Which predicates are factive? An empirical investigation

Author(s)

April 2, 2020

Abstract

Properties of the content of the clausal complement have long been assumed to distinguish factive predicates like *know* from non-factive ones like *think* (Kiparsky and Kiparsky 1970, i.a.). There is, however, disagreement about which properties define factive predicates as well as uncertainty about whether the contents of the complement of particular predicates exhibit the properties attributed to the content of the complements of factive predicates. This has led to a lack of consensus about which predicates are factive, a troublesome situation given the central role in linguistic theorizing that the distinction between factive and non-factive predicates has played. This paper reports six experiments designed to investigate two critical properties of the content of the complement of clause-embedding predicates, namely projection and entailment, with the goal of establishing whether these properties categorize clause-embedding predicates. We find that factive predicates are more heterogeneous than previously assumed and that there is little empirical support for the assumed categorical distinction between factive and non-factive predicates. We discuss implications of our findings for presuppositions, an area where the factive/non-factive distinction has played a central role. We propose that projection is sensitive to more fine-grained meaning distinctions between clause-embedding predicates than the factive/non-factive distinction. (Word count: 17,109)

1 Introduction

Kiparsky and Kiparsky's 1970 seminal paper categorized clause-embedding predicates like *know, think* and *acknowledge* into three classes based on whether the content of the clausal complement is presupposed, that is, whether "[t]he speaker presupposes that the embedded clause expresses a true proposition" (p.147). The paper categorized *know* as 'factive' because a speaker who utters the affirmative sentence in (1a) with *know*, its negative variant in (1b), or the question variant in (1c) presupposes that it is raining; projection of the content of the complement (henceforth abbreviated as CC) from under negation or out of a polar question was taken to diagnose the presuppositionality of the CC. The predicate *think*, on the other hand, was categorized as 'non-factive' because a speaker who utters the variants in (1a-c) with *think* does not presuppose the CC. Finally, the paper categorized *acknowledge* as neither factive nor non-factive because a speaker who utters the variants in (1a-c) with *acknowledge* may but need not presuppose that it is raining (p.163).[1] In the following, we refer this third class of predicates as 'optionally factive'; the need to invent a name for this class comes from the fact that, nowadays, optionally factive predicates are typically lumped in with the non-factive ones.

(1)   a.   Sam { knows / thinks / acknowledges } that it is raining.

     b.   Sam doesn't { know / think / acknowledge } that it is raining.

     c.   Does Sam { know / think / acknowledge } that it is raining?

---

[1]Clause-embedding predicates also differ in syntactic properties. We ignore syntax here because Kiparsky and Kiparsky (1970:fn.3) already pointed out that syntactic properties do not align with projection: *know* and *realize*, for instance, were taken to be syntactically non-factive but semantically factive. For recent discussion see White and Rawlins 2018 and references therein.

Today, some 50 years after Kiparsky and Kiparsky 1970, properties of the CC, including projection, are still taken to categorically distinguish factive predicates from others, in English as well as in other languages. However, as we show below, there is disagreement about which properties define factive predicates and uncertainty about whether the CCs of particular predicates exhibit the properties attributed to factive predicates. As a result, there is no consensus on which predicates are factive. This lack of consensus is problematic given the central role that the categorization plays in linguistic theorizing on topics as diverse as, for instance, projection (e.g., Karttunen and Peters 1979, van der Sandt 1992), embedded questions (e.g., Hintikka 1975, Guerzoni and Sharvit 2007, Spector and Egré 2015), extraction (e.g., Hukari and Levine 1995, Rooryck 2000, Abrusán 2014), exclamatives (e.g., Zanuttini and Portner 2003), control (e.g., Landau 2001), mood (e.g., van Gelderen 2004, Givón 1995, Heycock 2006, Giannakidou and Mari 2015), negative polarity and free choice (e.g., Giannakidou 1998, 2001), predicates of personal taste (e.g., Lasersohn 2009) and acquisition of language and theory of mind (e.g., de Villiers 2005). This paper presents the findings of experiments designed to investigate critical properties of the CC of clause-embedding predicates, with the goal of understanding whether these properties categorize clause-embedding predicates. The remainder of this introduction shows that there is disagreement about which properties define factive predicates (section 1.1) and uncertainty about whether the CCs of particular predicates exhibit the properties attributed to the CCs of factive predicates (section 1.2).

