Prior probability predicts projection

Author(s)

January 4, 2021

Abstract

Interpreters' beliefs about the world have been shown to influence utterance processing and interpretation in a variety of empirical domains. This paper reports three experiments designed to investigate the hypothesis that such beliefs influence projection, that is, inferences about speaker commitment to non-entailed utterance content. We find, in support of the hypothesis, that the higher the prior probability of content, the more projective it is. Prior probabilities predict projection both at the group and at the by-participant level. **JT: (4,098 words)**

1 Introduction

Interpreters' beliefs about the world (including beliefs about the speaker) 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). This paper provides empirical evidence from American English that interpreters' beliefs influence projection, that is, inferences about speaker commitment to non-entailed utterance content.

Inferences about speaker commitment are illustrated in (1). A speaker who utters (1a) is typically taken to be committed to the truth of the content of the complement of *know*, that it is raining. This inference is captured by assuming that the content of the complement (henceforth, CC) of *know* is an entailment. In (1b) and (1c), *know* is realized in a polar question and in the scope of negation, respectively: Because the CC is not entailed here, inferences that a speaker who utters (1b) or (1c) is committed to the CC can therefore not be due entailment. Instead, such inferences are attributed to the projection of the CC (e.g., Langendoen and Savin 1971, Beaver and Geurts 2014).

(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.

Projection is gradient, as interpreters' inferences about speaker commitment to utterance content can vary in strength and several factors, including the discourse status of the content and the prosody of the utterance, have been observed to influence projection (for an overview see Tonhauser et al. 2018). The hypothesis that interpreters' beliefs influence projection was initially put forth in Stevens, de Marneffe, Speer, and Tonhauser 2017 and Tonhauser, Beaver, and Degen 2018[jd: also mahler and alexandra's work?]**JT: added "initially" and left out those works: Stevens et al and TBD put forth the hypothesis, without investigating it. please change as you see fit.**. These works observed by-item projection variability such that, for instance, the CC of *know* is more projective when it is *Kim flew to New York* than when it is *Kim* *flew to the moon* [jd: this was just an example, right? not one that we explicitly tested?]**JT: yes.**. Both works hypothesized that "the projectivity of content may depend on the prior probability of the event described" (Tonhauser et al. 2018:500), such that more a priori likely content may be more likely to project:

(2) **Hypothesis:** The higher the prior probability of content, the more projective the content.

One piece of evidence for the hypothesis in (2) comes from Mahler (2020), who investigated the projection of politically charged CCs of two classes of English clause-embedding predicates: factive ones (represented by *know, discover, realize, see*) and non-factive ones (represented by *believe, think, feel*). For example, the politically charged content in (3) is that Obama improved/damaged the American economy. The prior probability of the content was manipulated by the speaker (Cindy in (3)) speaking at the club meeting of either the College Republicans or Democrats.

(3) Cindy, at the College Republicans/Democrats club meeting:
Ben doesn't know that...
 a. ...Obama improved the American economy.
 b. ...Obama damaged the American economy. (Mahler 2020:784f.)

Consistent with (2), Mahler (2020) found that higher prior probability content (e.g., a liberal content like (3a) uttered by a Democrat) was more projective than a lower prior probability content (e.g., a liberal content uttered by a Republican). The effect was observed for both classes of predicates.

In contrast to Mahler 2020, Lorson 2018 did not find empirical support for the hypothesis in (2). Lorson (2018) investigated the influence of prior probability on the projection of the pre-state content of the English change of state verb *stop*. Prior probability was manipulated through gender stereotypes reported in Boyce et al. 2018. For instance, because men are more likely than women to be plumbers, the pre-state content of (4a), that James has worked as a plumber, was hypothesized to be more projective than the pre-state content of (4b), that Linda has worked as a plumber.

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

Several differences between Mahler's and Lorson's investigations could be implicated in the differential support for the hypothesis in (2): a) the projective content investigated (contents of clausal complements vs. pre-state content of *stop*), b) the entailment-canceling environment (negation vs. polar question), c) the manipulation of the prior probability (political party affiliation vs. gender stereotypes), and d) how explicitly the prior-manipulating information was provided to the participants (statement of political party affiliation vs. use of a male or female name to indicate gender). The three experiments we report on in this paper investigate the hypothesis in (2) on the basis of the projection of the contents of complements of clause-embedding predicates that are embedded in polar questions. In contrast to Mahler 2020, our experiments include 20 clause-embedding predicates (rather than just 7) and we do not assume a factivity classification of the predicates because of the lack of empirical support for a class of factive predicates (Tonhauser and Degen ms). Our experiments manipulate the prior probability of content in a variety of ways. The prior-manipulating information was explicitly provided.

Both Mahler 2020 and Lorson 2018 measured projection using the 'certain that' diagnostic (see, e.g., Tonhauser 2016, Tonhauser et al. 2018, de Marneffe et al. 2019): Participants were asked whether the speaker is certain of the relevant content and they responded on a sliding scale, with one end was labeled negatively and the other positively. Both investigations assumed that the more positive the response, the more projective the content. Our experiments also used the 'certain that' diagnostic with such a sliding scale. Exp. 1 (section 2) investigated the hypothesis in (2) by measuring prior probability and projection in a within-participant design. This design allows us to investigate whether the hypothesis in (2) holds at the by-participant level. In Exps. 2 (section 3), prior probability and projection were measured from separate groups, as in Mahler 2020 and Lorson 2018. We find support for the hypothesis that prior probability influences projection in both Exps. 1 and 2. The paper offers concluding remarks in section 4.

