Prior probability predicts projection

Author(s)

October 15, 2020

Abstract

Interpreters' beliefs about the world have been shown to influence utterance interpretation in a variety of empirical domains, including ambiguity resolution, reference resolution, genericity and scalar implicatures. This paper provides experimental evidence that such beliefs also influence projection, that is, the phenomenon whereby speakers are committed to non-entailed utterance content. What's more, the experimental evidence shows that there is a by-participant effect: interpreters content priors differ, but the stronger an interpreter's prior belief in content, the more the content projects.

1 Introduction

Interpreters' beliefs about the world have been shown to influence sentence processing (e.g., Chambers et al. 2002, Hagoort et al. 2004, Hald et al. 2007) and utterance interpretation in a variety of empirical domains, including ambiguity resolution (e.g., Chambers et al. 2004, Bicknell and Rohde 2014), reference resolution (e.g., Hanna and Tanenhaus 2004), genericity (e.g., Tessler and Goodman 2019) and scalar implicature (e.g., Degen et al. 2015).[1] This paper provides empirical evidence that interpreters' beliefs also influence projection. Projection is characterized as the speaker being committed to content, based on their utterance, which then also influences whether the listener is committed: the content becomes part of the common ground between the speaker and the listener. A speaker who utters the polar question in (1b) may be committed to the truth of the content of the complement, that it is raining, even though it is realized in a polar question, that is, even though it is not entailed content.

(1) a. Sam knows that it's raining.
b. Does Sam know that it's raining?
c. Sam doesn't know that it's raining.

The hypothesis that the prior probability of content influences its projection has been put forth in Stevens et al. 2017 and Tonhauser et al. 2018. Stevens et al. 2017 investigated the projection of the prejacent of utterances of manner adverb sentences, illustrated in (2). An utterance of (2a) implies both the prejacent, that Masha ran, and that her running was done in a quick manner. The prejacent is projective because it may also be implied by an utterance of the negated variant of (2a) given in (2b), especially when the adverb is produced with prosodic emphasis.

(2) Stevens et al. 2017:1144
a. Masha ran quickly.
b. Masha didn't run QUICKLY.

[1] Different kinds of beliefs have been investigated. Some of these are called world knowledge, but it's not clear that it's knowledge. We use the more cautious term belief.

Stevens et al. 2017 observed by-item variability in the projection of the prejacent and suggested that this variability may be due to interpreters having different prior probabilities about the prejacent.

Tonhauser et al. 2018 investigated the projection of content associated with 19 American English expressions, including the content of the complement of clause-embedding predicates, like *know*, illustrated in (1), and the prejacent of utterances of sentences with *stop*, illustrated in (3). The pre-state content of (3a), that Mary's daughter has been biting her nails, survives embedding in the question version in (3b).

(3) a. Mary's daughter stopped biting her nails.
    b. Has Mary's daughter stopped biting her nails?

In addition to observing by-expression projection variability, Tonhauser et al. 2018 also found by-item variability for the same expression: for instance, the pre-state content of *stop* was numerically less projective when it was the content that Mary's daughter has been biting her nails than when it was the content that Jack was playing outside with the kids. Tonhauser et al. 2018:500 hypothesized that "the projectivity of content may depend on the prior probability of the event described by the expression that conveys the content, such that content conveyed by expressions that describe more a priori likely events may be more likely to project".

JT: 1. Prior research shows that prior probability influences the empirical phenomenon, at a population level. E.g., in general, if a container is big enough, then people will be have in a particular way. we also show this: in general, content with higher prior probability projects more. but don't we go beyond this in a significant way, by showing that a) there are variable priors and b) that there is a positive relationship between these variable priors and projection?

JT: 2. In this paper, we investigate the hypothesis that the prior probability of content predicts its projection, that is, that the beliefs of participants about the relevant content prior to observing the utterance influence the extent to which they take the speaker to be committed to that content.

JT: Lorson (2018) investigated this hypothesis. The relevant contents are 'James/Linda worked as a plumber'. Prior work established that people believe that it is more likely for a man to be a plumber than for a woman to be a plumber. Lorson exploited these beliefs to see whether prior probability predicts projection.

JT: Mahler (2020) investigated the projection of politically charged content like 'Obama improved/damaged the American economy'. Participants will differ in their beliefs about these contents. But that is not what Mahler manipulated (or measured). Rather, she manipulated whether the speaker is Democrat or Republication, that is, what she manipulated is the speaker's presumed belief in that content (before making the utterance). She then found that participants rate speakers with a higher prior belief as more committed than speakers with a lower prior belief.

Our investigation builds on previous work that has investigated this hypothesis, with different results, namely Lorson 2018 and Mahler 2020. Lorson (2018) investigated the influence of prior probability on the projection of the pre-state content of the English change of state verb *stop*: in the polar questions in (4), the pre-state contents are that James has worked as a plumber, in (4a), and that Linda has worked as a plumber, in (4b).

(4) a. Did James stop working as a plumber?
    b. Did Linda stop working as a plumber? (Lorson 2018:38)

Gender stereotypes reported in Boyce et al. 2018 were used to manipulate the prior probability of the pre-state content, that is, the prior beliefs of the interlocutors about the pre-state content: it is, for instance, more likely for a man (like James) to be a plumber than for a woman (like Linda) to be a plumber. The hypothesis was that higher probability pre-state contents, as in (4a), were more projective than lower probability pre-state contents, as in (4b). Lorson (2018) did not find empirical support for this hypothesis.

Mahler (2020) investigated the projection of politically-charged content, which was realized as the complement of a clause-embedding predicate: in (5), for example, the politically-charged content is that Obama improved/damaged the American economy, which is realized as the complement of the clause-embedding predicate *know*, which is embedded under negation.

(5) Cindy doesn't know that...
  a. ...Obama improved the American economy.
  b. ...Obama damaged the American economy. (Mahler 2020:779)

In Mahler 2020,

The speaker's political affiliation was used to manipulate the prior probability of the content of the complement: the speaker was attending a a meeting of the Democrat or the Republican party and was thereby implied to be a Democrat or a Republican. Thus, the content of the complement had a higher prior

The combination of the political party affiliation of the interlocutors and the content of the complement was used to manipulate the prior probability of the content: when the content of the complement was consistent with a liberal political perspective, as in (5a), it had a higher prior probability with Democrat interlocutors than with Republican interlocutors, and when the content of the complement was consistent with a conservative political perspective, as in (5b), it had a higher prior probability with Republican interlocutors than with Democrat interlocutors. Mahler (2020) found that prior probability influenced projection: the content of the complement was more projective when it had a higher prior probability than when it had a lower prior probability.

So, there is currently a conflicting state in the literature about the hypothesis that prior probability predicts projection. The difference in findings may be due to

- the projective content investigated (pre-state content of *stop* versus the content of the complement of clause-embedding predicates),

- the ways in which the prior was manipulated (gender stereotypes versus political party affilliation).

Both Lorson 2018 and Mahler 2020 used the certainty diagnostic for projection (Tonhauser et al. 2018): sliding scale in both. Lorson 2018 relied on gender stereotypes established in previous work (no separate norming study) while Mahler 2020 conducted a norming study: Lorson recruited participants on Prolific and social media (no info on their demographics) while gender stereotypes were established on AMT; Mahler ran both the norming study and the experiment on AMT. Prior probability entered Lorson's model in a non-categorical fashion in Lorson 2018 (consistency measure) and in a categorical fashion in Mahler 2020 (political party (2 levels) and political orientation of CC (liberal vs. conservative). Lorson, p.36: "The consistency measure is a value that ranges from 0 to 1, where a value close to 0 means that the gender of an individual is inconsistent with a certain occupation, according to the collected gender-bias that was found to be associated with that occupation. A value close to 1 means that the gender of an individual is consistent with an occupation."