1.1 Disagreement about the definition of factive predicates

One reason for the lack of consensus about which predicates are factive is disagreement about the definition of factive predicates: whereas Kiparsky and Kiparsky 1970 defined factive predicates as those for which the CC is presupposed (see also, e.g., Karttunen 1971a,b), researchers more recently defined factive predicates as those for which the CC is both presupposed and entailed (e.g., Gazdar 1979a:119-123, Chierchia and McConnell-Ginet 1990:355, van der Sandt 1992:345, Abbott 2006:3, Schlenker 2010:139, Anand and Hacquard 2014:77, Spector and Egré 2015:fn.7). For instance, according to Beaver (2001:66f.), Gazdar (1979a:119-123) "describes the inferences associated with factive verbs [...] as being indefeasible in simple affirmative sentences", that is, "entailments". Chierchia and McConnell-Ginet (1990:355) illustrated with *realize* that "[a] sentence can both entail and presuppose another sentence": *Joan realizes that syntax deals with sentence structure* "both entails and presupposes" (*ibid.*) that syntax deals with sentence structure. And Schlenker (2010:139), in discussing CCs, pointed to "the pattern of inference which is characteristic of presuppositions: an entailment of the positive sentence is preserved under negation and in questions".

These two definitions of factive predicates come apart on several predicates, including *take into account, bear in mind* and *make clear*. These predicates were considered factive in Kiparsky and Kiparsky 1970: thus, if Rahim utters one of the affirmative sentences in (2a) or one of the polar questions in (2b), he presupposes the truth of the CC, that the groom is unreliable. But true affirmative sentences with these predicates do not entail the CC, as shown by the acceptability of (2c). Thus, these predicates are factive on a definition of factive predicates according to which the CC is presupposed, but not on a definition of factive predicates according to which the CC is both presupposed and entailed.

(2) a. Rahim: "Joan { took into account / bore in mind / made clear } that the groom is unreliable."

b. Rahim: "Did Joan { take into account / bear in mind / make clear } that the groom is unreliable?"

c. Joan { took into account / bore in mind / made clear } that the groom is unreliable, but she was wrong to do so, because the groom is a very reliable person.

There are more predicates on which the two definitions come apart, as shown in section 1.2, which illustrates disagreements about whether the CC of particular clause-embedding predicates is presupposed and entailed.

1.2 Properties of the content of the complement of clause-embedding predicates

Projection. Under both definitions, the CC of factive predicates is presupposed, with projection taken to diagnose presupposed CCs. It is unclear, however, whether projection leads to a categorical distinction between clause-embedding predicates. As noted above, Kiparsky and Kiparsky (1970) and Karttunen (1971a) assumed that projection results in a three-way categorical distinction of clause-embedding predicates: the CC of factive predicates was assumed to be presupposed, that is, to categorically project; that of non-factive predicates was assumed to be not presupposed, that is, to not project; and that of optionally factive predicates was assumed to be sometimes presupposed – optionally factive predicates "can be used with or without presupposition of the truth of their complement sentence" (Karttunen 1971a:340).[2] But Karttunen (1971b) already challenged this classification when he pointed out that there are factive predicates that "lose their factivity" (p.64) in polar questions and the antecedents of conditionals; he dubbed this subclass of factive predicates 'semi-factive'. To illustrate, consider the negated examples in (3a), the polar questions in (3b) and the conditionals in (3c). Karttunen (1971b) suggested that "all the examples in [(3a)] presuppose that John had not told the truth" (p.63) and that a speaker who utters (3b) or (3c) with *regret* presupposes the CC, that is, the CC projects over negation, the polar question, or the antecedent of the conditional. However, he pointed out, a speaker who utters the variants of (3b) and (3c) with *realize* or *discover* does not necessarily presuppose the CC: the predicates "permit both factive and non-factive interpretations" (p.63), that is, the CC may but need not project. Accordingly, *realize* and *discover*, but also *find out, see* and *notice*, were taken to be semi-factive, and predicates like *regret, forget, resent* and *make clear* were taken to be properly factive.