2 Experiment 1

This experiment investigated the hypothesis in (2), that the higher the prior probability of content, the more projective the content. Prior probability and projection ratings were collected for the contents of 20 clauses that realized the complements of 20 clause-embedding predicates.[1]

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.

Materials The prior probability and the projection of the contents of 20 clauses were measured. As shown for the sample clause in (5), each clause was paired with two facts. One fact was hypothesized to result in higher prior probability for the content and the other in a lower prior probability. See Supplement A for the full set of clauses and facts.

(5) Sample clause: Julian dances salsa.
Higher probability fact: Julian is Cuban.
Lower probability fact: Julian is German.

Prior probability and projection were measured in separate blocks. In the prior block, the 20 clauses were realized as the complements of *How likely is it that... ?* questions. As shown in (6), each target stimulus consisted of one of the two facts for that clause and the *How likely is it that... ?* question. There were a total of 40 target stimuli in the prior block.

(6) 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, the target stimuli consisted of a fact and a polar question that was uttered by a named speaker, as shown in (7). The fact was one of the two facts that the clause was paired with. The polar questions were formed by realizing the 20 clauses as the complements of the 20 clause-embedding predicates in (8). There were a total of 800 target stimuli in the projection block.

(7) 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?

(8) *be annoyed, discover, know, reveal, see, acknowledge, admit, announce, confess, confirm, establish, hear, inform, prove, pretend, suggest, say, think, be right, demonstrate*

[1]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]. Exp. 1 was pregistered: [link removed for review]. All experiments were conducted with approval from the IRB of [university redacted] and informed consent was obtained.

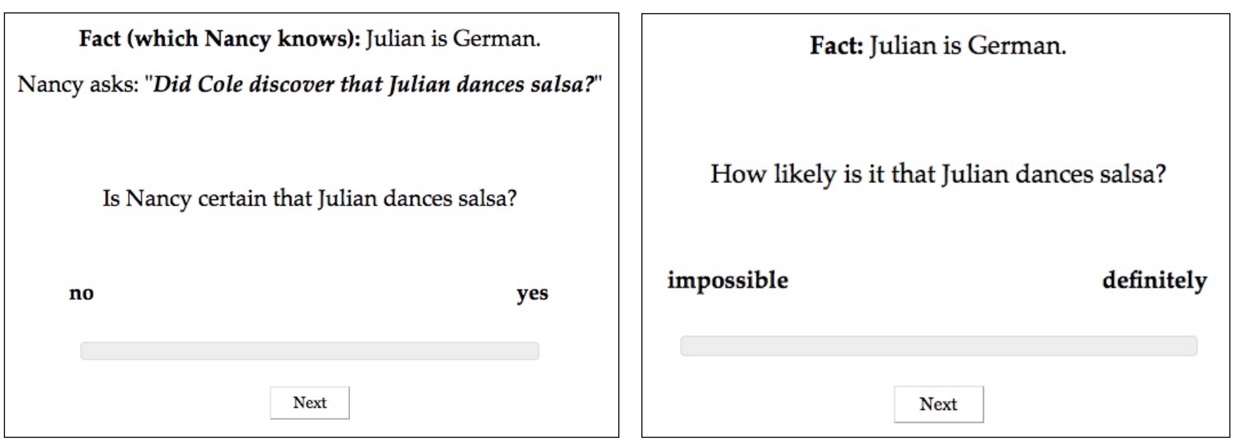
The 20 predicates in (8) include a cross-section of English clause-embedding predicates: They include cognitive predicates (e.g., *know*), emotive predicates (e.g., *be annoyed*), communication predicates (e.g., *announce*), and inferential predicates (e.g., *prove*), as well as so-called factive and non-factive predicates (e.g., *know* vs. *think*); as mentioned above, Tonhauser and Degen ms challenge this latter classification.

The projection block also included 6 control stimuli, which were used to assess participants' attention to the task. The content of these stimuli was hypothesized to not project: For example, in (9), the speaker is not committed to the main clause content, that Zack is coming to the meeting tomorrow. The same 6 main clauses were also used to form the 6 filler stimuli in the prior block; these were not used to assess participants' attention. For the full set of stimuli see Supplement A.

(9) 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?

For each participant, a set consisting of the 20 clauses with one of their facts was randomly generated, with 10 higher 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 prior block, participants rated the prior probability of 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. Here, participants rated the projection of the contents of the 20 clauses given the facts. 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 Data was excluded based on self-declared non-native speaker status and other criteria given in Supplement B, leaving data from 286 participants to be analyzed (ages 18-82; median: 35.5; 116 female, 186 male, 1 other, 1 undeclared).

2.1 Results

The manipulation of the prior probability of the 20 contents was successful. Figure 2 plots the mean prior probabilities of the 20 contents by fact. For each content, the mean prior probability rating was higher when it was presented with the higher probability fact than when it was presented with the lower probability fact.

