# Projection variability: Is the family of sentences really a family?[1]

Lisa HOFMANN — *University of Stuttgart*
Marie-Catherine DE MARNEFFE — *UCLouvain*
Judith TONHAUSER — *University of Stuttgart*

**Abstract.** We present experimental evidence **(i)** that the projection of the content of clausal complements varies across entailment-canceling operators, **(ii)** that the effect of operator varies between clause-embedding predicates, and **(iii)** we extend a result of Degen and Tonhauser 2022, that projection ratings in polar questions do not categorically distinguish factive from non-factives predicates, to cases with negation, the epistemic possibility modal *perhaps*, and conditional antecedents. The observed variability is not captured by existing theoretical accounts of projection (e.g., Heim 1983, van der Sandt 1992, Abrusán 2011, Schlenker 2021). Our results suggest that an analysis must consider interactions between predicates and operators and raise important questions for future research on projection.

**Keywords:** Projection variability, entailment-canceling operators, (non)factive predicates.

## 1. Introduction

The so-called 'family of sentences' diagnostic is the standard way of diagnosing whether a content is projective in formal semantics/pragmatics (e.g., Chierchia and McConnell-Ginet 1990). For instance, the clausal complement of *discover* is diagnosed as projective content, if its content (here: that Julian dances salsa) follows not just from (1), but also from the variants in (2), where (1) is embedded under an entailment-canceling operator, such as negation (2a), a polar question (2b), an epistemic possibility modal (2c), or in a conditional antecedent (2d).

(1) Cole discovered that Julian dances salsa.

(2) a. **Negation:** Cole didn't discover that Julian dances salsa.
b. **Polar Question:** Did Cole discover that Julian dances salsa?
c. **Modal:** Perhaps Cole discovered that Julian dances salsa.
d. **Conditional:** If Cole discovered that Julian dances salsa, Logan will be joyful.

Some research, however, suggests that entailment-cancelling operators may affect projection differentially. For instance, Karttunen (1971) proposed distinguishing English factive predicates (e.g., *regret*) from semi-factives (e.g., *discover*). Based on (3), he argued that the content of the complement (CC) of true factives consistently projects across the operators in (2), but that of semi-factives does not always project from under polar questions, modals, or conditionals.

(3) Karttunen 1971: (22, 24–26)
a. John didn't {regret/discover} that he had not told the truth.
b. Did you {regret/discover} that you had not told the truth?
c. If I {regret/discover} later that I have not told the truth, I will confess it to everyone.
d. It is possible that I {regret/discover} later that I have not told the truth.

[1]We thank Taylor Mahler for assistance in collecting the data presented here as well as valuable comments. We gratefully acknowledge the National Science foundation grant #1452674 (to Marie-Catherine de Marneffe, Craige Roberts, and Judith Tonhauser), which provided financial support for this project.

There have been two experimental investigations of by-operator projection variation. First, Smith and Hall (2014) investigated projection from under negation and conditional antecedents for various types of English projective contents. They found that the expressive content of epithets (e.g., *idiot*) and the CC of *know* was more projective under negation than conditionals. In contrast, the content of appositive relative clauses and the preparatory content of *win* showed the opposite pattern, and the existential presupposition of clefts showed no difference.

Second, Sieker and Solstad (2022) compared the projection of the CCs of German factives (*wissen* 'know', *bereuen* 'regret', *enthüllen* 'reveal') and semi-factives (*bemerken* 'notice', *entdecken* 'discover', *herausfinden* 'find out') from under the four operators in (2). The results suggest that the CCs project more from under negation than from under the other three operators. Contrary to what Karttunen (1971) suggested, a comparison of the factive and semi-factive predicates did not reveal that the CCs of factive predicates project more from under polar questions, modals, or antecedents of conditionals than the CCs of semi-factive predicates.[2]

This paper reports on the results of a set of experiments that were designed to compare projection from under the four entailment-canceling operators in (2) in English. Our experiments extend the empirical scope of prior research on by-operator projection variation by investigating projection for a larger set of contents, namely the contents of the complements of the 20 English clause-embedding predicates in (4), from Degen and Tonhauser (2022).

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

The five predicates in (4a) have been characterized as factive or semi-factive. Our set of predicates also includes the 15 non-factives in (4b). Including non-factive predicates in investigations of projection is motivated by the empirical investigations in De Marneffe et al. 2019 and Degen and Tonhauser 2022, which suggest that the CCs of non-factives may also project and that projection ratings do not categorically distinguish factive and non-factive predicates.

The results of our study suggest that the projection of the CCs of these 20 predicates is affected differently by the four entailment-cancelling operators in (2), but not in a way that is consistent with Karttunen's 1971 factive/semi-factive distinction. The results also replicate a result of Degen and Tonhauser 2022, namely that projection from under polar questions does not categorically distinguish factive and non-factive predicates. Finally, the results of our experiment extend this result of Degen and Tonhauser 2022 to projection from under the other three entailment-canceling operators, thereby solidifying their claim that "research on projective content has a much broader empirical scope than previously assumed" (p.585), as this scope includes the CCs of both factive and non-factive predicates.

[2]The factive/semi-factive distinction is also called into question by naturally occurring examples where the CCs of factive predicates do not project from under the four operators (see Beaver 2010, de Marneffe, Manning, and Potts 2012, De Marneffe, Simons, and Tonhauser 2019). For experimental research on the distinction, see **?** and Djärv, Zehr, and Schwarz 2018.

## 2. Experiments

To assess the effect of entailment-cancelling operator and clause-embedding predicate on projection, we collected projection judgments for the CCs of the 20 clause-embedding predicates in four sets of experiments. The predicates were embedded under negation in Exps. 1, under polar questions in Exps. 2, under the epistemic possibility modal *perhaps* in Exps. 3, and in conditional antecedents in Exps. 4. Each set of experiments contained three experiments, that differed in an at-issueness measure used in a separate block. In this paper, we limit our attention to the projection ratings collected in these twelve experiments.[3]

In all twelve experiments, projection was measured with the 'certain that' diagnostic, which has been used to measure projection with both polar interrogative and declarative sentences (see, e.g., Tonhauser 2016, Djärv and Bacovcin 2017, Stevens, de Marneffe, Speer, and Tonhauser 2017, Lorson 2018, Tonhauser, Beaver, and Degen 2018, Mahler 2019, 2020, De Marneffe et al. 2019, Sieker and Solstad 2022).[4] Under this diagnostic, participants are presented with utterances like those in (5), and asked to rate whether the (named) speaker is certain of the CC.

(5) a. Christopher: "Cole didn't discover that Julian dances salsa."
b. Christopher: "Did Cole discover that Julian dances salsa?"
c. Christopher: "Perhaps Cole discovered that Julian dances salsa."
d. Christopher: "If Cole discovered that Julian dances salsa, Logan will be joyful."
Projection question: Is Christopher certain that Julian dances salsa?

We assume, following Tonhauser et al. (2018) and Degen and Tonhauser (2022), that judgments of speaker certainty about the embedded content reflect speaker commitment to that content, that is, projection. If a participant interprets utterances like (5a–d) in a way that the speaker (here, Christopher) is certain of the CC, the CC is assumed to project. If a participant does not take the speaker to be certain of the CC, the CC is taken to not project.

### 2.1. Methods

**Participants.** We recruited 250-300 participants for each of the 12 experiments. Participants for one experiment were recruited on Amazon's Mechanical Turk platform. These participants were required to have U.S. IP addresses and at least 99% of previously approved HITs. Participants for the remaining experiments were recruited on Prolific. These participants were required to reside in the US, to be born in the US, to have English as their first language, and to have an approval rating of at least 99%. See Supplement D (in the repository linked to in footnote 3) for further information about the participants.

**Materials.** The target sentences consisted of the 400 combinations of the 20 clause-embedding predicates in (4) with 20 embedded clauses (provided in Supplement A). As mentioned above, the predicates were embedded under negation in Exps. 1, under polar questions in Exps. 2, under the epistemic possibility modal *perhaps* in Exps. 3, and in conditional antecedents in Exps. 4, for a total of 400 target stimuli in each of the four sets of experiments.

[3]The experiments, data and analysis scripts, as well as supplements referred to in this paper can be found in the following GitHub repository: `https://github.com/judith-tonhauser/CommitmentBankPlus`.

[4]For other diagnostics of projection see, e.g. Smith and Hall 2011, Xue and Onea 2011, and Tonhauser, Beaver, Roberts, and Simons 2013; and discussion in Tonhauser et al. 2018.

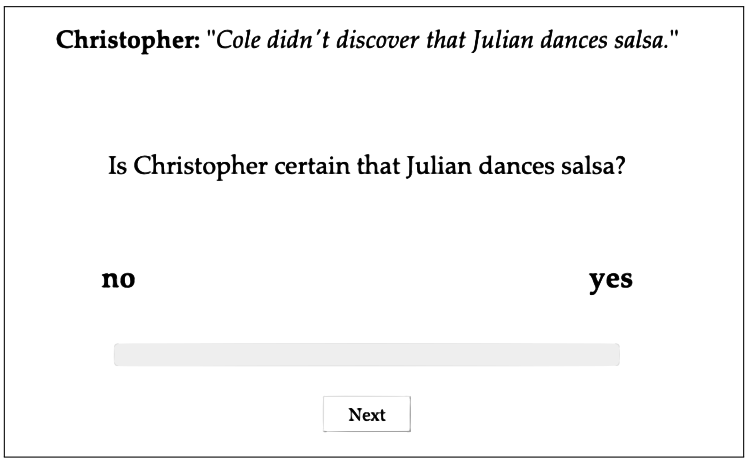


Figure 1: A sample trial in Exps. 1. In the corresponding trials in the other experiments, participants were presented with an utterance with a different entailment-cancelling operator.

To assess whether participants were attending to the task, each experiment included six control stimuli. For details on the six control stimuli, see Supplement C.

Each participant saw a random set of twenty-six stimuli: Each set contained one target stimulus for each of the 20 clause-embedding predicates (each with a unique complement clause) and the same six control stimuli.[5] Trial order was randomized.

