of a sentence in which a content is contributed in the scope of an entailment-canceling operator. The empirical evidence pointed to in §1 shows that different interpreters draw projection inferences of varying strength. For instance, as shown in Fig. 5, the participants of Degen and Tonhauser's 2022 Exp. 1a drew projection inferences of varying strength for the CC that Julian dances salsa from utterances of the interrogative *Did Scott discover that Julian dances salsa?* It is uncontroversial that this utterance content variability is gradient, not categorical. It is gradient because the projection of utterance content is sensitive to several factors on which interpreters differ, including their prior beliefs about the speaker or about the utterance content, their perception and interpretation of the prosodic realization of the utterances, and their assumptions about the

---

[13]These datasets were the CommitmentBank (de Marneffe et al. 2019), the VerbVeridicality dataset (Ross and Pavlick 2019), and the MegaVeridicality dataset (White and Rawlins 2018).

question under discussion addressed by the utterance (e.g., Djärv and Bacovcin 2020, Degen and Tonhauser 2021, Mahler 2020, 2022, Tonhauser 2016, Tonhauser et al. 2018, 2019).
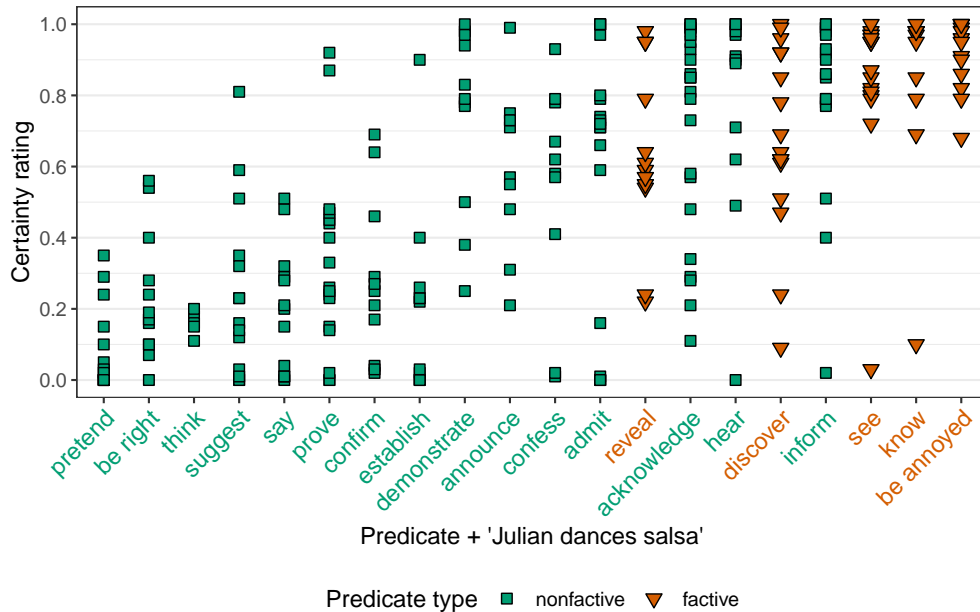

Figure 5: Participants' certainty ratings (measuring projection) of the CC that Julian dances salsa for the 20 factive and nonfactive predicates investigated in Exp. 1a of Degen and Tonhauser 2022.

What the question about whether projection is gradient or categorical is about is *content*, where a content is, for instance, the CC of *discover*, the CC of *inform*, or the pre-state content of *stop*. The question, thus, is about whether projection is gradient or categorical when we abstract away from variation that is attributable to interpreter variability and consider the observed by-content variation (for instance that the CC of *discover* is less projective than the CC of *know*; see Fig. 1): Is this by-content variation derived from underlying representations that either categorically trigger projection inferences or not, or from underlying representations that license projection inferences in combination with other factors in the uttered sentence?[14]

Both positions have already been advocated for in the literature. Categorical underlying representations are assumed in analyses of presuppositions like Heim 1983 and van der Sandt 1992, as they analyze presupposition triggers as being conventionally specified to trigger projection, whereas nontriggers do not trigger projection. To account for examples in which presuppositions do not project, such analyses rely on local accommodation. As discussed in detail in Tonhauser et al. 2018, it is not clear that local accommodation can capture the observed by-content variation in a principled way, as analyses would need to assume that the CC of *discover* is more amenable to local accommodation than the CC of *know*. Another problem such analyses face is that the CC of nonfactive predicates is also projective, sometimes even more so than the CC of factive predicates. Thus, as also discussed in Degen and Tonhauser 2022, it is not clear which expressions would be analyzed as triggering projection, and which ones not.