Our experiments manipulate the prior in much more variable ways than Mahler and Lorson. Like Mahler, we investigated the hypothesis that prior probability influences projection based on the content of the complement of clause-embedding predicates. But we do so on a broader set of 20 predicates than she did (she had 7 predicates, which she grouped into two groups and she found the effect in both groups). Our findings consistent with Mahler's but we have a broader set of predicates and don't artificially group the predicates. We have shown elsewhere (Tonhauser and Degen ms) that this binary distinction is not supported by the data. Also we have a broader manipulation of the prior.

The hypothesis we investigate is that people's prior beliefs about content affect projection, i.e., the extent to which the speaker is **JT: taken to be** committed to the content. People's individual beliefs are indicative of their prior probability.

[2]

Experiment 1: We measured the belief of each participant and also measured how much they take the speaker to be committed.

Experiment 2: We measured the belief of one group of people, test how the average belief of that group predicts something about projection of the second group.

[2] The experiments, data and R code for generating the figures and analyses of the experiments reported on in this paper are available at [redacted for review].

2 Experiment 1

This experiment investigated the hypothesis that the prior probability of content influences its projection, specifically, that the higher the prior probability of content is, the more projective it is. Prior probability and projection ratings were collected for the contents of 20 clauses that realized the complements of clause-embedding predicates.

Participants 300 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform (ages: 18-82, median: 35.5; 119 female, 179 male, 1 other, 1 undeclared). They were paid $1.80 for participating.

Materials We measured the prior probability of the contents of 20 clauses, which are given in (6) together with two facts that influence the prior probability of the content of the clause.

(6) 20 clauses with lower and higher probability facts, respectively

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

The experiment had two blocks: a prior block and a projection block. In the prior block, the 20 clauses were realized as polar questions and paired with the higher or the lower probability fact, as shown in the sample target stimuli in (7). There were a total of 40 target stimuli in the prior block.

(7) Sample target stimuli in the prior block

a. **Fact:** Julian is Cuban.
How likely is it that Julian dances salsa?

b. **Fact:** Julian is German.
How likely is it that Julian dances salsa?

In the projection block, target stimuli consisted of a fact and a polar question, as shown in (8). The polar questions were formed by realizing the 20 clauses in (6) as the complements of the 20 clause-embedding predicates in (9). The fact of each target stimulus was one of the two facts that the clause realized in the polar question was paired with. There were a total of 800 target stimuli in the projection block.

(8) Sample target stimuli in the projection block

a. **Fact (which Carol knows):** Julian is Cuban.
Carol: Does Sandra know that Julian dances salsa?

b. **Fact (which Carol knows):** Julian is German.
Carol: Does Sandra know that Julian dances salsa?

(9) 20 clause-embedding predicates
be annoyed, discover, know, reveal, see, acknowledge, admit, announce, confess, confirm, establish, hear, inform, prove, pretend, suggest, say, think, be right, demonstrate

The 20 predicates include both so-called factive predicates, like *know, be annoyed* and *see*, as well as so-called non-factive predicates, like *confirm, say* and *think*. This latter set of predicates is not typically considered in research on projection because the CC of such predicates is not assumed to be presupposed to be true by the speaker. However, Tonhauser and Degen (ms) show that there is no empirical support for a categorical distinction between factive and non-factive predicates and that the CC of non-factive predicates is projective, contrary to assumption. We therefore follow Tonhauser and Degen's ms proposal here that non-factive predicates be included in projection investigations.

The projection block also included 6 control stimuli, which we used to assess whether participants were attending to the task. The content of these control stimuli was hypothesized to not project. For example, in the control stimulus in (10), the speaker is not committed to the main clause content, that Zack is coming to the meeting tomorrow. The control stimuli were presented with a fact that we hypothesized did not influence the prior probability of the content. The same 6 control stimuli were presented to all participants; for the full set see Supplement A.

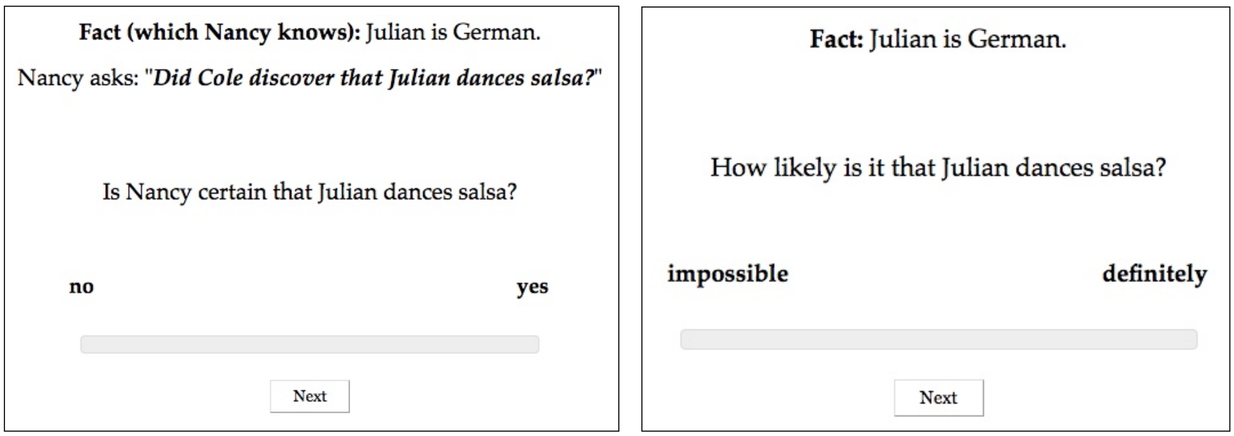
(10) Sample control stimulus in the projection block
Fact (which Margaret knows): Zack is a member of the golf club.
Margaret: Is Zack coming to the meeting tomorrow?

The prior block included 6 filler stimuli, which were formed from the same 6 main clauses as the control stimuli in the projection block. For example, the filler stimulus in (11) is formed from the main clause content of the control stimulus in (10). We expected participants to give mid-range likelihood ratings to these contents, but did not use these stimuli to assess participants' attention to the task. The same 6 filler stimuli were presented to all participants; for the full set see Supplement A.

(11) Sample filler stimulus in the prior block
Fact: Zack is a member of the golf club.
How likely is it that Zack is coming to the meeting tomorrow?

For each participant, a random combination of the 20 clauses with a fact was generated, with 10 higher probability facts and 10 lower probability facts, for a total of 20 clause/fact combinations. Participants saw a total of 52 stimuli: 20 target stimuli in each block, 6 control stimuli in the projection block and 6 filler stimuli in the prior block. In the projection block, participants gave certainty ratings for the contents of these 20 clauses given the facts; in the prior block, participants gave prior probability ratings for the contents of these 20 clauses given the facts. In the projection block, the 20 clauses were randomly realized as the complements of the 20 clause-embedding predicates, so that each participant gave a certainty rating for the content of the complement of each predicate. Block order and within-block trial order were randomized.

Procedure In the projection block, participants were told to imagine that they are at a party and that, on walking into the kitchen, they overhear somebody ask somebody else a question. Participants were asked to rate whether the speaker was certain of the content of the complement, taking into consideration the fact that was presented. They gave their responses on a slider marked 'no' at one end (coded as 0) and 'yes' at the other (coded as 1), as shown in Figure 1a. In the prior block, participants were told to read facts and to assess the likelihood of events, given those facts. They gave their responses on a slider marked 'impossible' at one end (coded as 0) and 'definitely' at the other (coded as 1), as shown in the sample trial in Figure 1b



(a) Projection block. (b) Prior block.

Figure 1: Sample trials in Exp. 1.

After completing the experiment, participants filled out a short, optional survey about their age, their gender, their native language(s) and, if English is their native language, whether they are a speaker of American English (as opposed to, e.g., Australian or Indian English). To encourage them to respond truthfully, participants were told that they would be paid no matter what answers they gave in the survey.

Data exclusion Prior to analysis, the data from 3 participants who did not self-identify as native speakers of American English were excluded. To assess whether the remaining 297 participants attended to the task, we inspected their responses to the control stimuli. We excluded the data from 11 participants whose response means were more than 2 standard deviations above the group mean. The data from 286 participants (ages 18-82; median: 35.5; 116 female, 186 male, 1 other, 1 undeclared) were analyzed.