(3) Karttunen 1971b:63f.

    a. John didn't { regret / realize / discover } that he had not told the truth.

    b. Did you { regret / realize / discover } that you had not told the truth?

    c. If I { regret / realize / discover } later that I have not told the truth, I will confess it to everyone.

Over the past 50 years, the assumption that projection categorizes clause-embedding predicates has been further challenged. First, contrary to Karttunen's (1971b) assumption, non-projection of the CC of semi-factive predicates is possible not just in polar questions and conditionals. For instance, the naturally occurring examples in (4a) and (4b) show that the CC of the semi-factive predicates *discover* and *realize* need not project when embedded under negation or a possibility modal, respectively.

(4)   a. [When did you discover that you liked stormwater?]
       I didn't discover that I liked stormwater, I discovered that I liked fly-fishing for trout.[3]

    b. Perhaps she realized that she was unable to write anything better than "Goblin Market," or perhaps her "failure" to surpass herself is explained by her turn away from poetry to children's stories and religious materials.   (Beaver 2010:87)

Furthermore, even the CC of properly factive predicates need not project, that is, are not categorically presupposed to be true by the speaker. Consider *be aware* and *know*, which Karttunen (1971b) took to be properly factive: the CC does not project in the naturally occurring examples in (5) from Beaver 2010.

(5)   a. Mrs London is not aware that there have ever been signs erected to stop use of the route, nor that there ever has been any obstruction to stop use of the route.   (Beaver 2010:83)

---

[2] According to Kiparsky and Kiparsky (1970) and Karttunen (1971a), optionally factive predicates include *anticipate, acknowledge, suspect, report, emphasize, announce, admit, deduce* and *remember*; this last predicate is taken to be a factive predicate in other research (e.g., Simons 2007, Kastner 2015, Abrusán 2016, Karttunen 2016, Aravind and Hackl 2017, Cremers 2018); in Karttunen and Zaenen 2005 and Karttunen 2016, *acknowledge, admit* and *confess* are considered factive; *announce* is considered non-factive in Karttunen and Zaenen 2005.

[3] http://science.unctv.org/content/whats-my-story-water-quality-engineer

b. Perhaps she knows that she really is to be married, and she, too, is now sad at the end of childhood. (Beaver 2010:86)

c. If Susan knows that her eyes are dark brown, then A) She believes her eyes are dark brown B) Her belief that her eyes are dark brown is justified C) This belief is true D) All of the above E) None of the above. (Beaver 2010:84)

These observations suggest, contrary to what was originally assumed in Kiparsky and Kiparsky 1970, that projection does not categorically distinguish factive predicates from optionally factive ones (or from non-factive ones, given that researchers nowadays typically lump optionally factive and non-factive predicates together as non-factives). One possibility is that the CC of factive and optionally factive predicates may project, but that, overall, the CC of factive predicates is more likely to project than that of optionally factive ones; in other words, speakers who utter sentences with factive and optionally factive predicates can presuppose the truth of the CC, but are more likely to presuppose the truth of the CC of factive predicates than of optionally factive ones. Evidence for this position initially comes from Tonhauser, Beaver, and Degen 2018, whose Experiment 1b investigated the projection of the CC of 12 clause-embedding predicates in polar questions, including the factive predicates *reveal, discover, notice* and *be annoyed* and the optionally factive predicates *establish* and *confess*.[4] They found that the CC of all predicates investigated was projective: participants were more likely to take speakers to be committed to the truth of the CC than to the truth of non-projective (main clause) content. They also found that the CC of the factive predicates was more projective than that of the optionally factive predicates. However, there also were significant differences between the factive predicates: the CC of *discover* was less projective than that of *be annoyed, notice* and *know*, and the CC of *reveal* was less projective than all other factive predicates. Given that there are significant differences in projection not just between factive and optionally factive predicates, but also among factive predicates, it is an open question whether projection categorically distinguishes factive predicates from optionally factive and non-factive ones.