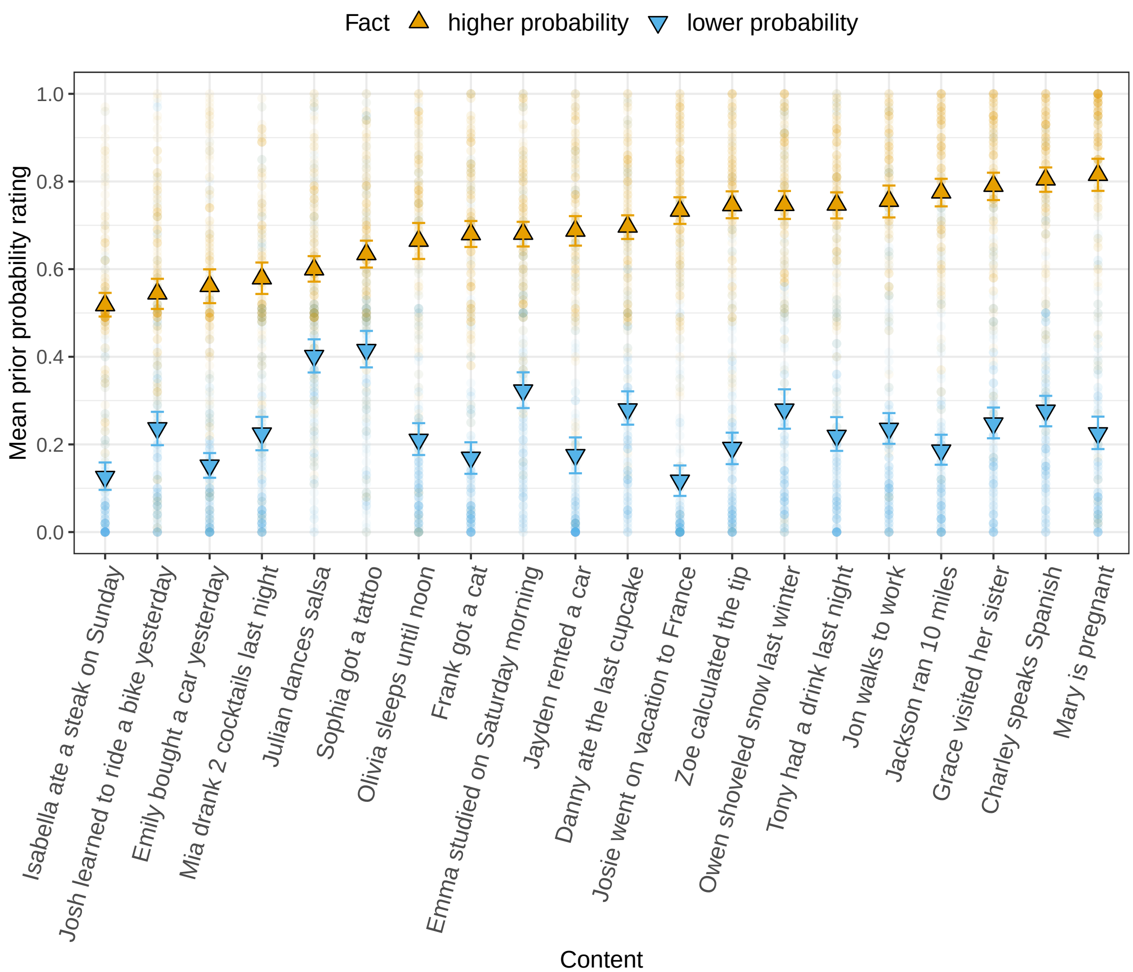


Figure 2: Mean prior probability by content and fact in Exp. 1. Error bars indicate 95% bootstrapped confidence intervals. Light dots indicate participants' ratings.

To address the question of whether prior probability influences projection, Figure 3 plots the mean certainty ratings for the CCs by predicate and by fact, as well as the mean certainty rating for the main clause controls (abbreviated 'MC'). As shown, the mean certainty ratings were higher for contents presented with higher probability facts than for contents presented with lower probability facts. This result, which holds for all 20 clause-embedding predicates, suggests that prior probability influences projection.