2.1 Results

The manipulation of the prior probability of the 20 contents was successful. As shown in Figure 2, which plots the mean prior probabilities of the 20 contents by fact (participants' ratings are given as light dots), the mean prior probability rating for each content was higher when the content was presented with the higher probability fact than when it was presented with the lower probability fact.

We now address the research question of whether prior probability influences projection. Figure 3 plots the mean certainty ratings for the contents of the clausal complements by clause-embedding predicate and by fact, as well as the mean certainty rating for the main clause content (abbreviated 'MC'). Light dots indicate participants' ratings. As shown, the mean certainty ratings were higher for contents presented with the higher probability facts than for contents presented with the lower probability fact. This finding, which holds for all 20 clause-embedding predicates, suggests that prior probability influences projection.

We also replicate Tonhauser and Degen's ms finding that there is by-predicate variation in the projection

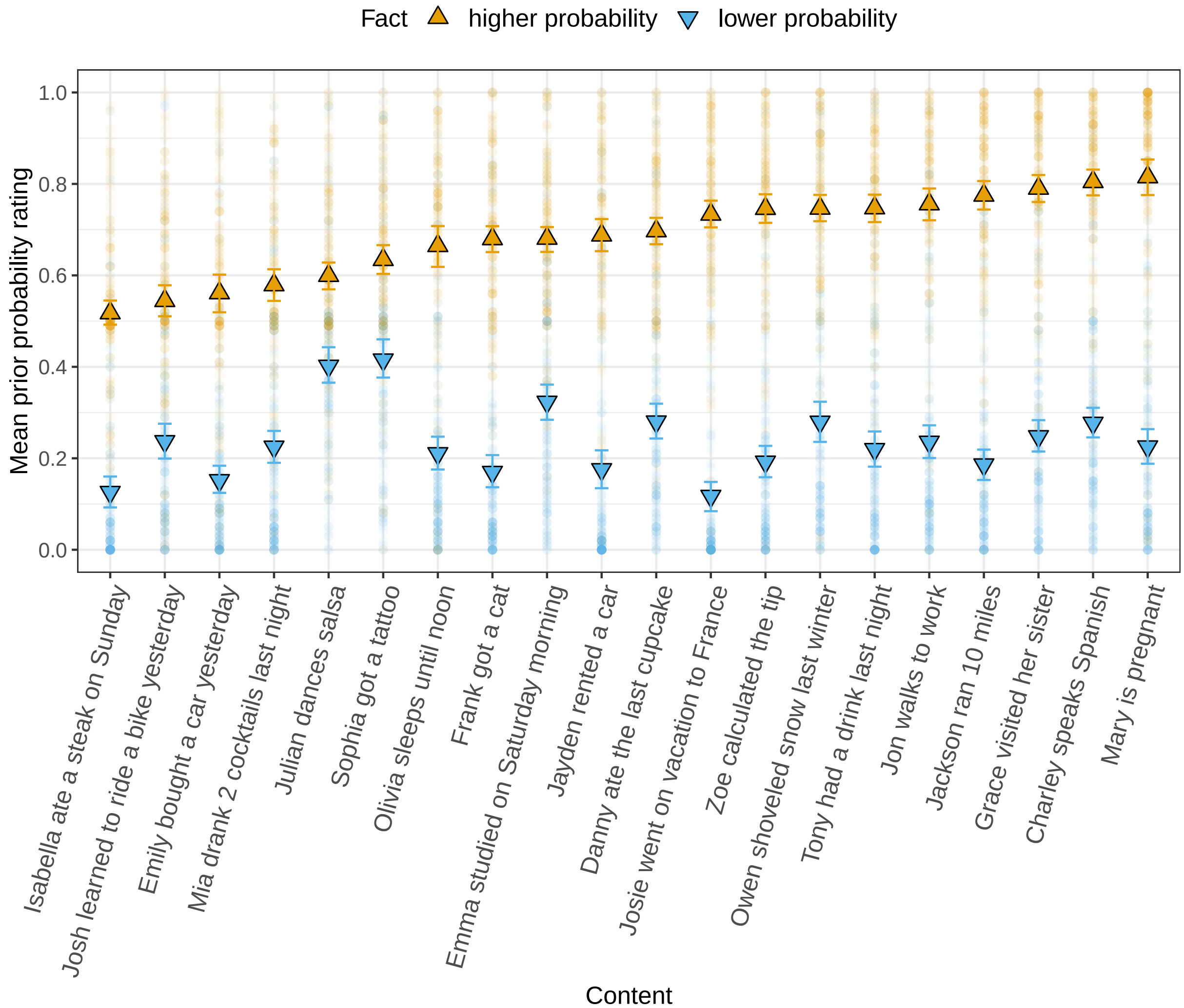


Figure 2: Mean prior probability by content and fact in Exp. 1. Error bars indicate 95% bootstrapped confidence intervals. Light dots indicate participants' ratings.

of the content of the complement: for instance, the content of the complement of *be annoyed* is more projective than that of *discover*, which in turn is more projective than that of *announce*. The Spearman rank correlation between the mean certainty ratings in Exp. 1 and Tonhauser and Degen's ms Exp. 1a is .991 (collapsing over facts; see Supplement C for a visualization). Exp. 1 thereby provides further evidence for the systematic influence of the predicate on projection.

Figure 3 plots the certainty ratings according to whether the fact that was presented to the participant was a higher or a lower probability fact. And although Figure 2 shows that higher and lower probability facts resulted in the expected differences in the mean prior probability ratings for the contents of the 20 clauses, this figure also reveals by-participant variation in prior probability ratings. As a consequence, individual participants' prior probability ratings need not align with assumed prior probability classification. For example, given a particular content (e.g., that Julian dances salsa), it is possible that one participant's prior probability rating was lower than that of another participant, even though the first participant was presented with the higher probability fact (Julian is Cuban) and the second one with the lower probability fact (Julian is German). To investigate whether prior probability influences projection at the by-participant level,