**Procedure.** Participants were asked to imagine that they are at a party and that, when walking into the kitchen, they overhear somebody say something to somebody else. On each trial, they read an utterance and gave a response to the 'certain that' question on a slider marked 'no' (coded as 0) at one end and 'yes' (coded as 1) on the other. A sample trial from Exps. 1 is shown in Figure 1.

Following Tonhauser et al. (2018), higher ratings of speaker certainty could reflect one of two things: On the one hand, higher ratings could reflect greater speaker commitment towards the CC, and therefore greater projection. This assumes that speaker commitment is interpreted in a gradient way. Conversely, higher certainty ratings could reflect the probability that an interpreter takes the speaker to be committed to the CC. On this interpretation, speaker commitment may be a binary, categorical property and projection variation is a result of the uncertainty about possible interpretations. In this paper, we remain agnostic about the underlying interpretation of projectivity as a gradient property, though our discussion of projection variability will be in line with the first interpretation (for discussion, see also Grove and White 2023).

After completing the experiment, 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.

**Data exclusion** was done based on self-declared non-native speaker status and other criteria given in the Supplements D. The data from 2,682 participants entered into the analysis.

[5]Each participant saw their set of 26 stimuli twice, once in the projection block and once in the at-issueness block. Block order was randomized. As mentioned above, we focus on the projection ratings here.

## 2.2. Results and discussion

We first address the question of whether there is by-operator variation (section 2.2.1) and then the question of whether there is by-predicate variation in the observed by-operator variation (section 2.2.2). Finally, in section 2.2.3, we relate the results to the result of Degen and Tonhauser 2022, that projection from under polar questions does not categorically distinguish factive and non-factive predicates.

### 2.2.1. By-operator variation

Figure 2 shows the mean certainty ratings by entailment-canceling operator, aggregating over the clause-embedding predicates. As shown, there is projection variability by operator: The CCs of the clause-embedding predicates were relatively more projective when embedded in the antecedent of a conditional than in a polar interrogative, where they were relatively more projective than when they were embedded under negation or the epistemic modal *perhaps*.

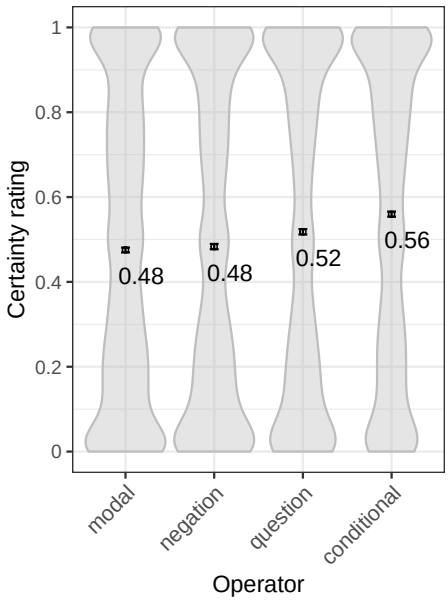


Figure 2: Mean certainty ratings by operator. Error bars indicate 95% bootstrapped confidence intervals and violin plots indicate the kernel probability density of the ratings.

These generalizations are supported by a post-hoc pairwise comparison of the estimated means for each entailment-canceling operator using the `emmeans` package (**?**) in R (**?**). The input to the pairwise comparison was a Bayesian mixed-effects beta regression model that was fit using the `brms` package (**?**) with weakly informative priors. The model predicted certainty ratings[6] from a fixed effect of entailment-canceling operator (with treatment coding and 'modal' as the reference level) and included a random by-predicate intercept.[7] The output of the pairwise comparison were 95% highest density intervals (HDIs) of estimated marginal mean differences

[6]To model the certainty 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.
[7]See Supplement E for more details on the model.

between each of the operators. We assume that two operators differ in certainty ratings if the HDI of their pairwise comparison does not include 0.

Table 1 provides the output of the pairwise comparison on a logit scale. As shown, the analysis suggests differences between each pair of operators, though the estimated marginal mean differences differ across the pairs. That is, certainty ratings are higher for CCs embedded in conditional antecedents than those embedded under polar questions; certainty ratings for CCs embedded under polar questions are higher than those for CCs embedded under negation; and finally certainty ratings for CCs embedded under negation are higher than those for CCs embedded under the epistemic modal *perhaps*.

| contrast | estimate | lower 95% CI | upper 95% CI |
|---|---|---|---|
| conditional - negation | 0.21 | 0.18 | 0.24 |
| conditional - question | 0.13 | 0.10 | 0.16 |
| modal - conditional | -0.26 | -0.29 | -0.23 |
| modal - negation | -0.05 | -0.08 | -0.02 |
| modal - question | -0.14 | -0.17 | -0.11 |
| negation - question | -0.09 | -0.12 | -0.06 |

Table 1: Output of the pairwise comparison of entailment-canceling operators. The 'contrast' column identifies the pair of entailment-canceling operators, the 'estimate' column the estimated marginal mean difference, and the 'lower 95% CI' and 'upper 95% CI' columns provide the lower and upper bounds of the HDIs.

These results suggest that certainty ratings for the CCs of of the English clause-embedding predicate we investigated vary by entailment-canceling operator. In contrast to Sieker and Solstad (2022) for German, the results of our experiments do not suggest that projection is strongest from under negation. Recall, however, that they only investigated projection of the CCs of (semi-)factive predicates. Since there is by-predicate variation in the effect of entailment-canceling operator on projection (as we show in the next section), this difference between the results of their experiment and ours might very well be due to the types of predicates investigated. Finally, the differences in mean certainty ratings between the four entailment-canceling operators are very small. This suggests that, when abstracting away from individual predicates and contents, projection from under the four entailment-canceling operators is very similar. In other words, when abstracting away from individual contents, the family of sentences really are a family.

#### 2.2.2. By-predicate variation in the effect of entailment-canceling operator

Figure 3 shows mean certainty ratings by entailment-canceling operator for the 20 predicates, with predicates ordered by their overall mean certainty rating. As shown, there is by-operator projection variation for the CCs of all 20 clause-embedding predicate, that is, there is not a single predicate for which there is no by-operator variation. Further, the effect of operator differs between predicates. Examining the observed variation suggests that it does not align with widely assumed predicate classes.

For instance, the five (semi-)factive predicates exhibit four different patterns: First, the CC of *be annoyed* projects most from under questions, less from under negation, followed by condi-


Figure 3: Mean certainty ratings by predicate ((semi-)factive, non-factive) and operator (modal, negation, polar questions, conditional antecedents) with 95% bootstrapped confidence intervals. Violin plots indicate kernel probability density of the individual participants' ratings. Facets for predicates are ordered by the predicate mean certainty rating (aggregating across operators). Below each facet, a line spanning two operators indicates a non-zero difference according to the pairwise comparison of operators. The linetype indicates whether the difference is at least 1 (solid line: —), below 1 but at least 0.5 (dashed line: – –), or below 0.5 but above 0 (dotted line: . . . )

tionals, and least from under the modal *perhaps* (Q > N > C > M). Second, the CC of *know* projects most from under questions, less from conditionals and negation, and least from under *perhaps* (Q > {N, C} > M). The CCs of *discover* and *see* exhibit a third pattern: They project most from under questions and conditionals, less from under negation, and least from under *perhaps* ({Q, C} > N > M). Finally, the CC of *reveal* projects most from conditionals, less from questions, and least from under negation and *perhaps* (C > Q > {N, M}). Thus, contrary to what Karttunen 1971 proposed, there is no predicate whose CC projects uniformly from

under all four entailment-canceling operators (as he suggested for factive predicates) and the purported semi-factive predicates *discover* and *reveal* do not project more from under negation than the other three entailment-canceling operators.

We also observe by-operator projection variation for non-factives predicates. Some of this variation aligns with that observed for factive predicates: For instance, the CC of *inform* exhibits the same pattern as the CC of *know*, and the CC of *acknowledge* exhibit the same pattern as that of the CC of *discover*. Other non-factives predicates exhibit other patterns: For instance, the CCs of *admit*, *confess* and *announce* project most from the antecedents of conditionals than the other three operators.

These generalizations are supported by the statistical analysis. We fitted Bayesian mixed-effects beta regression models for each the by-predicate subset of the data containing (using *brms*, Bürkner 2017; in R, **?**). Each of the 20 models used low-information priors and estimated the effect of the 'operator' (fixed effect, treatment-coded with reference-level 'modal') on (scaled) certainty ratings, with by-item random intercepts and random operator slopes.[8] For each predicate, the differences between operators were established by a pairwise comparison of posterior estimates of marginal means for the four operators (using *emmeans*, Lenth, Bolker, Buerkner, Giné-Vázquez, Herve, Jung, Love, Miguez, Riebl, and Singmann 2024). Here we report those differences between posterior samples where the 95% HDI does not include 0. These are indicated in Figure 3, where non-zero differences between two operators are indicated by lines spanning the two operators below each predicate facet.

More generally, we found by-operator differences for the CC of each of the 20 clause-embedding predicates, and that the effect of operator differs between predicates. Therefore, our findings align with those of Smith and Hall 2014, who also observed by-expression variation in the effect of operator. However, while they found that the CC of *know* projects more from under negation than the antecedent of a conditional, we did not find a difference here. We hypothesize that this difference is due to the difference in projection diagnostic used.

Our results, however, differ from those of Sieker and Solstad 2022. While their work did not find differences in by-operator projection variation between factive and semi-factive predicates, our results suggested five different patterns of by-operator variation for the five (semi)factive predicates. As Sieker and Solstad 2022 also used the 'certain that' diagnostic for projection, this difference in results is not likely due to the diagnostic. Other factors that varied between our experiments are the language under investigated (German vs. English), the clause-embedding predicates investigated, and the CCs that the predicates were paired with. Future research will need to establish which of these factors are implicated in the observed differences.