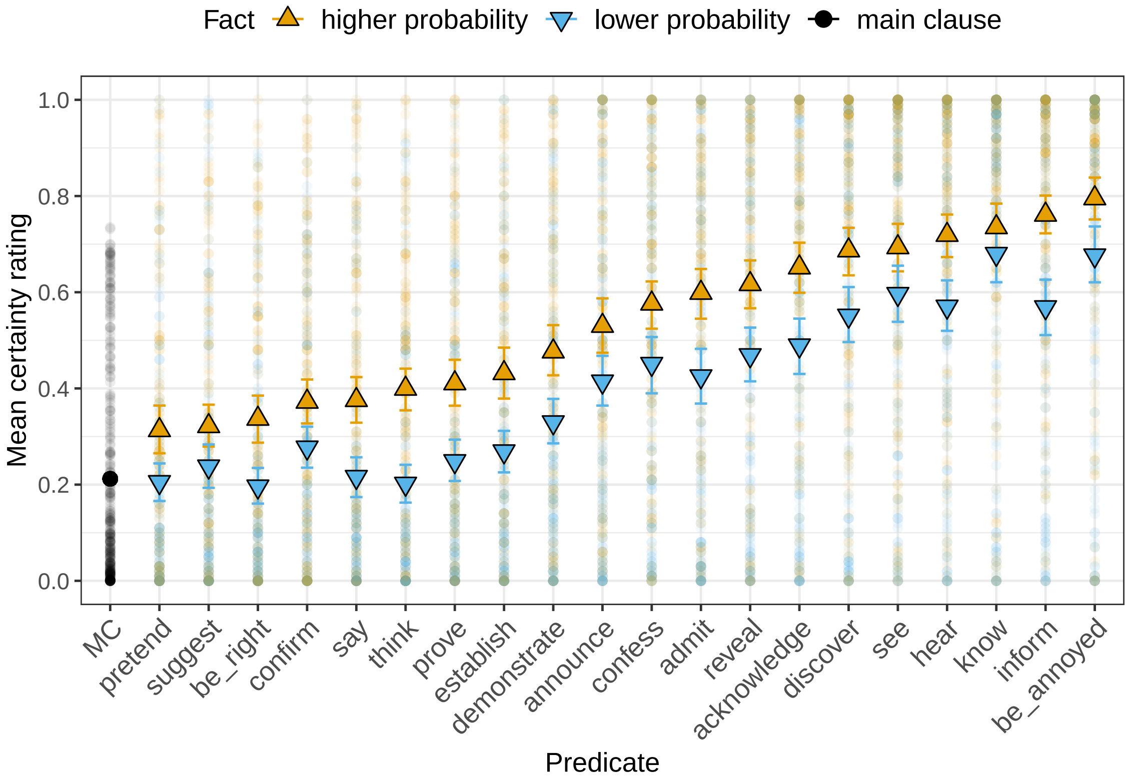


Figure 3: Mean certainty ratings by predicate and prior probability of the content of the complement in Exp. 1. Error bars indicate 95% bootstrapped confidence intervals. Light dots indicate participants' ratings.

We also replicate Tonhauser and Degen's ms result that there is by-predicate variation in the projection of the CC: for instance, the CC 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 (collapsing over facts) and Tonhauser and Degen's ms Exp. 1a is .991; see Supplement D for a visualization. Exp. 1 thereby provides further evidence for the systematic influence of the predicate on projection.

Figure 2 revealed by-participant variation in prior probability ratings, which means that participants' prior probability ratings need not align with the prior probability classification assumed in Figure 3. 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 4 plots the participants' certainty ratings by their prior probability ratings. (The color coding here merely represents the type of fact the participant was presented with. 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 result suggests that prior probability influences projection even at the by-participant level.