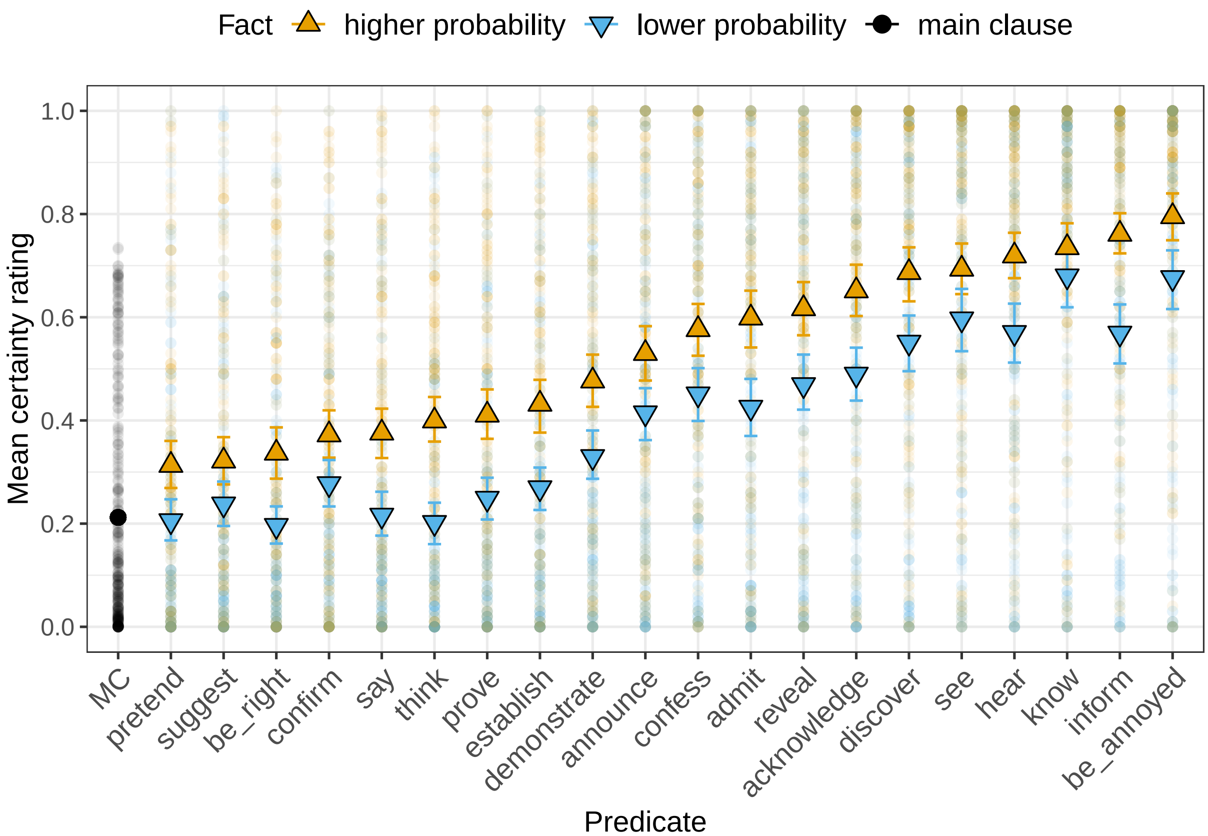


Figure 3: Mean certainty ratings by predicate and prior probability of the content of the complement. Error bars indicate 95% bootstrapped confidence intervals. Light dots indicate participants' ratings.

Figure 4 plots the participants' certainty ratings (indicating projection) by their prior probability ratings for the contents of the complements of the 20 clause-embedding predicates. (The color coding here merely represents the type of fact the participant was presented with, but no classification is imposed.) The linear smoothers suggest a positive correlation for each predicate between prior probability and certainty ratings such that contents with higher prior probability ratings receive higher certainty ratings. This finding suggests that prior probability predicts projection even at the by-participant level.

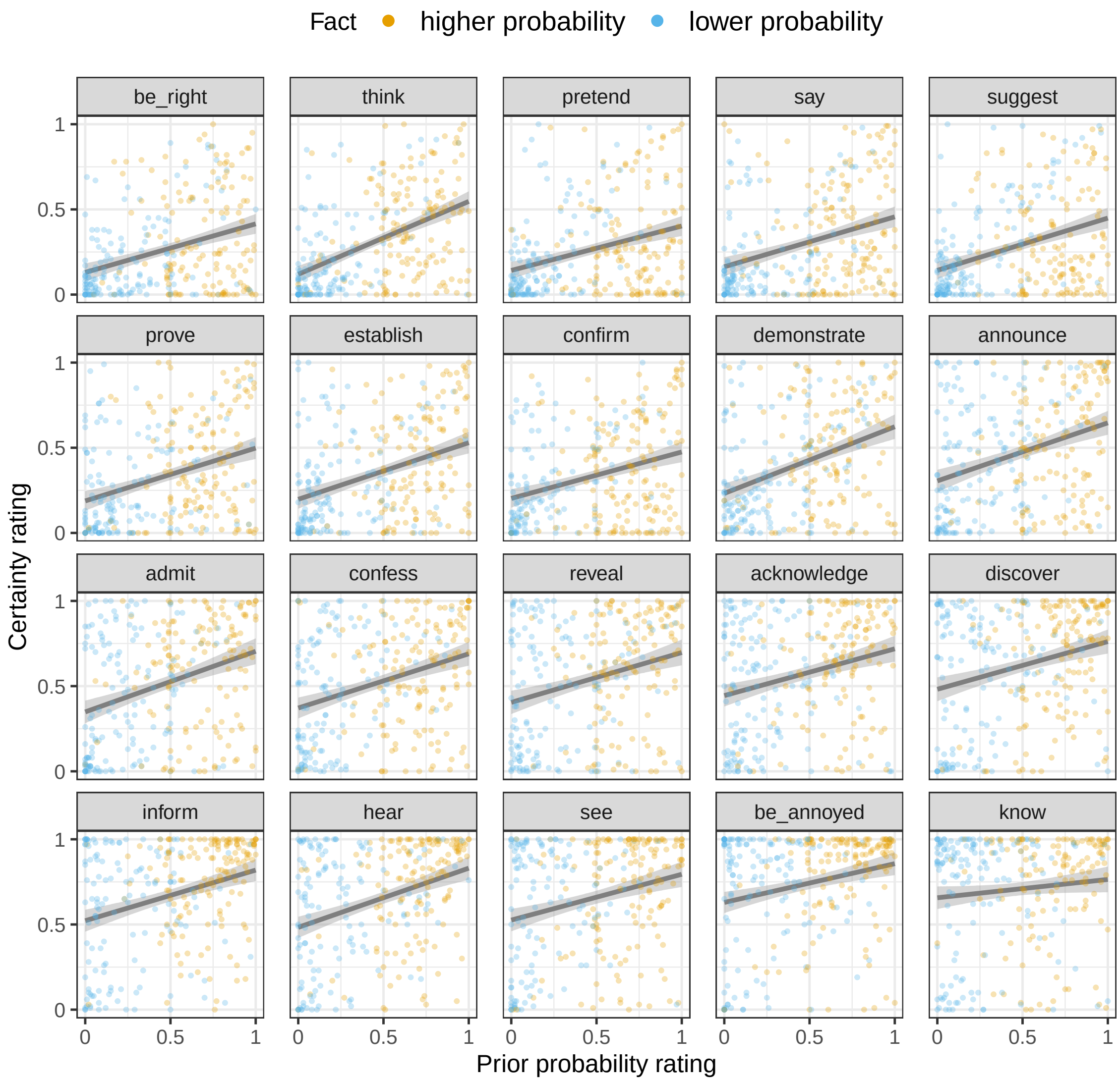


Figure 4: Certainty ratings by prior probability ratings by predicate in Exp. 1. Linear smoothers with 95% confidence intervals are overlaid.

The qualitative observations about the relations between prior probability, clause-embedding predicate and projection were borne out statistically. We fitted a Bayesian mixed effects Beta regression model with weakly informative priors using the `brms` (Bürkner 2017) package in R (R Core Team 2016) on the target data (5,720 data points). The model predicted the certainty ratings from a fixed effect of prior probability and included the maximal random effects structure justified by the design, namely random by-participant and by-item intercepts (where an item is a combination of a predicate and a complement clause). A Beta regression model estimates the mean of the outcome distribution (like a linear regression model).[3] We thus obtain a 95% credible interval for **JT: the mean effect of prior probability on certainty?**. Supplement

[3]Beta regression models also estimate a second parameter, namely the precision, which is a measure of dispersion: the greater the precision, the more concentrated the ratings are around the mean. In this paper, we rely on the estimated mean to identify whether prior probability predicts projection. Both the estimated mean and precision are reported in the full model output table in Supplement B.

B motivates the use of Beta regression over linear regression, provides a brief primer on how to interpret Bayesian mixed effects Beta regression models, and reports the full model output.

JT: According to the Beta regression model, the estimated mean for each predicate was higher than that of the main clause controls, i.e., the 95% credible intervals for the estimated adjustment to the main clause control mean did not contain 0 for any predicate. This suggests that the content of the complement of each of the 20 predicates is projective compared to non-projective main clause content.[4] Thus, to distinguish factive predicates from optionally factive and non-factive ones in this set of 20 predicates, one would need to arbitrarily distinguish one group of projective CCs from another group of projective CCs.

2.2 Discussion

The findings of Exp. 1 provide empirical support for the hypothesis, put forth in Stevens et al. 2017 and Tonhauser et al. 2018, that the prior probability of content influences its projection. Specifically, we observed that a participant's certainty ratings for the content of the clausal complement of a predicate like *think, announce* and *know* are influenced by the prior probability rating for such content.