### 2.2.3. Factive vs. non-factives predicates

Lexical approaches to projection assume that factive predicates are ones that presuppose the CC, while the CC of non-factives predicates is not presupposed (e.g., Kiparsky and Kiparsky 1970, Karttunen 1971, Schlenker 2010, Abrusán 2011).[9] Because presuppositions are assumed to project from under entailment-canceling operators, this definition predicts that factive pred-

[8]Further details about the models, their outputs, and the analysis are provided in Supplement F.

[9]Some of these works additionally assume that the CC of factive predicates is entailed.

icates are distinguished from non-factives by the projection of their CCs: The CCs of factive predicates are expected to be categorically more projective than those of non-factives.

This expectation was investigated in Degen and Tonhauser 2022 based on the 20 clause-embedding predicates in (4) embedded in polar questions. Contrary to expectation, Degen and Tonhauser's (2022) Exps. 1 found that the CCs of the five (semi-)factive predicates varied in projection, that the CCs of the 15 non-factive predicates were projective compared to the non-projective main clause contents, and that the CCs of some non-factives were as projective, or even more projective, than those of some factive predicates. In short, projection of the CC from under polar questions did not categorically distinguish factive from non-factives predicates. Further support for this result came from the analysis of projection ratings in three additional datasets, namely the CommitmentBank (De Marneffe et al. 2019), the VerbVeridicality dataset (Ross and Pavlick 2019), and the MegaVeridicality dataset (White and Rawlins 2018).

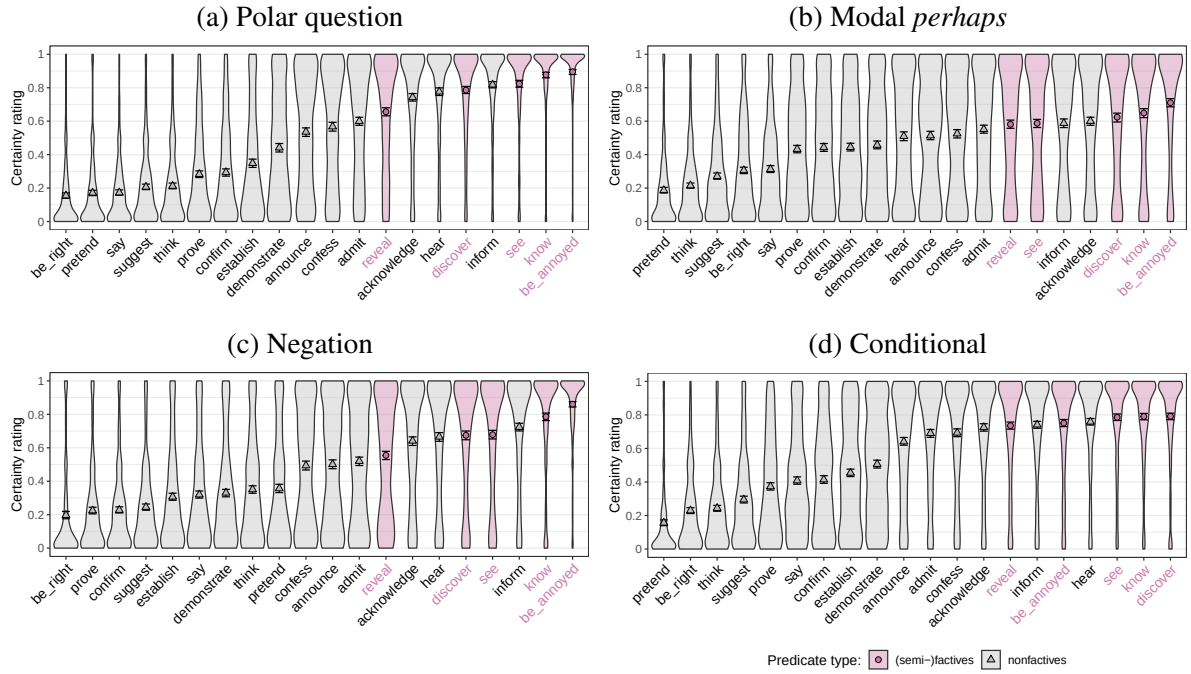


Figure 4: Mean certainty ratings by predicate for (a) polar questions, (b) the modal *perhaps*, (c) negation, (d) conditional antecedents. Error bars indicate 95% bootstrapped confidence intervals. Violin plots indicate kernel probability densities of the individual participants' ratings.

The experiments reported on in this paper replicate this result. As shown in Fig. 4a, there is variation between the CCs of the five factive predicates in the polar question condition, and projection from under polar questions does not categorically distinguish factive from non-factive predicates. Furthermore, our results suggest that this result can be extended to the three other entailment-canceling operators as well. As shown in Figs. 4b-4d, there is variation in the projection of the CCs of the five factive predicates from under *perhaps*, negation, and conditional antecedents, and projection ratings in these conditions do not show a categorical difference between factive and non-factive predicates here, either. Instead, we see a gradient effect of predicate on projection ratings in all four conditions. These results lend further support

to the conclusion of Degen and Tonhauser 2022 that there is, to date, no empirical evidence for a coherent class of factive predicates.

#### 2.2.4. Summary

The results of our experiments suggest that there is little by-operator variation when aggregating over clause-embedding predicates, but that the CCs of different clause embedding predicates exhibit by-operator projection variation. Crucially, the effect of operator on projection differs by predicate, but not in ways that align with prior claims about differences between factive and semi-factive predicates (e.g., Karttunen 1971). Finally, the results of our experiments provide further support for the results of Degen and Tonhauser 2022, who did not find empirical support for a class of factive predicates based on the projection of the CC from under polar questions. Our experiments suggest that projection of the CC from under negation, the antecedent of conditionals and epistemic modals also do not provide empirical support for a natural class of factive predicates.

Before discussing the methodological and theoretical implications of these results in section 4, we provide converging evidence from a different dataset in section 3.

## 3. Converging evidence for the by-operator variation

This section provides converging evidence for the by-operator variation based on the MegaVeridicality dataset (White and Rawlins 2018). This dataset contains projection ratings for the CCs of 517 English clause-embedding predicates. The stimuli that participants rated consisted of combinations of these predicates with what White and Rawlins (2018) referred to as 'low content arguments', as shown in (6) for *know*. The predicates were embedded under negation in stimuli like (6a), in the antecedent of a conditional and a polar question in stimuli like (6b), and under negation, in the antecedent of a conditional, and in a question in stimuli like (6c). To assess projection, participants were asked to respond to the question *Did that thing happen?* for stimuli like (6a) and to respond to the question posed by stimuli like (6b) and (6c). The response options were 'yes', 'maybe or maybe not', and 'no'.

(6) a. Somebody didn't know that a particular thing happened. Did that thing happen?
b. If somebody knows that a particular thing happened, did that thing happen?
c. If somebody didn't know that a particular thing happened, did that thing happen?

To investigate by-operator projection variability, we recoded 'yes' responses as 1, 'no' responses as -1, and 'maybe or maybe not' responses as 0. We then calculated the mean projection rating for 16 predicates under the three types of operator combinations shown in (6). The 16 predicates include ones also used in our experiment as purported factives (*be annoyed, know, reveal*) and semi-factives (*discover, see*), and eleven further predicates commonly characterized as factive or semi-factive (*amuse, find out, forget, learn, love, notice, realize, recognize, regret, remember, understand*).

The mean projection ratings by embedding context and predicate are presented in Figure 5, which shows that effect of embedding context on projection differs by predicate.

For instance,[10]

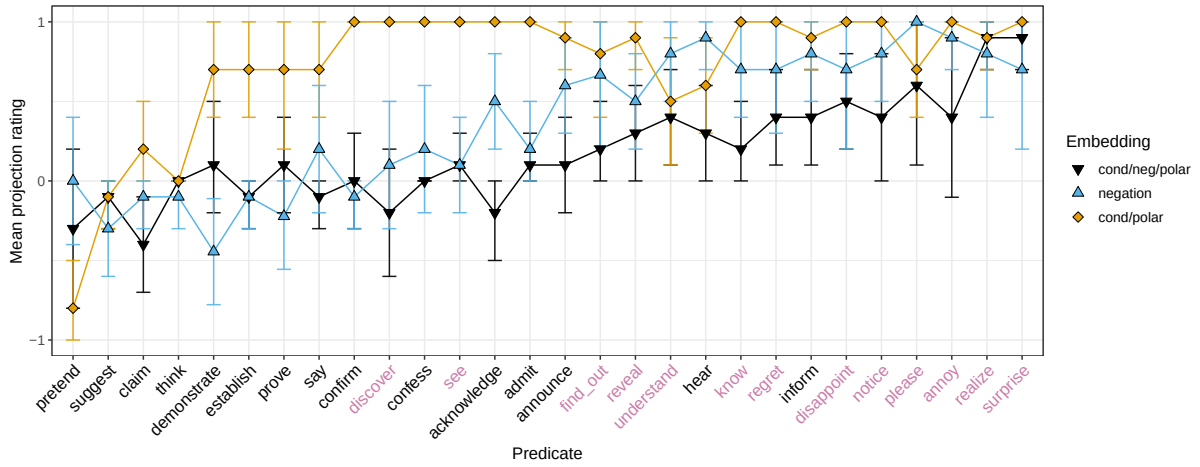


Figure 5: Mean projection ratings by embedding context and predicate in MegaVeridicality.

It is worth noting that the task assesses global projection only in case of (6a), but not (6b+c), where the question *did it happen?* is embedded in the conditional consequent. It assesses whether the projective inference holds in the conditional local context. Nevertheless, the data provides supporting evidence for by-predicate variation in the effect of operators on projection. Crucially, the by-predicate variation in the effect of negation in the global context is different from the by-predicate variation in the effect of negation in the conditional context.

## 4. General discussion

In this section, we point out methodological implications of our results (section 4.1), discuss whether contemporary projection analyses can capture the observed variation (section 4.2), and speculate about lexical differences between the clause-embedding predicates that might predict the by-operator variation observed (section 4.3).