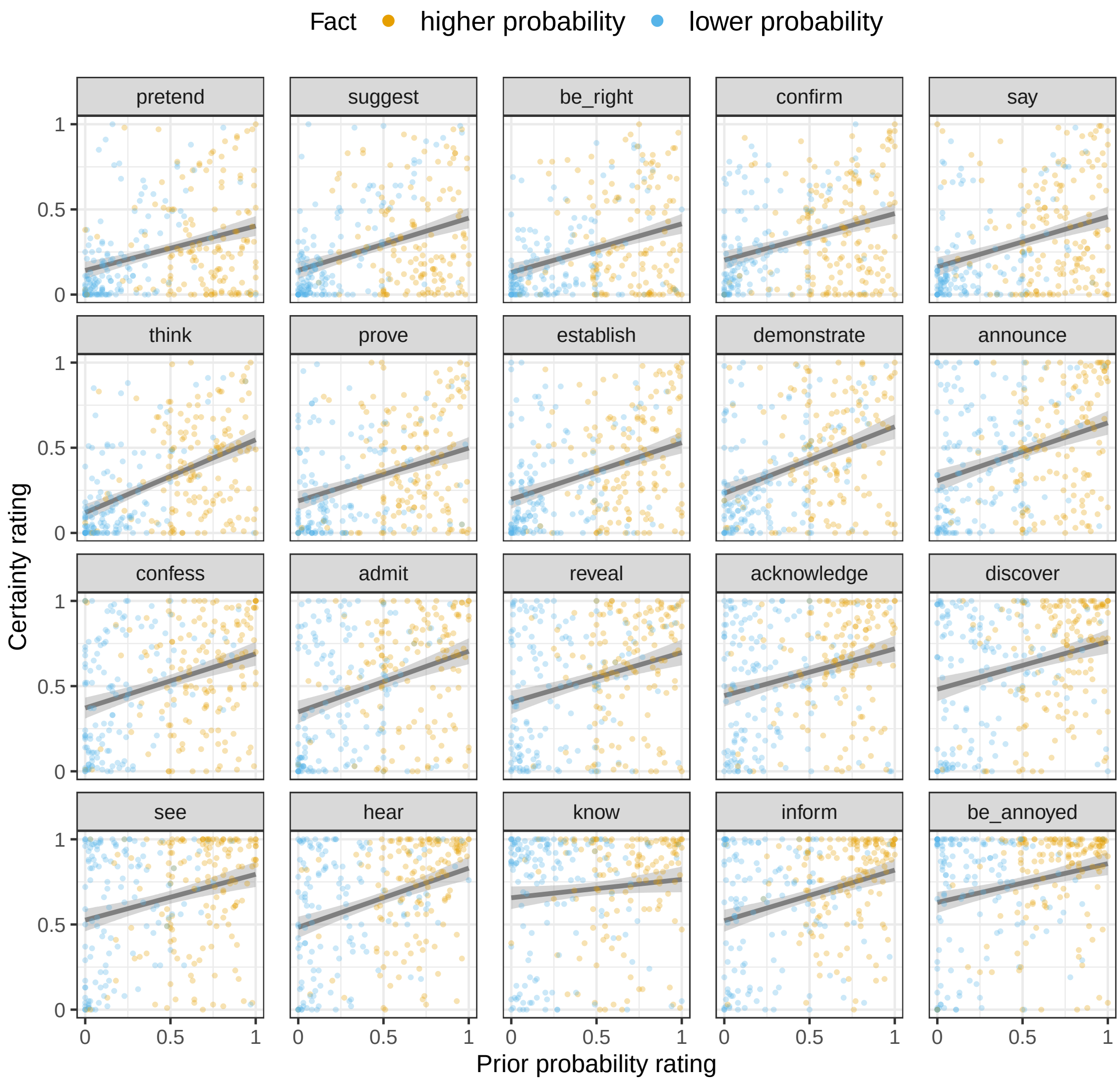


Figure 4: Certainty ratings by prior probability ratings by predicate in Exp. 1. Linear smoothers with 95% confidence intervals are overlaid.

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).[2] We thus obtain a 95% credible interval for **JT: the mean effect of prior probability on certainty?**. Supplement

[2]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 C.

C 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: We need to report two models here: predicting projection from prior at by-participant level, but also predicting mean projection from mean prior, to allow for comparison to Exps. 2

JT: According to the Beta regression model, the estimated mean for each predicate was higher when....

2.2 Discussion

The results of Exp. 1 provide empirical support for the hypothesis in (2), advanced in Stevens et al. 2017 and Tonhauser et al. 2018, that the higher the prior probability of content, the more projective it is.

These results confirm Mahler's 2020 results and expand on them in several ways. First, while Mahler 2020 manipulated only the political party affiliation of the speaker, the manipulation in Exp. 1 relied on 20 distinct properties of the individuals denoted by the subjects of the 20 clauses (e.g., whether Julian is more likely to dance salsa if he is German or Cuban, or whether Zoe is more likely to have calculated the tip, if she is 5 years old or a math major). Thus, the result of Exp. 1 suggests a general effect of prior probability on projection.[3]

Second, while Mahler 2020 observed an influence of prior probability of the projection of the CCs of two classes of predicates (so-called factive and non-factive predicates), Exp. 1 observed such an influence for the CCs of 20 predicates. Thus, the result of Exp. 1 supports the assumption that the effect of prior probability on projection is more general than was suggested in Mahler 2020. In fact, the results of Exp. 1 motivate the hypothesis that prior probability influences projection across the set of English clause-embedding predicates. Finally, prior probability and projection were measured in a within-participant design in Exp. 1, in contrast to Mahler's experiment, which only measured projection. Exp. 1 thus supports the claim that projection is influenced not only by the average prior probability of content but also by the prior probability that individual interpreters assign to content. This result suggests that by-participant variability in projection experiments (see, e.g., Tonhauser et al. 2018, Tonhauser and Degen ms) may be due to participants assigning different prior probabilities to the contents under investigation. It is possible, however, that the within-participant design resulted in participants' responses on either block influencing their responses on the other block. To mitigate against this possibility, we conducted Exps. 2, where prior probability and projection ratings were collected from different populations.

3 Experiments 2

Exps. 2a and 2b measured the prior probability and the projection of the 20 contents of Exp. 1, respectively.

3.1 Methods

Participants Participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform. The 95 participants in Exp. 2a (ages: 21-75, median: 33; 45 female, 50 male) were paid 55 cents. The 300 participants in Exp. 2b (ages: 21-72, median: 36; 145 female, 154 male, 1 undeclared) were paid 85 cents.[4]

[3]None of our prior probability manipulations relied on gender stereotypes because we only became aware of Lorson 2018 after running Exp. 2.

[4]28 participants took both experiments. Given that the experiments were run two weeks apart (Exp. 2a on November 13, 2017 and Exp. 2b on November 28, 2017, it is unlikely that these 28 participants' prior ratings influenced their projection ratings.

Materials and procedures The 40 target stimuli of Exp. 2a were identical to those of the prior block of Exp. 1. Each participants saw the two control stimuli in (9), which were included to assess attention to the task. We expected high prior probability ratings for (9a) and low ones for (9b).

(10) 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?

The materials of Exp. 2b were identical to those of the projection block of Exp. 1. Trial order in both experiments was random. The procedures of Exps. 2a and 2b were identical to those of the prior and projection blocks of Exp. 1, respectively.

Data exclusion We excluded data based on the criteria given in Supplement B, leaving data from 75 participants to be analyzed in Exp. 2a (ages 21-75; median: 35; 34 female, 41 male) and from 266 participants in Exp. 2b (ages 21-72; median: 36; 129 female, 136 male, 1 undeclared).

3.2 Results and discussion

Exp. 2a replicated the prior probability manipulation of Exp. 1. Figure 5 plots the mean prior probability ratings in Exp. 2a against those of Exp. 1. The Spearman rank correlation was high, at .977. For a visualization of the by-content prior ratings see Supplement E.