Mahler also manipulated clause-embedding predicate: two categories (factive: *know, realize, see, discover* and non-factive: *believe, think, feel*). CCs of factive predicates more projective than CCs of non-factive predicates. In addition to wanting to identify whether prior probability influences projection, she also wants to find out whether this only happens for factive predicates or for factive and non-factive predicates alike. Later on in the paper she adds the RQ of whether projection is influenced by factivity. She finds that prior probability influences projection for both factive and non-factives. She also finds that projection is higher with factive than non-factive predicates.

Lorson: There are several differences between Exp. 1 and Lorson's 2018 investigation that may have contributed to why Lorson 2018 did not observe an influence of prior probability on projection.

1. Difference: the prior probability of content was manipulated through gender stereotypes in Lorson 2018 but assumptions about what individuals with a variety of dispositions, professions, ages or nationalities are more or less likely to do. As shown in Figure A3, the effect of prior probability on projection is observed for all contents, i.e., regardless of how prior probability was manipulated.

2. Difference: we investigated the projection of the content of the clausal complement of clause-embedding predicates whereas Lorson 2018 investigated the pre-state content of *stop*. **JT: How much to investigate this possibility?**

Mahler: found an effect based on a manipulation of prior probability based on content that was more or less likely depending on the political party affilliation.

1. We replicate her finding, showing that it generalizes to prior probability manipulated based on a broader variety of properties of individuals. **JT: May also be a good reason to include Fig 5**

2. Mahler used the content of the complement of clause-embedding predicates, like us. She had 7 predicates; we had 20. She found that the CCs of factive predicates more projective than CCs of non-factive predicates, and finds the effect for both groups (no by-predicate discussion). We go beyond her by showing that this holds for 20 predicates.

[4] **JT: A Bayesian mixed effects linear regression with the same fixed and random effects structure yielded qualitatively identical results, except that the contrast between *pretend* and the main clause controls was only marginally significant. See the Github repository mentioned in footnote 2 for the model code.**

3. “CCs were more projective when the predicate was factive compared to when it was non-factive, regardless of the speaker?s political affiliation. This finding is compatible with the assumption that factive predicates lexically-encode their complements as presuppositions, whereas non-factive predicates do not.” (p.788) **JT: We want to engage with this here: factives paper challenges factive/non-factive distinction so the fact that we find effect of prior on all predicates cannot be used to support such a differential analysis**

   in general our findings are compatible with Mahler?s (we both find effect). if one wanted to uphold the binary divide, we also find that factives more than non-factives, but the fact that the prior effect holds across the different groups we think further supports that no binary categorical distinction necessary

In Exp. 1, prior probability and projection were measured in a within-participant design in order to investigate whether prior probability predicts projection at the by-participant level. It is possible, however, that participants' responses on either block primed their responses on the other block. To mitigate against this possibility, we conducted Exp. 2, where prior probability and projection ratings were collected from different population

3 Experiment 2

This experiment collected prior probability ratings for the 20 contents. **JT: write this experiment up as if we collected both prior and projection ratings?**

3.1 Methods

Participants 95 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform (ages: 21-75, median: 33; 45 female, 50 male). They were paid 55 cents for participating.

Materials The 40 target stimuli were identical to the target stimuli of the prior block of Exp. 1. For each participant, a random combination of 20 clause/fact combinations was generated, for 20 target stimuli. Each participant also saw the two control stimuli in (10), which were included to assess whether participants were attending to the task: we expected participants to rate the prior probability of the content in (10a) as high and that of the content in (10b) as low. Trial order of the 22 stimuli was random.

(12) a. **Fact:** Barry lives in Germany.
How likely is it that Barry lives in Europe?

b. **Fact:** Tammy is a rabbit.
How likely is it that Tammy speaks Italian and Greek?

Procedure The procedure was identical to the procedure of the prior block of Exp. 1. After completing the experiment, participants filled out the same optional survey as in Exp. 1.

Data exclusion Prior to analysis, the data from 8 participants who did not self-identify as native speakers of American English were excluded. To assess whether the remaining 87 participants attended to the task, we inspected their responses to the 2 control stimuli. The response means of 12 participants were more than 2 standard deviations below the group mean of the control in (10a) or more than 2 standard deviations above the group mean of the control in (10b). We excluded the data from these participants, leaving data from 75 participants (ages 21-75; median: 35; 34 female, 41 male).

3.2 Results and discussion

JT: Is the effect of prior on projection an artifact of the ratings collected in Exp1? predict projection mean from prior mean in both Exp1 and with Exp2 prior ratings, for comparing the two experiments; use cohen?s d?

Figure 5 plots the mean prior probability ratings in Exp. 2 against the mean prior probability ratings in Exp. 2. **JT: what do we see?**

Spearman rank correlation: .98

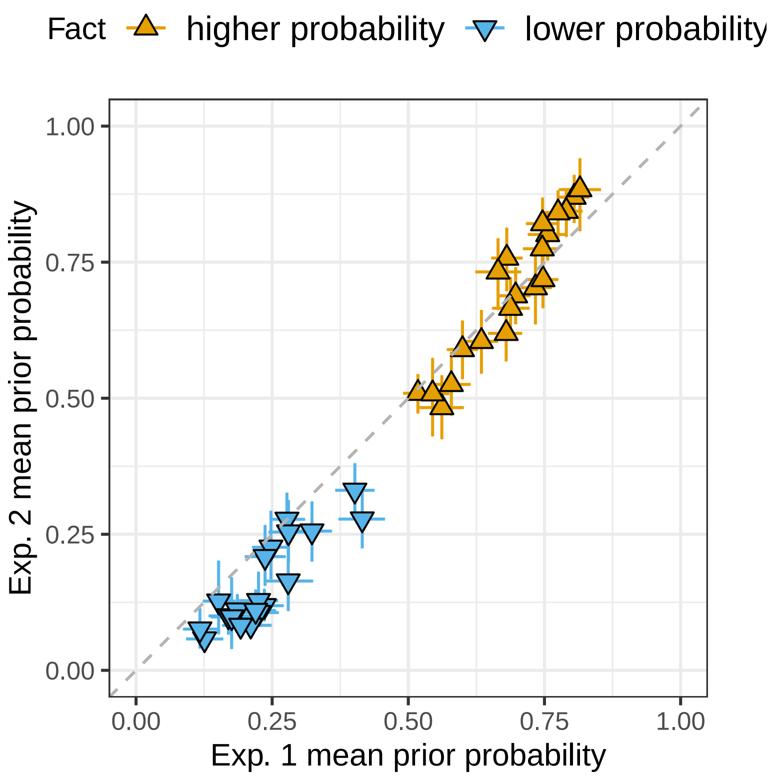


Figure 5: Mean prior probability ratings in Exp. 2 against those of Exp. 1. Error bars indicate 95% bootstrapped confidence intervals.

JT: Report model that predicts projection of Exp. 1 projection data from the new prior probability means.

4 Discussion

as well as for the main clause stimuli (abbreviated 'MC'), in increasing order from left to right. The mean certainty ratings are largely consistent with impressionistic judgments reported in the literature. First, the ratings for main clause content are lowest overall, as expected for non-projective content. Second, the ratings for factive predicates are among the highest overall, suggesting comparatively high projectivity of the CC. Third, the mean certainty ratings of many optionally factive predicates are lower than those of many factive predicates and higher than those of main clauses as well as of non-veridical non-factives. However, Figure **??** also shows that the CCs of the 5 predicates assumed to be factive are not categorically more projective than the CCs of the optionally factive predicates, contrary to what is expected under the first definition of factive predicates. Specifically, the CCs of the optionally factive predicates *acknowledge, hear* and *inform* are at least as projective as the CCs of *reveal* or *discover*. This finding suggests that projectivity alone does not categorically distinguish factive predicates from optionally factive and non-factive ones.

These results support Tonhauser et al.'s (2018) hypothesis that prior content probability influences projectivity. The finding that the CC of many non-factive predicates is at least weakly projective, even with low prior probability CCs, confirms intuitions reported in, e.g., Schlenker 2010, Anand and Hacquard 2014 and Spector and Egré 2015. These findings motivate the development of projection analyses that derive the influence of prior content probability and make predictions for the CCs of a broad range of both factive and non-factive predicates.