Entailment. There is disagreement for many predicates about whether the CC is entailed, with consequences for whether the predicates are factive under the second definition of factive predicates. Schlenker (2010:139), for instance, took the CC of *inform* to be entailed and *inform* to be a factive predicate: he assumed that Mary being pregnant necessarily follows from (6a). Anand and Hacquard (2014:76), however, pointed out that *inform* can be modified by *falsely*, as in the naturally occurring example in (6b), which they took to show that the CC does not necessarily follow from a true affirmative sentence with *inform*; according to them, *inform* is not factive. Similarly, Swanson (2012) took the CC of *establish* to be entailed, but the naturally occurring example in (7) again suggests that the CC does not necessarily follow from affirmative sentences with *establish*.

(6) a. Mary has informed her parents that she's pregnant.

b. Doctors falsely informed my parents that I was dying.[5]

(7) Diplomacy was repudiated, but with the media's help it was falsely established that the 'villain' was refusing to negotiate.[6]

Emotive predicates like *regret, be glad* or *be annoyed* are also often taken to be factive, following Kiparsky and Kiparsky 1970 and Karttunen 1971b. This assumption has been challenged based on data like (8), where the truth of the CC does not follow from the affirmative sentences, that is, the CC is not entailed (and the writer/speaker does not presuppose the truth of the CC).

---

[4]We assume that these predicate may be taken to be optionally factive given works that do not take speakers to presuppose the truth of the CC of these predicates (e.g., Wyse 2010, Swanson 2012, Karttunen 2016).

[5]`https://www.quora.com/Has-a-doctor-ever-falsely-informed-you-that-you-re-dying`

[6]https://books.google.de/books?id=BnU1qPRO-34C, p.66

(8) a. Mary, who was under the illusion that it was Sunday, was glad that she could stay in bed. (Klein 1975, as cited in Gazdar 1979a:122 and Heim 1992:fn37)

b. Falsely believing that he had inflicted a fatal wound, Oedipus regretted killing the stranger on the road to Thebes. (Klein 1975)

c. John wrongly believes that Mary got married, and he is enraged that she is no longer single. (adapted from Egré 2008, based on Schlenker 2003b)

The existence of examples like (8) has led some authors, including Klein (1975), Giannakidou (1998), Schlenker (2003a) and Egré (2008), to conclude that emotive predicates are not factive: a speaker who uses one of these predicates merely presupposes that the subject of the attitude believes the truth of the CC (Heim 1992). Consistent with this position is a proposal by Karttunen (2016) according to which the CC is not entailed but taken to follow from utterances of affirmative sentences as long as there is no indication that the beliefs of the speaker/writer and the subject of the attitude are not aligned. Other authors maintained that emotive predicates are factive despite examples like (8) (e.g., Gazdar 1979a, Abrusán 2011, Anand and Hacquard 2014): they hypothesized that examples like those in (8) are licensed by a kind of free indirect discourse, that is, do not show that the CC of such predicates is not entailed (or presupposed).

There is similar disagreement about the cognitive predicate *know*: the examples in (9) show that the truth of the CC does not necessarily follow from true affirmative sentences with *know*. Abrusán (2011) assumed that these examples do not show that *know* is not factive because "they are understood as reporting a firm belief or feeling of "knowledge" on somebody's part". Abrusán maintained that these examples "are analogous to the free indirect discourse cases" (p.515) discussed above.

(9) a. Everyone knew that stress caused ulcers, before two Australian doctors in the early 80s proved that ulcers are actually caused by bacterial infection. (Hazlett 2010:501)

b. John suffers from paranoia. He falsely believes that the police is spying on him and what is more he knows they are listening to his phone calls. (Abrusán 2011:514)

In short, some clause-embedding predicates may not be factive according to a definition of factive predicates on which the CC is both presupposed and entailed, but they may be according to a definition on which the CC of factive predicates is merely required to be presupposed. There has, to date, not been a systematic investigation of which clause-embedding predicates are such that their CC is entailed.[7]

1.3 The goal of this paper and the investigated clause-embedding predicates

As illustrated, there is no consensus in the literature about which predicates are factive. This lack of consensus is due partly to differences in how factive predicates are defined, as discussed above. We reiterate the two discussed definitions in (10):