Other projection analyses do not assume underlying representations on which particular expressions are conventionally specified to trigger projection. For instance, in the analyses of Abusch 2002, 2010 and Romoli 2015, expressions may specify alternatives as part of their lexical entries and this specification can,

---

[14]Another potential source of projection variability is speaker belief. Tonhauser et al. 2018:498f. pointed out that speaker belief could be categorical (the speaker either believe that the content is true or they are uncertain) or gradient (the speaker believes that the content is true to some degree). On either view, interpreters might draw projection inferences of different strengths due to uncertainty about (the degree of) the speaker's belief.

depending on other factors in the uttered sentence, lead to a projection inference. In the analyses of Abrusán 2011, 2016 and Simons et al. 2010, 2017 it is entailed content that may project, depending on the utterance context. While it is not clear that these analyses predict the observed by-content projection variability or can be extended to the CCs of nonfactive predicates (see Tonhauser et al. 2018 and Degen and Tonhauser 2022 for discussion), what is important for our discussion here, is that such analyses do not assume categorical underlying representations.

Where does this leave us with the question of whether the projection of content is categorical or gradient? The challenge is clear: Proponents of either position must confront the observed by-content projection variability and develop analyses that capture the variability in a principled way. As none of the contemporary projection analyses are able to do so, one might say that the verdict is still out. The empirical evidence, however, seems to favor the hypothesis that projection is gradient, as there is, to date, no empirical evidence that projection is categorical, and empirical evidence from several different measures that it is gradient.

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 claim in Mandelkern et al. 2020 that there is an empirical measure that categorically 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, instead providing further evidence for variation among the set of contents traditionally referred to as presuppositions. We suggest retiring the term 'presupposition' or, at the least, confining its use to a particular subtype of projective content, such as those 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.

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: Routledge.

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* 5:5970.

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

Djärv, Kajsa and Hezekiah Akiva Bacovcin. 2020. Prosodic effects on factive presupposition projection. *Journal of Pragmatics* 169:61–85.

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. 2021. Presupposition projection from disjunction is symmetric. *Linguistic Society of America* 6:556–571.

Kalomoiros, Alexandros and Florian Schwarz. to appear. Presupposition projection from 'and' vs. 'or': Experimental data and theoretical implications. *Journal of Semantics* .

Karttunen, Lauri. 1971. Some observations on factivity. *Papers in Linguistics* 4:55–69.

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

Kiparsky, Paul and Carol Kiparsky. 1970. Fact. In M. Bierwisch and K. Heidolph, eds., *Progress in Linguistics*, pages 143–173. The Hague: Mouton.

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

Mahler, Taylor. 2020. The social component of projection behavior of clausal complements. *Linguistic Society of America* 5:777–791.

Mahler, Taylor. 2022. *Social identity information in projection inferences: a case study in social and semantic-pragmatic meaning*. Ph.D. thesis, The Ohio State University.

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.

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

Ross, Alexis and Ellie Pavlick. 2019. How well do nli models capture verb veridicality? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 2230–2240. Association for Computational Linguistic.

Schlenker, Philippe. 2010. Local contexts and local meanings. *Philosophical Studies* 151:115–142.

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

Simons, Mandy, David Beaver, Craige Roberts, and Judith Tonhauser. 2017. The best question: Explaining the projection behavior of factives. *Discourse processes* 54(3):187–206. Publisher: Taylor & Francis.

Simons, Mandy, Judith Tonhauser, David Beaver, and Craige Roberts. 2010. What projects and why. *Semantics and Linguistic Theory* 11:309–327.

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

Tonhauser, Judith. 2016. Prosodic cues to speaker commitment. *Semantics and Linguistic Theory* 26:934–960.

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.

Tonhauser, Judith, Marie-Catherine de Marneffe, Shari R. Speer, and Jon Stevens. 2019. On the information structure sensitivity of projective content. *Sinn und Bedeutung* 23:363–389.

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.

White, Aaron S. and Kyle Rawlins. 2018. The role of veridicality and factivity in clause selection. In *48th Meeting of the North East Linguistic Society*, pages 221–234.

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?