Current projection analyses, while limited to the CCs of factive predicates (e.g., Heim 1983, van der Sandt 1992, Abrusán 2011, Simons et al. 2010, Beaver et al. 2017), are compatible with the finding that prior content probability influences projectivity. Heim 1983, for instance, assumes default global accommodation when a presupposition is not entailed by the common ground (CG) when the trigger is uttered. This default is overridden when the presupposition is inconsistent with the CG. If we can assume that Julian dancing salsa is more likely to be consistent with the CG when Julian is Cuban than when he is German, Heim 1983 predicts that the presupposition that Julian dances salsa is more projective when it has a higher prior probability.

As shown in Fig. **??**, the CCs of several non-factives, including *inform, hear, acknowledge* and *admit*, are at least as projective as that of factive *reveal*. This finding challenges the long-standing assumption that the CCs of factives are more projective than those of non-factives. We suggest that this motivates constraint-based analyses that derive the projectivity of utterance content from the integration of multiple cues, including prior CC probability, the meanings of clause-embedding predicates, at-issueness (Tonhauser et al. 2018), and information structure (Tonhauser 2016).

discussion: measure prior and projection at participant-level (as we do in next work)

5 Conclusions

References

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

Anand, Pranav, Jane Grimshaw, and Valentine Hacquard. 2019. Sentence embedding predicates, factivity and subjects. In C. Condoravdi and T. Holloway King, eds., *Tokens of meaning: Papers in honor of Lauri Karttunen*. Stanford: CSLI Publications.

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

Anderson, Lloyd B. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In W. Chafe and J. Nichols, eds., *Evidentiality: The Linguistic Coding of Epistemology*, pages 273–312. New Jersey: Ablex Publishing Corporation.

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

Beaver, David, Craige Roberts, Mandy Simons, and Judith Tonhauser. 2017. Questions Under Discussion: Where information structure meets projective content. *Annual Review of Linguistics* 3:265–284.

Bicknell, Klinton and Hannah Rohde. 2014. Dynamic integration of pragmatic expectations and real-world event knowledge in syntactic ambiguity resolution. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 1216–1221. Austin, TX: Cognitive Science Society.

Boyce, V, Titus von der Malsburg, Tim Poppels, and Roger Levy. 2018. Implicit gender in the production and comprehension of pronominal references. In *Proceedings of the 31th Annual CUNY Conference on Human Sentence Processing*.

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

Chambers, Craig G., Michael K. Tanenhaus, Kathleen M. Eberhard, Hana Filip, and Greg N. Carlson. 2002. Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language* 47:30–49.

Chambers, Craig G., Michael K. Tanenhaus, and James S. Magnuson. 2004. Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology* 30:687–696.

Degen, Judith, Michael Henry Tessier, and Noah D. Goodman. 2015. Wonky worlds: Listeners revise world knowledge when utterances are odd. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, page ???? Austin, TX: Cognitive Science Society.

Egré, Paul. 2008. Question-embedding and factivity. In L. Lihoreau, ed., *Grazer Philosophische Studien*, pages 85–125. Amsterdam: Rodopi.

Hagoort, Peter, Lea Hald, Marcel Mastiaansen, and Karl Magnus Petersson. 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 304:438–441.

Hald, Lea A., Esther G. Steenbeek-Planting, and Peter Hagoort. 2007. The interaction of discourse context and world knowledge in online sentence comprehension. evidence from the n400. *Brain research* 1146:210–218.

Hanna, Joy E. and Michael K. Tanenhaus. 2004. Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science* 28:105–115.

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

Jaynes, Edwin and Oscar Kempthorne. 1976. Confidence intervals vs. Bayesian intervals. In W. L. Harper and C. A. Hooker, eds., *Foundations of probability theory, statistical inference, and statistical theories of science*, pages 175–257. Dordrecht: Springer Netherlands.

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.

Lorson, Alexandra. 2018. The influence of world knowledge on projectivity. M.A. thesis, University of Potsdam.

Mahler, Taylor. 2020. The social component of projection behavior of clausal complements. In *LSA 2020 XXX*.

Morey, Richard D., Rink Hoekstra, Jeffrey N. Rouder, Michael D. Lee, and Eric J. Wagenmakers. 2016. The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin and Review* 23(1):103–123.

Nicenboim, Bruno and Shravan Vasishth. 2016. Statistical methods for linguistic research: Foundational Ideas, Part II. *Linguistics and Language Compass* 10(11):591–613.

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

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

Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117:1034–1056.

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

Smithson, Michael and Jay Verkuilen. 2006. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods* 11(1):54.

Spector, Benjamin and Paul Egré. 2015. A uniform semantics for embedding interrogatives: *An* answer, not necessarily *the* answer. *Synthese* 192:1729–1784.

Stevens, Jon Scott, Marie-Catherine de Marneffe, Shari R. Speer, and Judith Tonhauser. 2017. Rational use of prosody predicts projectivity in manner adverb utterances. *Annual Meeting of the Cognitive Science Society* 39:1144–1149.

Swanson, Eric. 2012. Propositional attitudes. In K. von Heusinger, C. Maienborn, and P. Portner, eds., *Semantics: An International Handbook of Natural Language Meaning*, vol. 2, pages 1538–1561. Berlin: Mouton de Gruyter.

Tessler, Michael Henry and Noah D. Goodman. 2019. The language of generalization. *Psychological Review* 126:395–436.

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

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 and Judith Degen. ms. Which predicates are factive? An empirical investigation. Manuscript, The Ohio State U and Stanford U.

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

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*, page ??

Wyse, Brendan. 2010. *Factive/non-factive predicate recognition within Question Generation systems*. Master's thesis, Open University.

Supplemental material for *Higher-probability content is more projective than lower-probability content*

A Experiment 1 target and control stimuli

The target stimuli of the projection block of Exp. 1 consisted of 800 combinations of a predicate, a complement clause and a fact relative to which the content of the complement clause had a higher or lower probability. The following list gives the 20 clauses that served as complement clauses together with the two facts that the clause was paired with: first the lower probability fact, then the higher probability one. The target stimuli of the prior block of Exp. 1 consisted of the 40 combinations of a clause and a fact.

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

In the target stimuli of the projection block of Exp. 1, 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*. The subject of the clause-embedding predicate and the speaker of the target stimuli were realized by a proper name.

The following list gives the six control stimuli of the projection block of Exp. 1:

1. **Fact:** Zack is a member of the golf club. Is Zack coming to the meeting tomorrow?
2. **Fact:** Mary visited her aunt on Sunday. Is Mary's aunt sick?
3. **Fact:** Todd goes to the gym 3 times a week. Did Todd play football in high school?
4. **Fact:** Vanessa won a prize at school. Is Vanessa good at math?
5. **Fact:** Trish sent Madison a card. Did Madison have a baby?
6. **Fact:** Hendrick just bought a car. Was Hendrick's car expensive?

B Model details for Experiments 1, 2 and 3

This supplement provides details on the data analysis conducted for Exps. 1, 2, and 3. We first motivate the use of Beta regression rather than linear regression in Exps. 1a, 2a, and 3a (section B.1) and then provide a brief primer on how to interpret Bayesian mixed effects Beta regression models (section B.2). We then report the model outputs for Exps. 1, 2, and 3 (section B.3).