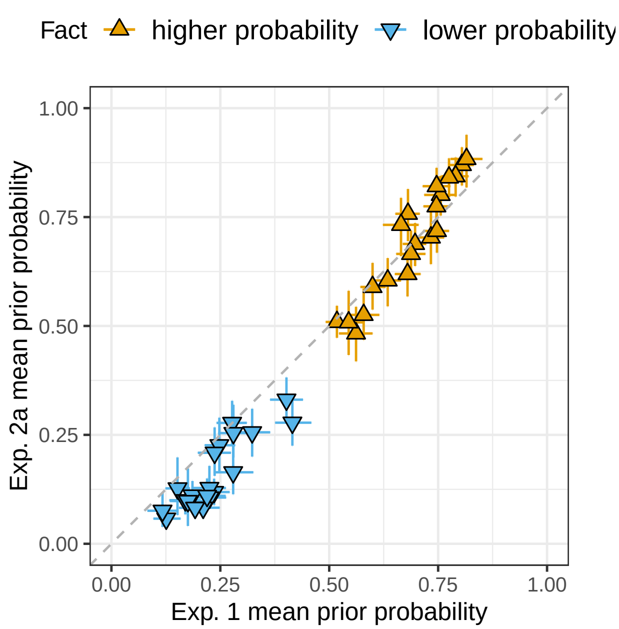


Figure 5: Mean prior probability ratings in Exp. 2a against those of Exp. 1. Error bars indicate 95% bootstrapped confidence intervals. The dotted line indicates **JT: what?**.

Exp. 2b replicated the critical result of Exp. 1, that prior probability influences projection.[5] Figure 6

[5]Exp. 2b also replicated Tonhauser and Degen's ms result that there is by-predicate variation in the projection of the content of the complement; see Supplement D.

shows that mean certainty ratings in Exp. 2b were higher for contents presented with higher probability facts than for contents presented with lower probability facts.

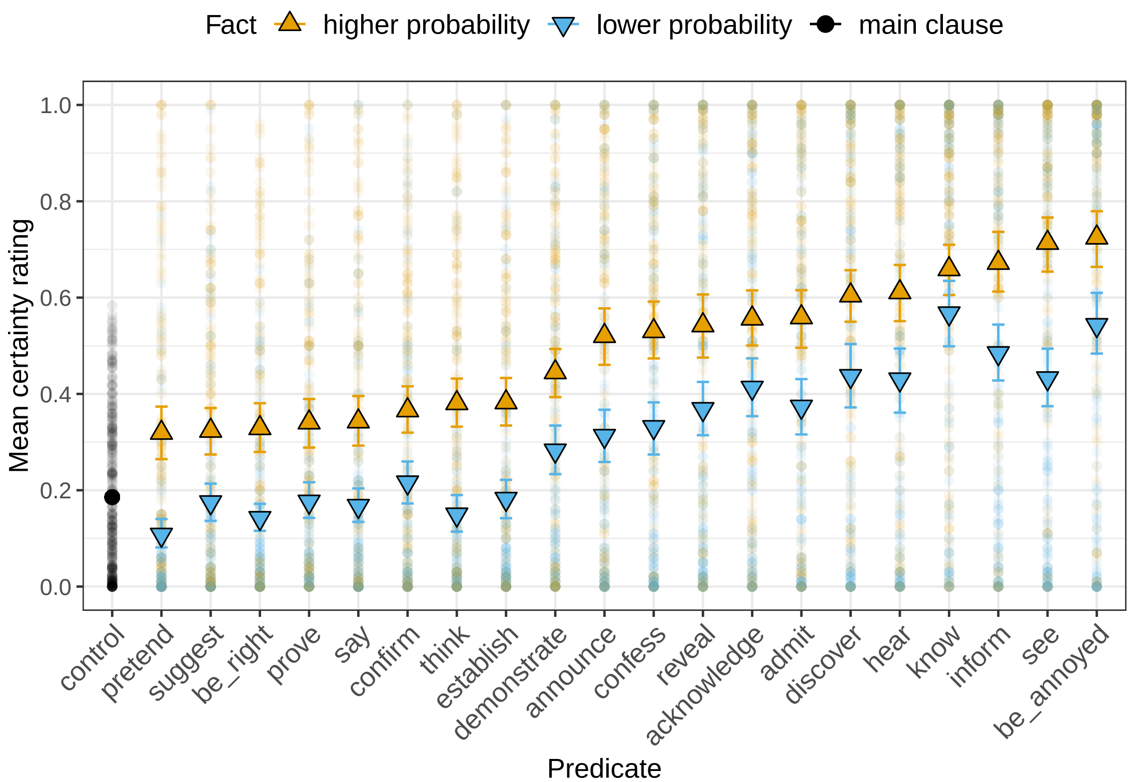


Figure 6: Mean certainty ratings by predicate and prior probability of the content of the complement in Exp. 2b. Error bars indicate 95% bootstrapped confidence intervals. Light dots indicate participants' ratings.

JT: report models predicting a) projection mean Exp 2b from prior mean Exp 2b, b) projection mean Exp 1 from prior mean Exp 2b, and c) projection mean Exp 2b from prior mean Exp 1). USE COHENS D? These results suggest that the result of Exp. 1 is not an artifact of the within-participant design of Exp. 1.