### 4.1. Methodological implications

The results of the experiments reported on in section 2 suggest that there is little by-operator projection variation when aggregating observations for the CCs of the clause-embedding predicates, but that there is by-operator variation that cannot be neglected when we do not aggregate. These results have two methodological implications. First, when initially investigating the projection of a content (or teaching projection to students), the family-of-sentences can indeed be treated as a family, as standard textbooks assume. However, it is advisable to apply the diagnostic with all four operators and attend to possible by-operator projection variation.

A second methodological implication of the results of our experiments is that research on projection should take into account possible by-operator variation when choosing which entailment-canceling operator to use. Further, results about projection variation should be relativized

[10]Give some examples of differences here, but first double check if lines are labeled correctly. Also angle labels on x-axis.

to the entailment-canceling operator under which the expressions and contents were embedded, as they may very well depend on this choice. For instance, two contents that exhibit projection variation when embedded under polar questions might not exhibit such differences from under a different entailment-canceling operator, and two contents that do not exhibit different projection across negation might exhibit variation when embedded under some other operator.

### 4.2. Theoretical implications

This section discusses whether contemporary analyses can capture the observed by-predicate projection variation and the interactions between predicate and entailment-canceling operator in projection ratings.

#### 4.2.1. Heim 1983 and van der Sandt 1992

On the analyses in Heim 1983 and van der Sandt 1992, the CCs of factive predicates are lexically specified as presupposed, in contrast to the CCs of non-factives predicates. Presupposed CCs project to the global context, except when that would produce an inconsistency, in which case they are accommodated to the local context of the operator.

These analyses do not predict the observed by-predicate variation, for several reasons. One reason is variation among the factive predicates: For instance, when embedded under polar questions, negation, or *perhaps,* the CC of *be annoyed* is more projective than that of *discover*, which in turn is more projective than that of *reveal*. (In conditional antecedents, the order of *be annoyed* and *discover* is reversed, see Figure 4). This variation is not predicted because the CCs of factive predicates are invariably specified as presupposed, with no mechanism to predict projection variation between factive predicates. (Recourse to local accommodation is not possible because the stimuli in our experiments were presented in minimal contexts that were not inconsistent with the CCs.) A second reason is that the CCs of some non-factives predicates (e.g., *inform, acknowledge*) are just as projective as those of some factive predicates, or even more projective. Analyses like those in Heim 1983 and van der Sandt 1992 fail to make predictions about the projection of the CCs of non-factives predicates. Thus, as discussed in detail in Degen and Tonhauser 2022, the factive/non-factive distinction is not sufficiently fine-grained to adequately predict the observed by-predicate projection variation.

The analyses in Heim 1983 and van der Sandt 1992 also do not predict differential effects of entailment-cancelling operators on projection. This is because negation and conditional antecedents are given a semantics that derives their behavior as presuppositional holes, in the sense of Karttunen 1971. For instance in Heim 1983, presupposition triggers under negation or in conditional antecedents are evaluated relative to a local context that is equal to the global context, with the consequence that presuppositions are not expected to project differently from under negation or conditional antecedents. While the analysis does not explicitly address epistemic possibility modals or polar questions, we might expect that they would be treated as presuppositional holes along the same lines, therefore also not predicting any by-operator projection variation, or by-predicate variation in the effect of entailment-canceling operators.

#### 4.2.2. Abrusán 2011 and Simons, Beaver, Roberts, and Tonhauser 2017

Abrusán 2011 and Simons et al. 2017 do not assume that presuppositions are lexically specified.

In Abrusán 2011, a lexical entailment of a (possibly embedded) sentence is a presupposition, if it is about a time that is not the event time of the matrix predication and it is not at-issue with respect to the Question Under Discussion (QUD, see Roberts 1996, 2012) addressed by the utterance. For instance, the CC of B's utterance in (7), that Phil's ballet class is canceled, is predicted to be a presupposition (and therefore to project) because it is a lexical entailment of the modal prejacent and not at-issue with respect to A's interrogative utterance.

(7) Adapted from Simons et al. 2017: 188
Context: It's early on Saturday morning. A and B are talking about their son.
A: Why is Phil up already?
B: Perhaps he forgot that his ballet class is canceled today.

In Simons et al. 2017, the CC of a clause-embedding predicate projects if it is entailed by the Current Question of the utterance (where the Current Question is the question that is congruent with the utterance).[11] In (7), the Current Question of B's utterance might be the set of propositions {Phil forgot that his ballet class is canceled today, Phil is aware that his ballet class is canceled today}. If so, the Current Question entails that Phil's ballet class is canceled today, and the CC may therefore project under Simons et al.'s 2017 analysis.

Under the analyses in Abrusán 2011 and Simons et al. 2017, CCs that are entailed based on the literal content of an utterance may or may not project, depending on whether they contribute to the main point of the utterance. This opens up the possibility for projection variation among veridical predicates (including those often characterized as factive), representing an improvement compared to the lexical specification of factivity assumed in Heim 1983 and van der Sandt 1992. However, these accounts still do not predict the observed by-predicate variation. First, to make explicit predictions about the observed by-predicate variation, these types of analyses would need to make explicit assumptions about how different predicates contribute to how the question under discussion is chosen. Second, the analyses do not make systematic predictions for non-entailed CCs, and hence they do not predict that the CCs of some non-factives predicates project at least as much or even more than those of some factive predicates.

Both of these analyses set aside the effect of entailment-cancelling operators in determining the main point of an utterance. Therefore they do not make predictions about by-operator differences or by-predicate variation in the effect of entailment-canceling operators.

#### 4.2.3. Schlenker 2021

Under the view put forward in Schlenker 2021, the CC of a sentence $S$ like (8a), is presupposed in a context $c$, if the CC is an epistemic precondition on the global meaning of $S$. Formally, Schlenker defines presuppositions relative to their local context (§6.2). For instance, in (8a), the negative prejacent $S'$, that is (8b), is presupposed in the local context under negation $c'$, if two conditions are met: i) $S'$ contextually entails the CC relative to $c'$; and ii) If we consider "a generic agent" who believes the propositions in $c'$ and who has now learned about the truth of S′, then the probability that this generic agent already believed the CC is above a

[11] The Current Question is defined in Simons et al. 2017: 194 as follows: "The Current Question for an utterance is a privileged subset of the focal alternative set of the uttered sentence (given a structural analysis of that sentence, including focus marking)" which meets the conditions that "i) the proposition expressed is a member of the Current Question and ii) the Current Question has at least one additional member."

contextual threshold $\alpha$; more colloquially, condition ii) requires that the generic agent "typically antecedently believes" the CC (p.6) upon interpreting S$'$ in $c'$. Based on Heim 1983 and Schlenker 2009, the the local context under negation $c'$ is assumed to be identical to the global context $c$. Therefore, epistemic preconditions relative to $c'$ will also apply to the global context.

Applied to (8), Schlenker's 2021 analysis predicts that the CC is presupposed if i) (8b) contextually entails that Julian dances salsa, and ii) if a generic agent would typically antecedently believe that Julian dances salsa upon interpreting (8b) in the minimal contexts we provided to our participants.

(8) a. Cole doesn't know that Julian dances salsa.
b. Cole knows that Julian dances salsa.

Condition i) is met under the assumption that the CC of (8b) is an entailment. Schlenker 2021 also assumes the condition ii) is met: "in many cases, one's knowledge of facts will precede one's knowledge of [Cole's] beliefs about them... believing that [Julian dances salsa] is often an epistemic precondition for believing that" Cole knows that Julian dances salsa (p.6). One might, however, challenge this assumption on the basis of the corpus study presented in Spenader 2002, which showed that the CCs of the majority of the utterances of sentences with factive verbs (namely 81 out of 109) had to be accommodated (i.e., were not contextually entailed). In other words, utterance of sentences with the factive predicates investigated by Spenader (2002), which included *know*, were "generally used to communicate information the speaker thought was hearer-new" (p.99). This result might therefore suggest that one cannot assume that a generic agent typically antecedently believes the CC of *know*.

Schlenker's 2021 analysis does not incorporate differential effects of entailment-cancelling operators on projection. As discussed for Heim 1983, above, the local contexts under negation and in conditional antecedents are both assumed to be the global context (and similarly in Schlenker 2009). Therefore, the analysis does not make predictions about by-operator projection variation or about by-predicate variation in the effect of entailment-canceling operator.

Does the account capture the observed by-predicate variation? Because the probabilities about previous beliefs and the threshold $\alpha$ are assumed to potentially vary based on the context, the formal proposal does not in principle preclude the possibility of by-predicate projection variation. However, the subjective conditional probabilities associated with expressions and the contextual threshold $\alpha$ are taken as given in the account, and therefore do not offer explicit or systematic predictions about projection ratings for different expressions. We, therefore, assume that it is an open, empirical question which predicates are such that the probability of a generic agent antecedently believing the CC is above the contextual threshold $\alpha$ in the minimal contexts we provided our participants (and, of course, what that threshold might be).

Throughout the discussion, however, Schlenker 2021 appears to assume that there are two classes of predicates: those where the probability is usually above the threshold (including *know, inform*, and *announce*), and those where it is not (including *demonstrate* and *establish*; see p.12 and appendix I). As this division does not fall along the lines of traditionally assumed classes of (non-)factives, one advantage of Schlenker's 2021 analysis over the analyses reviewed in sections 4.2.1 and 4.2.2 is that it predicts the projection of the CCs of (certain) non-factive predicates (modulo the open questions about condition ii)).

Based on these assumptions, the analysis correctly predicts that the CCs of *know, inform*, and *announce* are more projective than the CCs of *demonstrate* and *establish*, by virtue of the CCs of the former being usually presupposed, in contrast to the CCs of the latter. It is not clear, however, that the analysis is able to predict the observed variation, because the analysis – even though it does not divide predicates into factive and non-factives ones – nevertheless imposes a binary, categorical distinction between predicates. Crucially, the by-predicate variation reported in Sections 2.2.2+2.2.3 is not captured by an analysis that does not make more fine-grained distinctions between the meanings of clause-embedding predicates, as discussed extensively in Degen and Tonhauser 2022. First, there is projection variation between the supposedly presupposed CCs (that of *know* is more projective than the CC of *inform*, which is more projective than that of *announce*, under all four operators). Furthermore, as discussed in Degen and Tonhauser 2022, the CCs of the supposedly nonpresupposed CCs (of *establish* and *demonstrate*) are projective when compared to nonprojective main clause content. In the experiments reported on in section 2, the mean certainty rating of the CC of *announce*, whose CC is assumed to possibly be presupposed, is 0.55, and that of *demonstrate*, whose CC is not assumed to be presupposed, is 0.43 (both means are aggregated across entailment-canceling operators). It is not clear whether this particular difference in mean certainty rating motivates analyzing the CC of *announce* as presupposed in contrast to that of *demonstrate*.