B.1 Motivation for using Bayesian mixed effects Beta regression

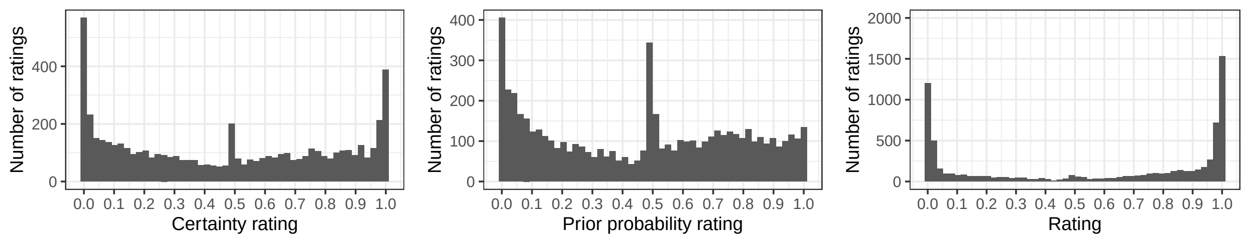
There are three separate pieces to motivate: the use of *mixed effects*, the use of a *Bayesian* rather than *frequentist* models, and the use of *Beta regression* rather than *linear regression*.

Using mixed effects refers to the practice of modeling the outcome variable, here slider ratings or proportions of 'yes' ratings, as a function of not just fixed effects of interest (i.e., predicate) but also as the result of possible random variability that is not of theoretical interest (e.g., random by-participant or by-item variability). This is standard practice in psycholinguistic studies and allows the researcher to trust that any observed effects of theoretical interest are true average effects rather than the result of idiosyncratic behavior (e.g., of participants or items). This is also the motivation for using mixed effects in Exps. 1b, 2b, and 3b.

Using Bayesian models rather than frequentist models is increasingly becoming the norm in psycholinguistic studies as computational power has increased and running Bayesian models has become more accessible with the introduction of R packages such as `brms` (Bürkner 2017). The presence of an effect in frequentist models is evaluated by checking whether the p-value is smaller than .05, where the p-value is defined as the probability of obtaining data that is as skewed or more skewed than the observed data if the null-hypothesis was true, i.e., if the hypothesized effect was absent. Parameter estimates in frequentist models are obtained via maximum-likelihood techniques, i.e., by estimating the parameter values that maximize the probability of observing the data. Bayesian models, by contrast, return a full posterior distribution over parameter values that take into account not just the probability of the data under the parameter values, but also the prior probability of parameter values. In order to evaluate the evidence for an effect of a predictor of interest, one can report 95% credible intervals and the posterior probability $P(\beta < 0)$ or $P(\beta > 0)$ that the predictor coefficient β is either lower or greater than zero, depending on the direction of the expected effect. A 95% credible interval (CI) demarcates the range of values that comprise 95% of probability mass of the posterior beliefs such that no value inside the CI has a lower probability than any point outside it (Jaynes and Kempthorne 1976, Morey et al. 2016). There is substantial evidence for an effect if zero is (by a reasonably clear margin) not included in the 95% CI and $P(\beta > 0)$ or $P(\beta < 0)$ is close to zero or one. Posterior probabilities indicate the probability that the parameter has a certain value, given the data and model – these probabilities are thus *not* frequentist p-values. In order to present statistics as close to widely used frequentist practices, and following Nicenboim and Vasishth 2016, we defined an inferential criterion that seems familiar (95%), but the strength of evidence should not be taken as having clear cut-off points (such as in a null-hypothesis significance testing framework).

Using Beta regression rather than linear regression was motivated by the violation of two of the assumptions of linear regression: first, that residuals be normally distributed (where "residuals" refers to the residual error for each data point after fitting the model), and second, that the error term exhibit homoscedasticity (that it be roughly the same across different conditions). Slider ratings data has the property of being bounded by its endpoints (which we code as 0 and 1, respectively). This often leads to "bunching" behavior at the endpoints (see Figure A1 for the distribution of raw ratings in Exps. 1a, 2a, and 3a).

This "bunching" behavior, in turn, can lead to the violation of both of the above assumptions of linear regression. Intuitively, these assumptions are violated because conditions that elicit ratings closer to endpoints necessarily have a compressed variance; consequently, a condition's mean and its variance are not independent. Beta regression is useful here because it allows for modeling an arbitrarily distributed outcome



(a) Exp. 1 certainty ratings. (b) Exp. 1 prior ratings. (c) Exp. 2 prior ratings.

Figure A1: Histograms of raw slider ratings in Exps. 1 and 2.

variable in the [0,1] interval. The Beta distribution is characterized by two parameters, one capturing the mean μ of the distribution and one capturing its precision ϕ, a measure of dispersion. The greater the precision, the more concentrated the values are around the mean, i.e., the lower the variance of the distribution. We follow Smithson and Verkuilen (2006) in modeling μ and ϕ separately for each predictor. That is, we allow each predictor to affect both the mean and the precision of the outcome variable's distribution.

B.2 Coding choices and interpreting model output

The outcome variable in Exps. 1a, 2a and 3a (slider ratings) contained the values 0 and 1, which Beta regression is undefined for. We therefore applied a common transformation to ratings before the main analysis that rescales values y to fall in the open unit interval (0,1) (Smithson and Verkuilen 2006). First, we apply $y' = (y - a)/(b - a)$, where b is the highest possible slider rating and a is the smallest possible slider rating. The range is then compressed to not include 0 and 1 by applying $y'' = [y'(N - 1) + 1/2]/N$, where N is the total number of observations.

The mean parameter μ is modeled via a logit link function (default for Beta regression in `brms`), though other links that squeeze μ into the [0,1] interval are possible. The dispersion parameter ϕ is modeled via a log link, which ensures that values of ϕ are strictly positive, which is necessary because a variance cannot be negative.

We allowed both μ and ϕ to vary as a function of predicate, with reference level set to main clause control in Exp. 1a, entailing control in Exp. 2a and contradictory control in Exp. 3a. We also allowed random intercept adjustments to each parameter by participant and by item, where item was defined as a unique combination of a predicate and a complement clause. Four chains converged after 2000 iterations each (warmup = 1000, $\hat{R} = 1$ for all estimated parameters) with a target acceptance rate of .95 and a maximum treedepth of 15.

B.3 Model outputs for Experiments 1, 2 and 3

The three tables in this section show the model outputs for Exps. 1, 2 and 3, respectively: Table **??** for Exps. 1a and 1b, Table **??** for Exps. 2a and 2b, and Table **??** for Exps. 3a and 3b. Each table shows maximum a posteriori (MAP) model estimates for projection ratings from the Beta regression model (left and middle column, mean μ and precision ϕ) and the logistic regression model (right column, β) with 95% credible intervals.

C Projection comparison

Projection: Comparison of findings of Exp. 1 and Tonhauser and Degen's ms Exp. 1a

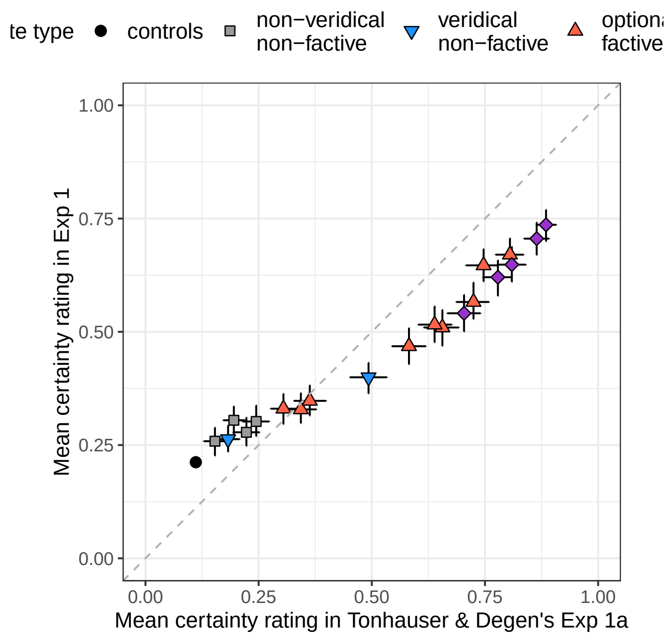


Figure A2: Mean by-predicate certainty ratings in Exp. 1 and Tonhauser and Degen's ms Exp. 1a. Error bars indicate 95% bootstrapped confidence intervals.