4 Concluding remarks

This paper provided empirical support for Stevens et al.'s (2017) and Tonhauser et al.'s (2018) hypothesis that interpreters' beliefs about utterance content influence their inferences about speaker commitment to that content. A pressing question for future research is how prior probabilities interact with other factors that have been shown to influence projection, such as the content's discourse status and utterance prosody. On the theoretical side, our result motivates the development of projection analyses that predict the influence of prior probability on the projection of the content of English clause-embedding predicates. Analyses currently on the market do not predict the results of our experiments because they are limited to subsets of clause-embedding predicates, like factive ones (e.g., Heim 1983, van der Sandt 1992, Abrusán 2011, 2016, Romoli 2015, Simons et al. 2017).

References

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

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

Beaver, David and Bart Geurts. 2014. Presupposition. In E. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Stanford University.

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.

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

Degen, Judith, 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.

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.

Langendoen, Terry and Harris Savin. 1971. The projection problem for presuppositions. In C. Fillmore and T. Langendoen, eds., *Studies in Linguistic Semantics*, pages 54–60. New York: Holt, Rinehart and Winston.

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. *Linguistic Society of America* 5:777–791.

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.
Romoli, Jacopo. 2015. The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of Semantics* 32:173–291.
Simons, Mandy, David Beaver, Craige Roberts, and Judith Tonhauser. 2017. The Best Question: Explaining the projection behavior of factive verbs. *Discourse Processes* 3:187–206.
Smithson, Michael and Jay Verkuilen. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11(1):54.
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.
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.

Supplemental material for *Prior probability predicts projection*

A Experiment 1: Target and control stimuli

This list gives the 20 clauses of the target stimuli with the 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

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 clauses that were used in the control and filler stimuli of Exp. 1, with their facts. In the prior block, these six clauses were embedded under *How likely is it that… ?*. The projection block featured polar questions variants of the clauses.

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

B Data exclusion

Table A1 presents how many participants' data were excluded from the analyses based on the exclusion criteria. The first column records the experiment, the second ('recruited') how many participants were recruited, and the final column ('remaining') how many participants' data entered the analysis. The 'Exclusion criteria' columns show how many participants' data were excluded based on the two exclusion criteria:

- 'language': Participants' data were excluded if they did not self-identify as native speakers of American English.
- 'controls': In Exps. 1 and 2b, participants' data were excluded if their response mean on the 6 control items was more than 2 sd above the group mean. In Exp. 2a, participants' data were excluded if their response mean was more than 2 sd below the group mean of the control in (9a) or more than 2 sd above the group mean of the control in (9b).

		Exclusion criteria		
	recruited	language	controls	remaining
Exp. 1	300	3	11	286
Exp. 2a	95	8	12	75
Exp. 2b	300	23	11	266

Table A1: Data exclusion in Exps. 1 and 2

C Model details for Experiments 1 and 2

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 C.1) and then provide a brief primer on how to interpret Bayesian mixed effects Beta regression models (section C.2). We then report the model outputs for Exps. 1, 2, and 3 (section C.3).

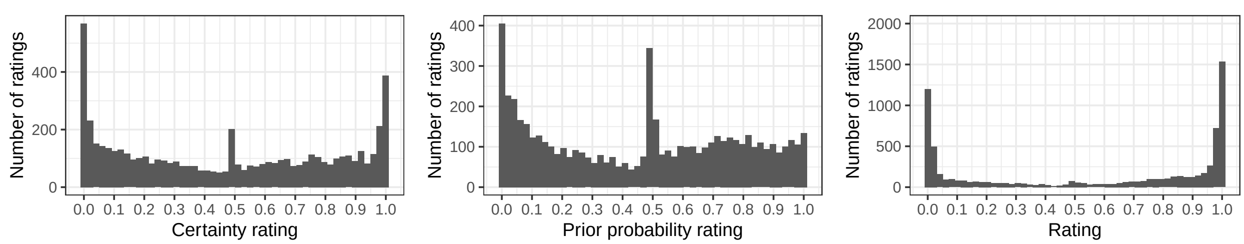
C.1 Motivation for using Bayesian mixed effects Beta regression

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).



(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.

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 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.

C.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.

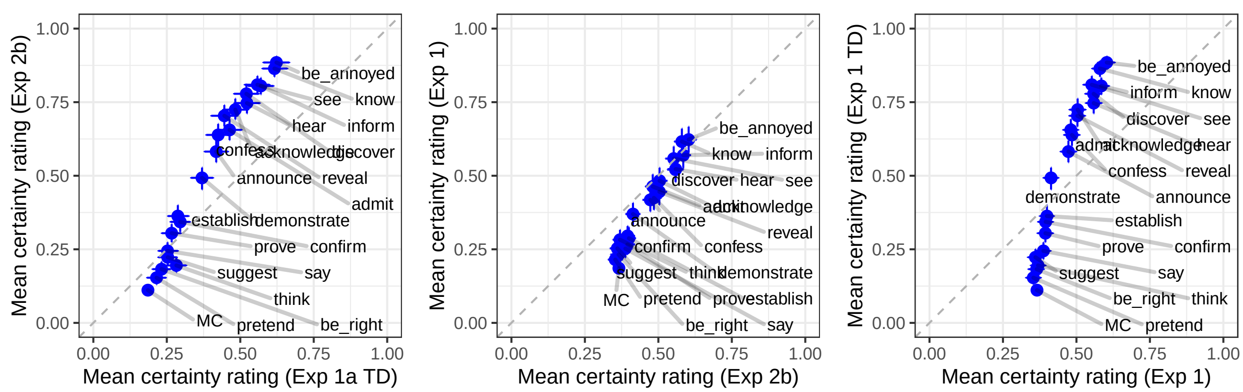
We allowed both μ and ϕ to vary as a function of predicate, with reference level set to main clause control in Exp. 1a, entailing control in Exp. 2a and contradictory control in Exp. 3a. We also allowed random intercept adjustments to each parameter by participant and by item, where item was defined 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.