#### 4.2.4. Interim summary

There is currently no projection analysis on the market that predicts the by-predicate and by-operator variation we observed in the experiments reported on in section 2. Degen and Tonhauser 2022 suggested that an empirically adequate analysis to capture the observed by-predicate variation requires consideration of "more fine-grained distinctions [between clause-embedding predicates] that are based on the lexical meaning and discourse use of clause-embedding predicates" (p. 585). Further, an account of the observed by-predicate variation in the effects of entailment-cancelling operators will need to take into account how semantic and pragmatic properties of predicate meanings interact with entailment-cancelling operators in varying ways. In the next section, we offer some hypotheses based on the data from our experiments.

### 4.3. Lexical patterns

Can the observed interactions between predicate and entailment-canceling operator be predicted from lexical semantic/pragmatic properties of the predicates? This is a pressing question for future research, to which our data offer some tentative answers.

We can find some initial generalizations over lexical properties, indicated in Figure 6, which gives the mean certainty ratings for the four operators by predicate, identifying four groups of predicates that show similar by-operator variation.

The non-veridical predicates *pretend* and *think* exhibit the 'Negation high' pattern, shown in panel (a) of Figure 6. These are the only predicates that are most projective under negation compared to all other operators. This generalization, supported by the statistical analysis in Section 2.2.2 (see Figure 3), could be related to assumptions that non-veridical doxastics like *think* can often be interpreted in comparison to a stronger veridical alternative (e.g., Heim 1991,

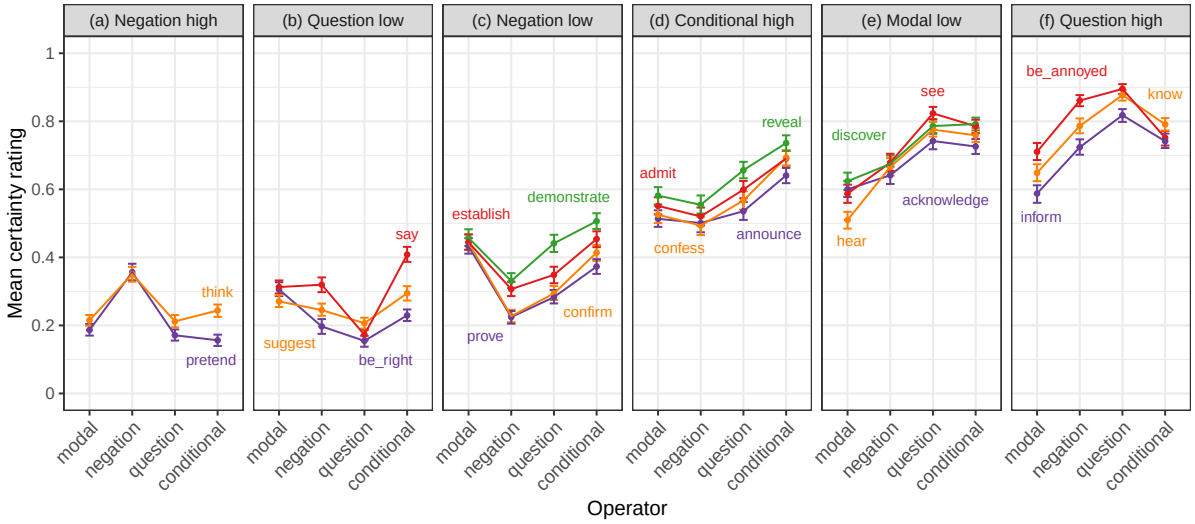


Figure 6: Mean certainty ratings by operator with 95% bootstrapped confidence intervals, for some groups of predicates ('predicate patterns').

Chemla 2008). We tentatively hypothesize that this can lead to an inference that the CC is false in upward-monotone contexts, but not under negation. A similar alternative seems to be salient for *pretend*, but not the non-veridical communicative predicates *say* or *suggest*.

These latter two predicates, along with *be right*, exhibit the 'Question low' pattern, shown in panel (b). They are the only predicates which are least projective in polar questions (again, compare Figure 3). We tentatively suggest that these communicative predicates can interact with the pragmatics of polar questions in a way that can lead to an inference of incredulity towards the CC. However, there are also some differences between these predicates. The veridical *be right* is most projective under modals (M $>$ C $>$ N $>$ Q), whereas *say* and *suggest* are most projective in conditional antecedents (C $>$ M, N $>$ Q).

The two u-shaped patterns in panels (c) and (d) are characterized by relatively low projection ratings with negation and high ratings with conditionals. While there is fine-grained variation with regards to which of the by-operator differences were statistically supported, (c) and (d) can be distinguished by their ratings for the modal. The change-of-state communication predicates *admit, announce, confess*, and *reveal*, exhibit relatively low projection ratings of the CC when embedded under *perhaps* (C $>$ Q $>$ M, N). In contrast, the inferential predicates *confirm, demonstrate, establish*, and *prove* exhibit relatively high ratings with *perhaps* (*confirm:* C $>$ M, Q $>$ N; *demonstrate:* C $>$ M, Q $>$ N; *establish:* C, M $>$ Q, N; *prove:* M $>$ C $>$ Q, N). For an explanation of the high conditional ratings, one might examine ways in which the discourse effect of a conditional interacts with the change-of-state component of these inferential and communicative predicates. Our tentative suggestion for the relatively low negation ratings is that these predicates can be interpreted relative to contextual assumptions that lead to a neg-raising type inference more readily than others (so that, for instance, not announcing $p$ amounts to communicating not $p$, or not proving $p$ amounts to inferring not $p$).

Finally, the two roof-shaped patterns in (e) and (f) are characterized by relatively high projection ratings with questions, and low ratings with *perhaps*. We can distinguish the two patterns

by their relative ratings for conditionals. The predicates *acknowledge, discover, hear,* and *see*, which are associated with a change of some informational state, show relatively high ratings with conditionals (Q, C > N > M). The ratings for conditionals are relatively lower for the predicates *be annoyed* (Q > N > C > M), and *inform*, and *know* (Q > N, C > M), whose CCs are among the most projective.

## 5. Conclusion

This paper investigated variation in projection from under the four entailment-canceling operators that have traditionally been used in the family-of-sentences diagnostic for projection, namely negation, polar questions, epistemic modals, and conditional antecedents. The results of our experiments suggest that the projection of the contents of the clausal complements of clause-embedding predicates varies across these operators. As discussed, there is currently no projection analysis on the market that is able to predict the observed by-predicate variation or the by-operator variation. The results of our experiments also extend a result of Degen and Tonhauser 2022, that projection ratings in polar questions do not categorically distinguish factive from non-factives predicates, to cases with negation, the epistemic possibility modal *perhaps*, and conditional antecedents. This results strengthens the conclusion of Degen and Tonhauser 2022 that there is, to date, no empirical evidence for a coherent class of factive predicates.

## References

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and philosophy 34*, 491–535.

Beaver, D. (2010). Have you noticed that your belly button lint colour is related to the colour of your clothing? In R. Bäuerle, U. Reyle, and T. Zimmermann (Eds.), *Presuppositions and Discourse: Essays offered to Hans Kamp*, pp. 65–99. Oxford: Elsevier.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software 80*(1), 1–28.

Chemla, E. (2008). An Epistemic Step for Anti-Presuppositions. *Journal of Semantics 25*(2), 141–173.

Chierchia, G. and S. McConnell-Ginet (1990). Meaning and grammar: An introduction to semantics. Publisher: MIT press Cambridge, MA.

de Marneffe, M.-C., C. D. Manning, and C. Potts (2012). Did It Happen? The Pragmatic Complexity of Veridicality Assessment. *Computational Linguistics 38*(2), 301–333. Publisher: MIT Press.

De Marneffe, M.-C., M. Simons, and J. Tonhauser (2019). The commitmentbank: Investigating projection in naturally occurring discourse. In *Proceedings of Sinn und Bedeutung*, Volume 23, pp. 107–124. Issue: 2.

Degen, J. and J. Tonhauser (2022). Are there factive predicates? An empirical investigation. *Language 98*(3), 552–591.

Djärv, K. and H. A. Bacovcin (2017). Prosodic effects on factive presupposition projection. *Semantics and Linguistic Theory 27*, 116–133. Publisher: Ithaca, NY: CLC Publications.

Djärv, K., J. Zehr, and F. Schwarz (2018). Cognitive vs. emotive factives: An experimental differentiation. In *Proceedings of Sinn und Bedeutung*, Volume 21, pp. 367–386. Issue: 1.

Grove, J. and A. S. White (2023). Factivity, presupposition projection, and the role of discrete knowledge in gradient inference judgments.

Heim, I. (1983). On the projection problem for presuppositions. In P. H. Portner and B. H. Partee (Eds.), *Formal Semantics: The Essential Readings*, pp. 249–260. John Wiley & Sons.

Heim, I. (1991). Artikel und Definitheit. In A. von Stechow and D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, pp. 487–535. Berlin: Mouton.

Karttunen, L. (1971). Some observations on factivity. *Research on Language & Social Interaction 4*(1), 55–69. Publisher: Taylor & Francis.

Kiparsky, P. and C. Kiparsky (1970). Fact. In M. Bierwisch and K. E. Heidolph (Eds.), *Progress in Linguistics*, pp. 143–173. The Hague: Mouton.