D Visualization of the effect of prior probability on projection by content

E Projection experiment

E.1 Methods

Participants 300 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform (ages: 21-72, median: 36; 145 female, 154 male, 1 undeclared). They were paid 85 cents for participating.

Materials and procedure The materials and procedure of this experiment were identical to those of the projection block of Exp. 1.

Data exclusion Prior to analysis, the data from 23 participants who did not self-identify as native speakers of American English were excluded. To assess whether the remaining 277 participants attended to the task, we inspected their responses to the 6 control stimuli. There were 11 participants whose response means were more than 2 standard deviations above the group mean. We excluded the data from these participants, too, leaving data from 266 participants (ages 21-72; median: 36; 129 female, 136 male, 1 undeclared).

E.2 Results

Figure A4 shows the mean certainty ratings for the target stimuli by predicate and by the prior probability of the content of the complement; the mean certainty rating for the non-projective main clause controls are also included. As shown, mean certainty ratings were higher when the content had a higher prior probability than when it had a lower prior probability, as hypothesized by Tonhauser et al. (2018). This finding suggests

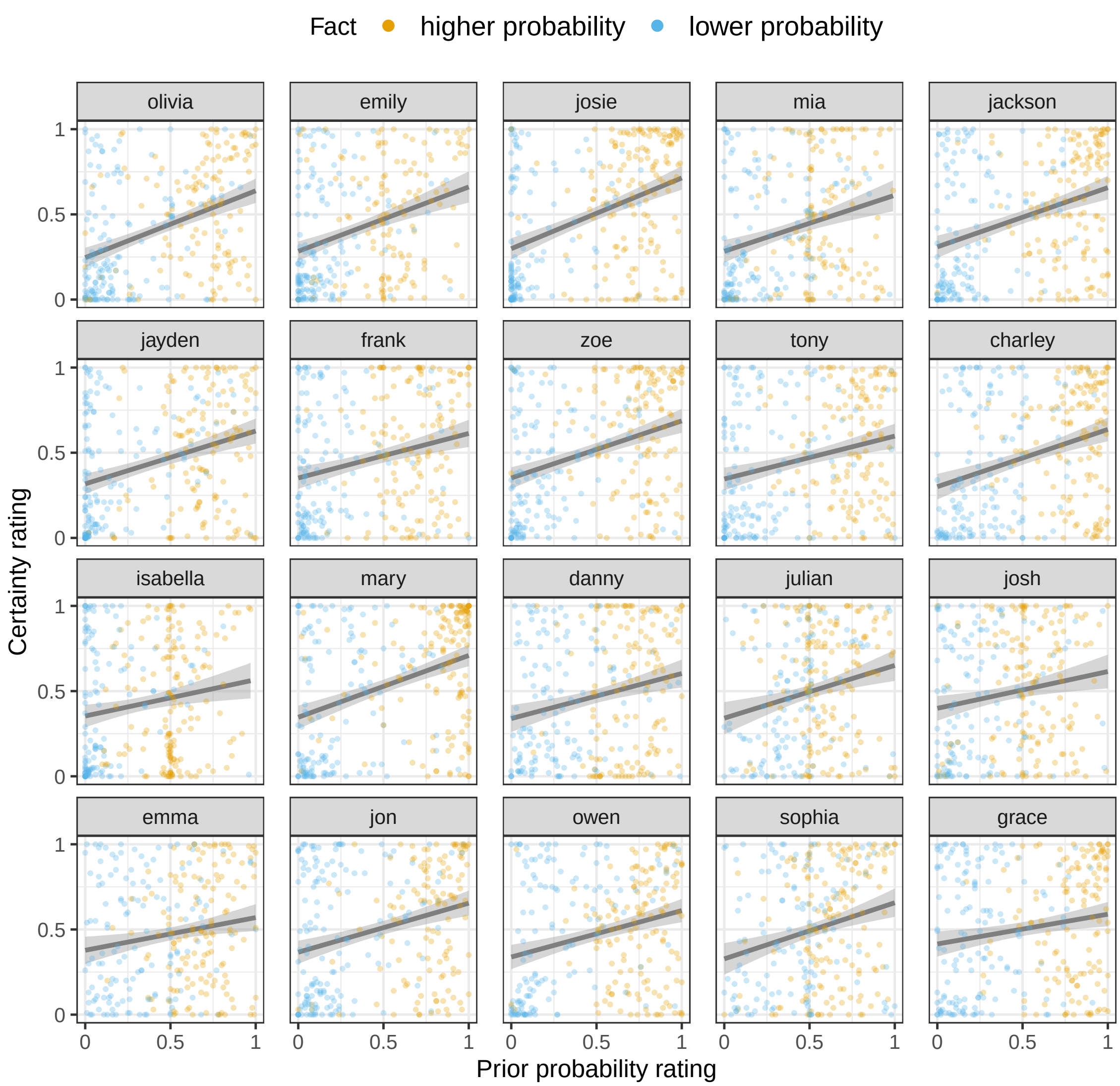


Figure A3: Certainty ratings by prior probability ratings by content in Exp. 1. Linear smoothers with 95% confidence intervals are overlaid.

that prior content probability influences projection. Furthermore, the effect of prior content probability on projectivity was present across all 20 clause-embedding predicates.

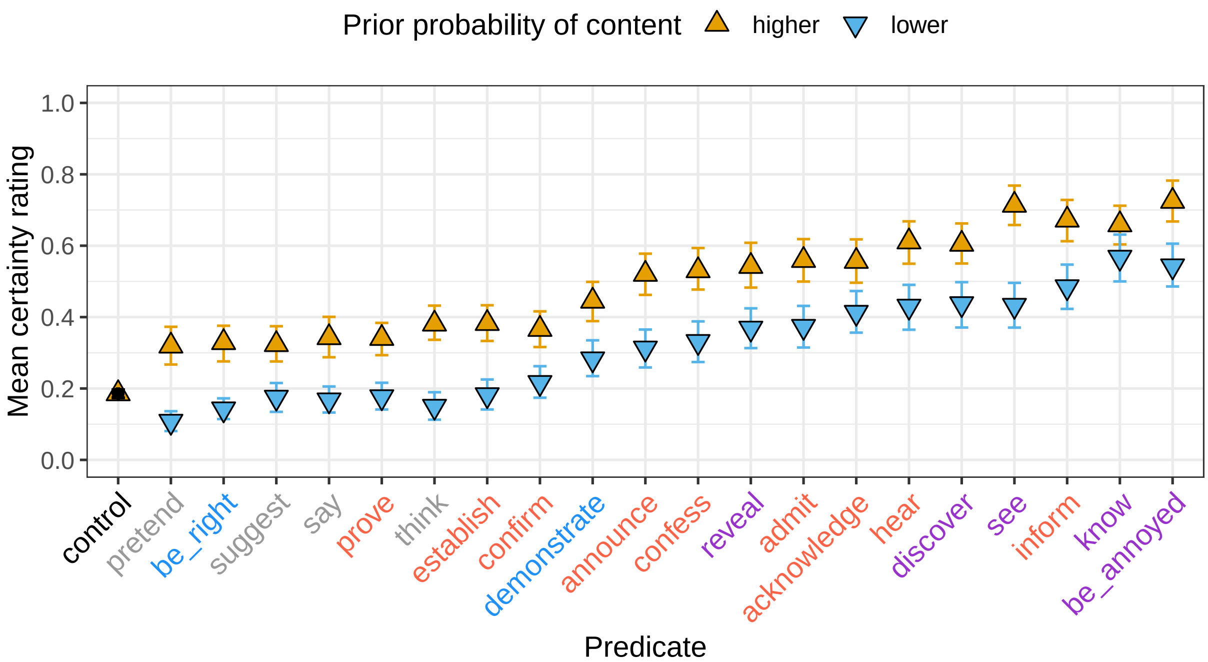


Figure A4: Mean certainty ratings by predicate and prior probability of the content of the complement. Error bars indicate 95% bootstrapped confidence intervals.

JT: report analyses here: prior from Exp1/projection from this experiment