C.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.

D Projection comparisons

Figure A2 compares the mean certainty ratings of the predicates and main clause controls in Exp. 1, Exp. 2b, and Tonhauser and Degen's ms Exp. 1a (abbreviated 'Exp. 1a TD'). The Spearman rank correlations were .986 (Exp. 2b vs. Exp. 1a TD), .971 (Exp. 1 vs. Exp. 2b) and .974 (Exp. 1a TD vs. Exp. 1).



(a) Exp. 2b vs. Exp. 1a TD.

(b) Exp. 1 vs. Exp. 2b.

(c) Exp. 1a TD vs. Exp. 1.

Figure A2: Comparisons of mean by-predicate certainty ratings from Exp. 1, Exp. 2b, and Tonhauser and Degen's ms Exp. 1a (abbreviated 'Exp. 1a TD'). Error bars indicate 95% bootstrapped confidence intervals.

E Prior probability results in Exp. 2a

Figure A3 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.

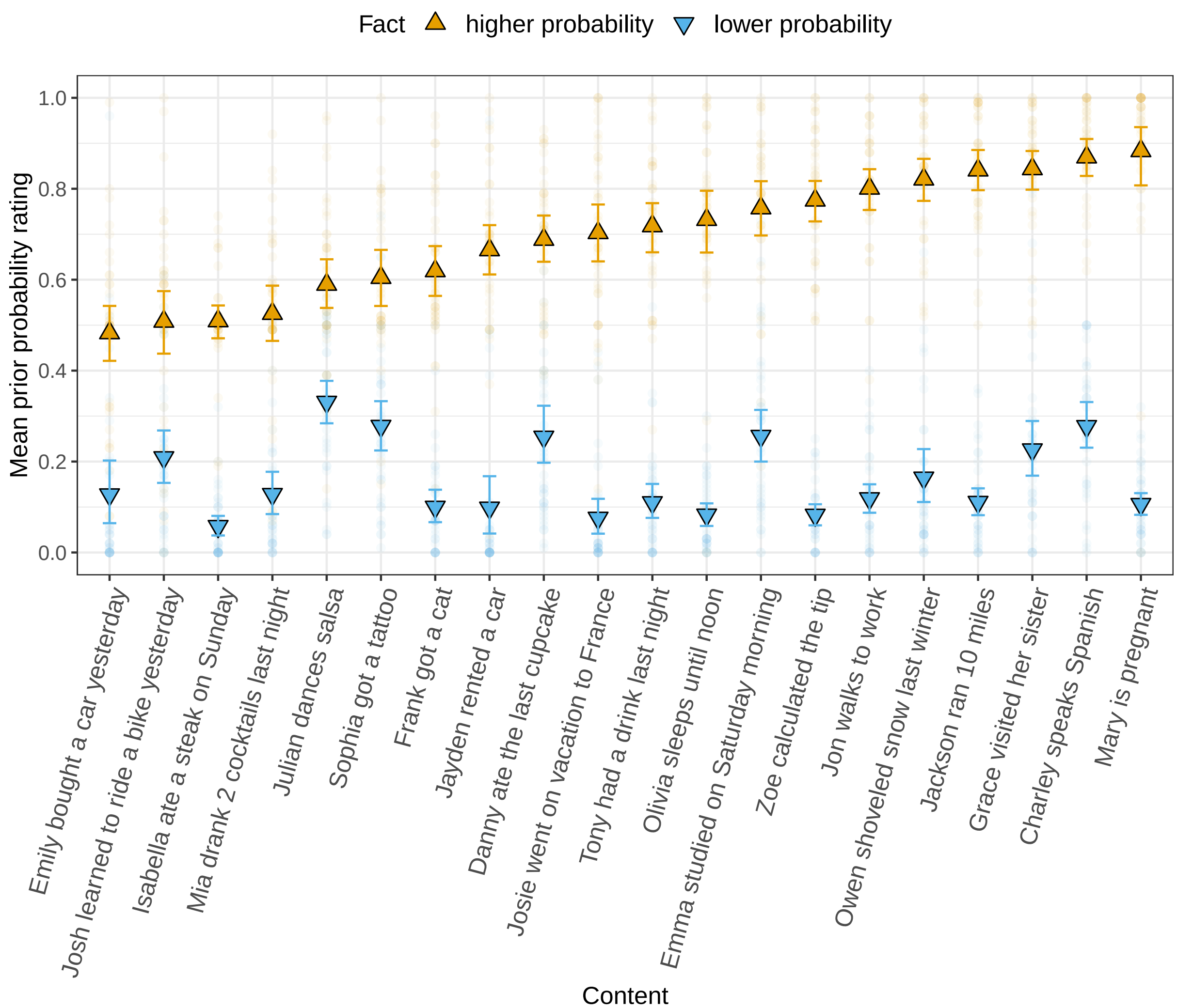


Figure A3: Mean prior probability by content and fact in Exp. 2a. Error bars indicate 95% bootstrapped confidence intervals. Light dots indicate participants' ratings.