Lenth, R. V., B. Bolker, P. Buerkner, I. Giné-Vázquez, M. Herve, M. Jung, J. Love, F. Miguez, H. Riebl, and H. Singmann (2024). emmeans: Estimated Marginal Means, aka Least-Squares Means.

Lorson, A. (2018). The influence of world knowledge on projectivity.

Mahler, T. (2019). Does at-issueness predict projection? It's complicated! *North East Linguistic Society 49*, 245–255.

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

Roberts, C. (1996). Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. In J. H. Yoon and A. Kathol (Eds.), *Ohio State University Working Papers in Linguistics*, Volume 49. The Ohio State University, Department of Linguistics.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics 5*, 1–69.

Ross, A. and E. 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 (EMNLP-IJCNLP)*, pp. 2230–2240.

Schlenker, P. (2009). Local contexts. *Semantics and pragmatics 2*, 3–1.

Schlenker, P. (2010). Local contexts and local meanings. *Philosophical Studies 151*, 115–142.

Schlenker, P. (2021). Triggering presuppositions. *Glossa: a journal of general linguistics 6*(1).

Sieker, J. and T. Solstad (2022). Projective variability of (semi) factive verbs in family of sentence contexts: A rating study. In *Proceedings of the 23rd Amsterdam Colloquium*.

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

Smith, E. A. and K. C. Hall (2011). Projection diversity: Experimental evidence. *2011 ESSLLI Workshop on Projective Content.*, 156–170.

Smith, E. A. and K. C. Hall (2014). The relationship between projection and embedding environment. In *Proceedings of the 48th Meeting of the Chicago Linguistics Society*. Citeseer.

Smithson, M. and J. Verkuilen (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods 11*(1), 54. Publisher: American Psychological Association.

Spenader, J. (2002). *Presuppositions in Spoken Discourse*. PhD Thesis, Stockholm University.

Stevens, J., M.-C. de Marneffe, S. R. Speer, and J. Tonhauser (2017). Rational use of prosody

predicts projection in manner adverb utterances. In *CogSci*.

Syrett, K. and T. Koev (2015). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics 32*, 525–577.

Tonhauser, J. (2016). Prosodic cues to presupposition projection. In *Semantics and Linguistic Theory*, Volume 26, pp. 934–960.

Tonhauser, J., D. Beaver, C. Roberts, and M. Simons (2013). Toward a taxonomy of projective content. *Language 89*(1), 66–109.

Tonhauser, J., D. I. Beaver, and J. Degen (2018). How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics 35*(3), 495–542.

van der Sandt, R. A. (1992). Presupposition projection as anaphora resolution. *Journal of semantics 9*(4), 333–377.

Warriner, A. B., V. Kuperman, and M. Brysbaert (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods 45*(4), 1191–1207.

White, A. S. and K. Rawlins (2018). The role of veridicality and factivity in clause selection. In *48th Meeting of the North East Linguistic Society*, Volume 3, pp. 221–234.

Xue, J. and E. Onea (2011). Correlation between presupposition projection and at-issueness: An empirical study. In *Proceedings of the ESSLLI 2011 workshop on projective meaning*, pp. 171–184.

## A. 20 clauses

The contents of the following 20 clauses, which realized the complements of the 20 clause-embedding predicates, were investigated in Exps. 1-4:

1. Mary is pregnant.
2. Josie went on vacation to France.
3. Emma studied on Saturday morning.
4. Olivia sleeps until noon.
5. Sophia got a tattoo.
6. Mia drank 2 cocktails last night.
7. Isabella ate a steak on Sunday.
8. Emily bought a car yesterday.
9. Grace visited her sister.
10. Zoe calculated the tip.
11. Danny ate the last cupcake.
12. Frank got a cat.
13. Jackson ran 10 miles.
14. Jayden rented a car.
15. Tony had a drink last night.
16. Josh learned to ride a bike yesterday.
17. Owen shoveled snow last winter.
18. Julian dances salsa.
19. Jon walks to work.
20. Charley speaks Spanish.

## B. Consequents for conditional target stimuli in Exps. 4

We created 20 consequent clauses for the three experiments in which the 20 clauses were embedded in the antecedent of a conditional (Exps. 4). Each of the 20 clauses was paired with a unique consequent clause, as shown in the list below. To minimize the variability of the effect of the contents of these consequent clauses on the projection of the contents of the 20 complement clauses, the consequent clauses all consist of a uniquely named subject and an adjectival predication in the future tense (*will be*), and the adjectives all denote an emotion. We selected the 20 emotion-denoting adjectives based on the valence and arousal values reported in Warriner, Kuperman, and Brysbaert 2013: 10 of the adjectives had a positive valence, and 10 had a negative valence; all 20 adjective had an arousal value between 4.7 and 6.5.

1. . . . that Mary is pregnant, Esther will be mad.
2. . . . that Josie went on vacation to France, Arnold will be frustrated.
3. . . . that Emma studied on Saturday morning, Liam will be proud.
4. . . . that Olivia sleeps until noon, Elijah will be embarrassed.
5. . . . that Sophia got a tattoo, Ariel will be giddy.
6. . . . that Mia drank 2 cocktails last night, Mariela will be worried.
7. . . . that Isabella ate a steak on Sunday, Liz will be delighted.
8. . . . that Emily bought a car yesterday, Kate will be excited.
9. . . . that Grace visited her sister, Henry will be surprised.
10. . . . that Zoe calculated the tip, Alex will be astonished.
11. . . . that Danny ate the last cupcake, Harper will be disgusted.
12. . . . that Frank got a cat, Lucas will be grouchy.
13. . . . that Jackson ran 10 miles, Kayla will be cheerful.
14. . . . that Jayden rented a car, Brittany will be furious.
15. . . . that Tony had a drink last night, Victoria will be ashamed.
16. . . . that Josh learned to ride a bike yesterday, Mason will be envious.
17. . . . that Owen shoveled snow last winter, Bianca will be jealous.
18. . . . that Julian dances salsa, Logan will be joyful.
19. . . . that Jon walks to work, Caleb will be suspicious.
20. . . . that Charley speaks Spanish, Jay will be happy.

## C. Control stimuli in Exps. 1-3

The control stimuli in Exps. 1-3 were the contents of main clauses. In Exps. 1q, 2q and 3q, the control stimuli consisted of the polar questions in (1). The non-restrictive relative clauses (NRRCs), given in parentheses in (1), were included in Exps. 2q and 3q, where at-issueness was measured with an assent diagnostic. The control stimuli here consisted of two clauses (like the target stimuli), to allow the relevant speaker to assent with one of two clauses.

(1) Sentences for control stimuli in in question embedding experiments (Exps. 1q, 2q and 3q)
   a. Do these muffins (, which are really delicious,) have blueberries in them?
   b. Does this pizza (, which I just made from scratch,) have mushrooms on it?
   c. Was Jack (, who is my long-time neighbor,) playing outside with the kids?
   d. Does Ann (, who is a local performer,) dance ballet?
   e. Were John's kids (, who are very well-behaved,) in the garage?

f. Does Samantha (, who is really into fashion,) have a new hat?

We expected participants to give low responses on the 'certain that' diagnostic for the control stimuli in (1), indicating that the speaker is not certain of the main clause content, because main clause content is hypothesized to not project out of polar questions. These expectations were borne out, as shown in the third column of Table 2 for Exps. 1q, 2q, and 3q. We also expected participants to give low responses on the at-issueness diagnostics for the control stimuli in (1), indicating that the main clause content is at-issue. These expectations were borne out for Exps. 1q and 2q, as shown in the fourth column of Table 2, but the ratings were higher than expected for Exp. 3q. See below for discussion.

In the remaining experiments, the control stimuli consisted of the positive declarative variants of (1) given in (2). The NRRCs in parentheses were realized in Exps. 2 and 3 for the reason explained above, as well as in Exp. 1c, to make the control stimuli more similar to the target stimuli (which also consisted of two clauses, namely the antecedent and the consequent).

(2) Sentences for control stimuli in negation, modal and conditional embeddings
   a. These muffins (, which are really delicious,) have blueberries in them.
   b. This pizza (, which I just made from scratch,) has mushrooms on it.
   c. Jack (, who is my long-time neighbor,) was playing outside with the kids.
   d. Ann (, who is a local performer,) dances ballet.
   e. John's kids (, who are very well-behaved,) were in the garage.
   f. Samantha (, who is really into fashion,) has a new hat.

We expected participants to give high responses on the 'certain that' diagnostic for the control stimuli in (2), indicating that the speaker is certain of the main clause content, because speakers are hypothesized to be committed to the asserted main clause content. These expectations were borne out, as shown in the third column of Table 2 for the 'n', 'm' and 'c' variants of Exps. 1, 2, and 3. We expected participants to give low responses on the at-issueness diagnostics for the control stimuli in (2), indicating that the main clause content is at-issue. These expectations were borne out for the 'n', 'm', and 'c' variants of Exps. 1, as shown in the fourth column of Table 2, but not for the 'n', 'm', and 'c' variants of Exps. 2 and 3. See below for discussion.

As mentioned above, the mean at-issueness ratings were higher than expected in the 'n', 'm', and 'c' variants of Exps. 2 and all four of the Exps. 3. We discuss these exceptions in more detail here because they allow us to further understand the various measures for at-issueness investigated in this paper. The example in (3a) illustrates the version of the assent diagnostic applied in Exps. 2n, 2m, and 2c: the assent particle *yes* is followed by the clause whose content is diagnosed, here, the content of the main clause. The related affirmation diagnostic applied in Exp. 2q (where the mean at-issueness rating was as low as expected) is illustrated in (3b).

(3) a. Sample control stimulus in Exps. 2n, 2m, and 2c
      A: These muffins, which are really delicious, have blueberries in them.
      B: Yes, that's true, they have blueberries in them.
      Question to participants: Does A's response to B sound good?
   b. Sample control stimulus in Exp. 2q
      A: Do these muffins, which are really delicious, have blueberries in them?
      B: Yes, they have blueberries in them.

| | | Mean ratings | | |
|---|---|---|---|---|
| Exp. | Control stimuli | Certainty | Not-at-issueness | At-issueness measure |
| 1q | (1) | .14 | .05 | asking whether *c* |
| 1n | (2) | .95 | .04 | sure that *c* |
| 1m | (2) | .96 | .03 | sure that *c* |
| 1c | (2) with NRRC | .94 | .08 | sure that *c* |
| 2q | (1) with NRRC | .18 | .07 | *yes*, *c* |
| 2n | (2) with NRRC | .96 | .22 | *yes, that's true*, *c* |
| 2m | (2) with NRRC | .96 | .25 | *yes, that's true*, *c* |
| 2c | (2) with NRRC | .96 | .22 | *yes, that's true*, *c* |
| 3q | (1) with NRRC | .17 | .28 | *yes*, but $\neg c'$ |
| 3n | (2) with NRRC | .94 | .44 | *yes, that's true*, but $\neg c'$ |
| 3m | (2) with NRRC | .93 | .50 | *yes, that's true*, but $\neg c'$ |
| 3c | (2) with NRRC | .93 | .53 | *yes, that's true*, but $\neg c'$ |

Table 2: Mean certainty and at-issueness ratings for control stimuli, for self-declared American English participants

Question to participants: Does A's response to B sound good?

As shown in Table 2, the group mean on the control stimuli is numerically higher (at .22 or .25) than for Exp. 2q (.07). We hypothesize that a possible explanation for this difference is that participants take A to assert both the content of the main clause and of the NRRC in (3a) but is only asking about the main clause content in (3b). Whereas B's affirmation in (3b) specifies the one content that is affirmed, B's assent in (3a) specifies only one of the two contents that A asserted, seemingly leaving out a specification of the second content, that of the NRRC. Participants may judge B's response in (3b) to be less acceptable than in (3a) because of this missing content specification. This hypothesis is consistent with the results of Syrett and Koev's (2015) Exp. 3, which suggests that content of a sentence-medial NRRCs can be the target of a direct denial, though the main clause content is preferred as the target of such a denial. That the group means on the control stimuli in our Exps. 2n, 2m, and 2c are still relatively low may be due to the fact that the one content that was specified is the at-issue main clause content. One would expect lower acceptability ratings in a version of this diagnostic in which the one content that was specified is the NRRC.

The mean at-issueness ratings for the control stimuli were also higher than expected in Exps. 3. The examples in (4) illustrate the versions of the affirmation and assent diagnostics used in these experiments:

(4) a. Sample control stimulus in Exp. 3q
A: Do these muffins, which are really delicious, have blueberries in them?
B: Yes, but they aren't really delicious.
Question to participants: Does A's response to B sound good?

b. Sample control stimulus in Exps. 3n, 3m, and 3c
A: These muffins, which are really delicious, have blueberries in them.
B: Yes, that's true, they but they aren't really delicious.
Question to participants: Does A's response to B sound good?

The mean at-issueness rating for the control stimuli in Exp. 3q was comparatively higher (at .28) than for Exp. 1q (.05) or Exp. 2q (.07). We hypothesize that that this difference (especially to Exp. 2q) is due the content of the NRRC being directly denied, even though it was presented as backgrounded content in A's question (as shown in (2), the content of none of the other NRRCs was a matter of personal taste).

The mean at-issueness ratings for the control stimuli in Exps. 3n, 3m, and 3c were numerically even higher than for Exp. 3q and, in fact, the highest across all 12 experiments (at .44, .50, and .53, respectively). There are two factors that could be implicated in the difference between Exp. 3q and Exps. 3n, 3m, and 3c: first, the NRRC is included in a polar question in the former but a declarative assertion in the latter; second, B utters an affirmation *yes* in the former but an assent *yes, that's true* in the latter. For instance, participants might judge B's direct denial of the content of the NRRC as less acceptable when A uttered a declarative assertion and B assented with that assertion using *yes, that's true* than when A uttered a polar question and B responded in the affirmative with *yes*. Both factors must be considered in future research on at-issueness measures.

## D. Participant information and data exclusion criteria

This supplement provides information on the participants of the 12 experiments and the criteria by which participants' data were excluded. The first three columns of Table 3 show, for each of the 12 experiments how many participants were recruited, their age range and mean age, and their self-reported gender; no gender data was collected in Exp. 1q. The next three columns provide information on the number of participants whose data were excluded based on the following criteria:

- 'multiple': Due to an experimental glitch, some participants participated more than once in Exp. 1q. Since no information was available on which one was their first take, those participants' data was removed.
- 'language': Participants' data were excluded if they did not self-identify as native speakers of American English.
- 'controls': Participants' data were excluded if their mean rating on the 6 main clause control items in the projection block was more than 2 sd above the group mean (in Exps. 1q, 2q, and 3q) or more than 2 sd below the group mean (in the remaining experiments). Participants' data were also excluded if their mean rating on the 6 main clause control items in the at-issueness block was more than 2 sd above the group mean (across all experiments).
- 'variance': Participants' data were excluded if they always selected roughly the same point on the response scale for the target stimuli. To identify such participants, we first identified participants whose mean variance on the target stimuli was more than 2 sd below the group mean variance and then manually inspecting their response patterns. The data of participants who used the full scale was not excluded.

The remaining columns of Table 3 provide information on the remaining participants, that is, the participants' data that entered into the analysis. Participants took around 9-11 minutes to

complete the various experiments. Participants were paid more in Exps. 1c, 2c, and 3c than the remaining experiments because the target stimuli in those experiments were longer (as they consisted of conditionals). More women than men were recruited in many of the experiments because the experiments were run at a time when Prolific went viral on TikTok, resulting in a large number of young women registering for the service (around July 24, 2021; see `https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/`, last accessed February 4, 2022).

## E. Analysis 1 — Estimating operator effects on projection

Models: Bayesian mixed effects beta regression, with:

- **Repsonse variable:** projection ratings scaled for beta-regression to transform from closed unit interval $[0,1]$ to open unit interval $(0,1)$, using method used in Degen and Tonhauser (2022), from Smithson and Verkuilen (2006), for proportional data.

$$y' = \frac{y * (n-1) + 0.5}{n}$$

- **Fixed effect:** operator (modal, negation, question, conditional; with reference level: modal)
- **Random effects:** Predicate random intercepts
- Assuming a beta-distributed response variable, these effects were estimated for the mean ($\mu$) and precision ($\phi$) parameters of the beta-distribution.

The model was fit using `brms` using low-information priors (`brms` defaults). The estimation used 4 chains, 3,000 iterations per chain, 500 warm-up, total 10,000 iterations saved. An excerpt from the model output is shown below, providing parameter estimates, and 95% credible intervals (CIs) for the fixed effects.

| | Estimate | 95% CI |
|---|---|---|
| Intercept (modal) | -0.12 | [-0.48,0.25] |
| phi_Intercept | -0.21 | [-0.28,-0.13] |
| operatorconditional | 0.26 | [0.23,0.29] |
| operatornegation | 0.05 | [0.02,0.08] |
| operatorquestion | 0.14 | [0.11,0.17] |
| phi_operatorconditional | 0.23 | [0.21,0.26] |
| phi_operatornegation | -0.06 | [-0.08,-0.03] |
| phi_operatorquestion | 0.06 | [0.03,0.08] |

12

[12]The analysis script can be found at `https://github.com/judith-tonhauser/CommitmentBankPlus/blob/main/results/main/meta-analyses/1_projection/rscripts/2_analysis-operators.R`.

## F. Analysis 2 — Estimating operator effects for each predicate

Models: 20 Bayesian mixed effects beta regression models for the each subset of the dataset by predicate, with:

- **Repsonse variable:** projection ratings scaled for beta-regression.
- **Fixed effect:** operator (modal, negation, question, conditional; with reference level: modal)
- **Random effects:** Item random intercepts and slopes by operator
- These effects were estimated for the mean ($\mu$) and precision ($\phi$) parameters of the beta-distribution.

The models were fit using `brms` using low-information priors (`brms` defaults). The estimation used 4 chains, 4,000 iterations per chain, 700 warm-up, total 13,200 iterations saved for each model.

For each predicate, differences between levels of operator were established by considering the pairwise difference between estimated marginal means for each level of operator, using the `emmeans` package (Lenth et al. 2024) in R (R Core Team 2016). The below table gives posterior means, as well as upper and lower values for 95% highest density intervals (HDIs) for the derived variables representing the pairwise differences.

| predicate | contrast | mean | lower | upper |
|---|---|---|---|---|
| acknowledge | conditional - negation | -0.50 | -0.70 | -0.31 |
| acknowledge | conditional - question | 0.19 | -0.03 | 0.42 |
| acknowledge | modal - conditional | -0.21 | -0.36 | -0.05 |
| acknowledge | modal - negation | -0.71 | -0.90 | -0.50 |
| acknowledge | modal - question | -0.01 | -0.23 | 0.21 |
| acknowledge | negation - question | 0.69 | 0.44 | 0.97 |
| admit | conditional - negation | 0.58 | 0.40 | 0.76 |
| admit | conditional - question | 0.30 | 0.10 | 0.49 |
| admit | modal - conditional | -0.50 | -0.67 | -0.33 |
| admit | modal - negation | 0.08 | -0.08 | 0.22 |
| admit | modal - question | -0.21 | -0.37 | -0.05 |
| admit | negation - question | -0.28 | -0.46 | -0.12 |
| announce | conditional - negation | 0.42 | 0.25 | 0.60 |
| announce | conditional - question | 0.32 | 0.14 | 0.48 |
| announce | modal - conditional | -0.48 | -0.63 | -0.34 |
| announce | modal - negation | -0.06 | -0.23 | 0.11 |
| announce | modal - question | -0.17 | -0.33 | -0.01 |
| announce | negation - question | -0.11 | -0.29 | 0.07 |
| be_annoyed | conditional - negation | -0.59 | -0.77 | -0.41 |
| be_annoyed | conditional - question | -0.90 | -1.08 | -0.70 |
| be_annoyed | modal - conditional | -0.33 | -0.48 | -0.16 |

| | | | | |
|---|---|---|---|---|
| be_annoyed | modal - negation | -0.92 | -1.08 | -0.76 |
| be_annoyed | modal - question | -1.23 | -1.40 | -1.05 |
| be_annoyed | negation - question | -0.31 | -0.50 | -0.12 |
| be_right | conditional - negation | -0.22 | -0.39 | -0.06 |
| be_right | conditional - question | 0.40 | 0.21 | 0.57 |
| be_right | modal - conditional | 0.43 | 0.28 | 0.58 |
| be_right | modal - negation | 0.21 | 0.05 | 0.37 |
| be_right | modal - question | 0.84 | 0.66 | 1.00 |
| be_right | negation - question | 0.62 | 0.43 | 0.81 |
| confess | conditional - negation | 0.65 | 0.48 | 0.82 |
| confess | conditional - question | 0.45 | 0.29 | 0.61 |
| confess | modal - conditional | -0.62 | -0.77 | -0.48 |
| confess | modal - negation | 0.03 | -0.14 | 0.18 |
| confess | modal - question | -0.17 | -0.32 | -0.02 |
| confess | negation - question | -0.20 | -0.38 | -0.03 |
| confirm | conditional - negation | 0.77 | 0.59 | 0.95 |
| confirm | conditional - question | 0.55 | 0.34 | 0.76 |
| confirm | modal - conditional | 0.12 | -0.04 | 0.28 |
| confirm | modal - negation | 0.89 | 0.73 | 1.05 |
| confirm | modal - question | 0.67 | 0.48 | 0.85 |
| confirm | negation - question | -0.22 | -0.42 | -0.02 |
| demonstrate | conditional - negation | 0.58 | 0.42 | 0.74 |
| demonstrate | conditional - question | 0.21 | 0.06 | 0.36 |
| demonstrate | modal - conditional | -0.17 | -0.31 | -0.03 |
| demonstrate | modal - negation | 0.41 | 0.25 | 0.56 |
| demonstrate | modal - question | 0.03 | -0.11 | 0.18 |
| demonstrate | negation - question | -0.37 | -0.53 | -0.21 |
| discover | conditional - negation | 0.46 | 0.27 | 0.65 |
| discover | conditional - question | -0.15 | -0.33 | 0.03 |
| discover | modal - conditional | -0.64 | -0.79 | -0.48 |
| discover | modal - negation | -0.18 | -0.36 | -0.01 |
| discover | modal - question | -0.79 | -0.95 | -0.61 |
| discover | negation - question | -0.61 | -0.81 | -0.42 |
| establish | conditional - negation | 0.44 | 0.26 | 0.62 |
| establish | conditional - question | 0.28 | 0.10 | 0.47 |
| establish | modal - conditional | -0.06 | -0.24 | 0.11 |
| establish | modal - negation | 0.37 | 0.22 | 0.52 |
| establish | modal - question | 0.22 | 0.07 | 0.37 |
| establish | negation - question | -0.15 | -0.32 | 0.01 |
| hear | conditional - negation | 0.44 | 0.27 | 0.61 |
| hear | conditional - question | -0.03 | -0.19 | 0.13 |
| hear | modal - conditional | -0.92 | -1.06 | -0.76 |
| hear | modal - negation | -0.47 | -0.63 | -0.31 |

| | | | | |
|---|---|---|---|---|
| hear | modal - question | -0.95 | -1.10 | -0.79 |
| hear | negation - question | -0.47 | -0.65 | -0.31 |
| inform | conditional - negation | 0.12 | -0.07 | 0.30 |
| inform | conditional - question | -0.33 | -0.53 | -0.13 |
| inform | modal - conditional | -0.63 | -0.80 | -0.45 |
| inform | modal - negation | -0.51 | -0.67 | -0.35 |
| inform | modal - question | -0.96 | -1.12 | -0.79 |
| inform | negation - question | -0.45 | -0.62 | -0.26 |
| know | conditional - negation | 0.13 | -0.06 | 0.31 |
| know | conditional - question | -0.56 | -0.79 | -0.35 |
| know | modal - conditional | -0.70 | -0.86 | -0.55 |
| know | modal - negation | -0.57 | -0.74 | -0.40 |
| know | modal - question | -1.26 | -1.49 | -1.04 |
| know | negation - question | -0.70 | -0.93 | -0.46 |
| pretend | conditional - negation | -1.00 | -1.22 | -0.76 |
| pretend | conditional - question | -0.09 | -0.33 | 0.12 |
| pretend | modal - conditional | 0.20 | -0.03 | 0.42 |
| pretend | modal - negation | -0.80 | -0.97 | -0.64 |
| pretend | modal - question | 0.11 | -0.06 | 0.27 |
| pretend | negation - question | 0.91 | 0.73 | 1.08 |
| prove | conditional - negation | 0.52 | 0.35 | 0.69 |
| prove | conditional - question | 0.38 | 0.20 | 0.54 |
| prove | modal - conditional | 0.24 | 0.09 | 0.40 |
| prove | modal - negation | 0.77 | 0.61 | 0.92 |
| prove | modal - question | 0.62 | 0.46 | 0.78 |
| prove | negation - question | -0.15 | -0.33 | 0.02 |
| reveal | conditional - negation | 0.63 | 0.46 | 0.79 |
| reveal | conditional - question | 0.29 | 0.10 | 0.49 |
| reveal | modal - conditional | -0.60 | -0.77 | -0.45 |
| reveal | modal - negation | 0.02 | -0.13 | 0.16 |
| reveal | modal - question | -0.32 | -0.49 | -0.13 |
| reveal | negation - question | -0.34 | -0.52 | -0.16 |
| say | conditional - negation | 0.30 | 0.04 | 0.54 |
| say | conditional - question | 1.23 | 1.01 | 1.44 |
| say | modal - conditional | -0.45 | -0.64 | -0.28 |
| say | modal - negation | -0.16 | -0.36 | 0.05 |
| say | modal - question | 0.77 | 0.59 | 0.95 |
| say | negation - question | 0.93 | 0.73 | 1.14 |
| see | conditional - negation | 0.55 | 0.33 | 0.78 |
| see | conditional - question | -0.15 | -0.35 | 0.08 |
| see | modal - conditional | -0.89 | -1.09 | -0.67 |
| see | modal - negation | -0.33 | -0.50 | -0.17 |
| see | modal - question | -1.03 | -1.20 | -0.87 |

| | | | | |
|---|---|---|---|---|
| see | negation - question | -0.70 | -0.89 | -0.52 |
| suggest | conditional - negation | 0.25 | 0.08 | 0.42 |
| suggest | conditional - question | 0.61 | 0.44 | 0.77 |
| suggest | modal - conditional | -0.27 | -0.42 | -0.11 |
| suggest | modal - negation | -0.02 | -0.19 | 0.14 |
| suggest | modal - question | 0.34 | 0.18 | 0.49 |
| suggest | negation - question | 0.36 | 0.19 | 0.52 |
| think | conditional - negation | -0.50 | -0.70 | -0.30 |
| think | conditional - question | 0.19 | -0.04 | 0.41 |
| think | modal - conditional | -0.21 | -0.36 | -0.05 |
| think | modal - negation | -0.71 | -0.91 | -0.52 |
| think | modal - question | -0.01 | -0.24 | 0.22 |
| think | negation - question | 0.70 | 0.44 | 0.97 |

| | Recruited participants | | | Exclusion criteria | | | | Remaining participants | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Exp. | total | ages (mean) | gender (f/m/o/u) | multiple | language | controls | variance | total | ages (mean) | gender (f/m/o/u) | payment |
| 1q | 300 | 19-74 (38.2) | – | 5 | 7 | 35 | 0 | 242 | 21-74 (39.2) | – | $1.70 |
| 1n | 300 | 18-74 (33.2) | 150/145/5/0 | 0 | 8 | 17 | 1 | 274 | 18-74 (33.3) | 141/128/5/0 | $1.70 |
| 1m | 300 | 18-74 (32.7) | 150/141/7/2 | 0 | 0 | 19 | 0 | 281 | 18-74 (32.7) | 144/129/7/1 | $1.70 |
| 1c | 300 | 18-58 (25.9) | 249/45/6/0 | 0 | 6 | 26 | 2 | 266 | 18-58 (24.8) | 235/25/6/0 | $2.15 |
| 2q | 250 | 18-58 (25.5) | 201/43/6/0 | 0 | 4 | 24 | 1 | 220 | 18-58 (24.8) | 187/28/5/0 | $1.70 |
| 2n | 250 | 18-69 (33.2) | 127/114/6/1 | 1 | 4 | 29 | 0 | 215 | 18-69 (33.1) | 113/95/6/1 | $1.70 |
| 2m | 251 | 18-74 (31.7) | 132/113/6/0 | 0 | 4 | 27 | 0 | 220 | 18-70 (31.9) | 116/98/6/0 | $1.70 |
| 2c | 250 | 18-56 (24.5) | 212/30/8/0 | 0 | 0 | 26 | 0 | 224 | 18-56 (24.4) | 195/24/5/0 | $2.15 |
| 3q | 250 | 18-66 (32.4) | 140/102/7/1 | 0 | 4 | 20 | 0 | 225 | 18-66 (32.6) | 125/93//7/0 | $1.70 |
| 3n | 250 | 18-70 (24.6) | 114/31/5/0 | 0 | 5 | 13 | 4 | 228 | 18-70 (24.3) | 198/25/5/0 | $1.70 |
| 3m | 250 | 18-63 (25.5) | 205/40/5/0 | 0 | 3 | 14 | 0 | 233 | 18-63 (24.8) | 197/31/5/0 | $1.70 |
| 3c | 250 | 18-59 (27.5) | 182/64/4/0 | 0 | 3 | 17 | 0 | 230 | 18-59 (26.7) | 177/49/4/0 | $2.15 |

Table 3: Recruited participants, excluded data, and remaining participants in Exps. 1, 2 and 3. Gender distinctions were ‘f’ = female, ‘m’ = male, ‘o’ = other, and ‘u’ = undeclared.