(10) Two definitions of factive predicates
A clause-embedding predicate is factive if and only if the content of its clausal complement is

a. presupposed. (e.g., Kiparsky and Kiparsky 1970, Karttunen 1971a,b)

b. presupposed and entailed. (e.g., Gazdar 1979a, Chierchia and McConnell-Ginet 1990, van der Sandt 1992, Abbott 2006, Schlenker 2010, Anand and Hacquard 2014, Spector and Egré 2015)

---

[7]White and Rawlins's 2018 MegaVeridicality dataset and Ross and Pavlick's 2019 VerbVeridicality dataset include entailment and projection ratings, but the authors did not present by-predicate analyses. We compare the MegaVeridicality and VerbVeridicality ratings to our data in sections 2.3 and 3.5. Djärv et al. (2016) set out to investigate whether emotive and cognitive predicates differ in whether the CC is entailed, but the predicates in their stimuli were realized under an entailment-canceling operator, namely polar questions (e.g., *Is Maria aware that Mike is moving back to Chicago?*). Consequently, the CC is not entailed and observed differences do not directly bear on whether the two types of predicates differ in whether the CC is entailed.

The lack of consensus is compounded by uncertainty about whether projection – which is taken to diagnose whether the CC is presupposed – categorically distinguishes predicates, and by disagreement about whether the CC of particular predicates is entailed.

This paper investigates the projection and entailment of CCs of clause-embedding predicates, with the goal of understanding whether these properties categorize into factive and non-factive predicates. To this end, we present the findings of three pairs of experiments designed to investigate the projection and entailment of the CCs of the 20 clause-embedding predicates that are listed in (11) alongside the categories they are typically taken to fall into:[8] factive predicates in (11a), non-factive predicates in (11b) and optionally factive predicates in (11c). Specifically, the 5 factive predicates in (11a) include the cognitive predicate *know*, the change-of-states predicates *discover* and *reveal*, the sensory predicate *see*, and the emotive predicate *be annoyed*. The 6 non-factive predicates in (11b) include 4 'non-veridical non-factive' predicates *pretend, suggest, say* and *think*, whose CC is typically taken to be neither presupposed nor entailed, as well as the 2 'veridical non-factive' predicates *be right* and *demonstrate*, whose CC is typically taken to be entailed but not presupposed.[9] The 9 optionally factive predicates in (11c) include *acknowledge, admit* and *announce*, which Kiparsky and Kiparsky 1970 listed as optionally factive, as well as *confirm, inform, confess, establish, hear* and *prove*.[10]

(11)   20 clause-embedding predicates

    a. factive: *be annoyed, discover, know, reveal, see*

    b. non-factive:

       i. non-veridical non-factive: *pretend, suggest, say, think*

      ii. veridical non-factive: *be right, demonstrate*

    c. optionally factive: *acknowledge, admit, announce, confess, confirm, establish, hear, inform, prove*

Under the definition in (10a), according to which the CC of factive predicates is presupposed, we expect to see a categorical difference in projection between factive predicates on the one hand, and optionally factive and non-factive predicates on the other. Section 2 reports two experiments designed to investigate the projection of the CCs of these 20 clause-embedding predicates, the results of which allow us to assess whether projection alone identifies a class of factive predicates. Under the definition in (10b), the CC of factive predicates is both presupposed and entailed. Section 3 reports four experiments designed to investigate whether the CC of the 20 clause-embedding predicates is entailed. We expect the CC of factive and veridical non-factive predicates to be entailed, in contrast to that of optionally factive and non-veridical non-factive predicates. These results allow us to assess whether projection and entailment jointly identify a class of factive predicates.

---

[8]The experiments, data and R code for generating the figures and analyses of the experiments reported in this paper are available at [Github repository redacted for review]. All experiments were conducted with IRB approval. Exps. 1b, 2b and 3b were preregistered: [link to OSF registration redacted for review].

[9]Anand and Hacquard (2014) assumed that *demonstrate* is veridical, in contrast to Anand et al. ms. This latter work also takes *reveal* to be non-factive, in contrast to, for instance, Egré 2008, Wyse 2010 or Tonhauser et al. 2018.

[10]Different categories have been assumed for these 6 predicates. For *confirm* and *inform*, Anand and Hacquard (2014) took them to be optionally factive, but recall from above that Schlenker (2010) took *inform* to be factive. For *prove*, White and Rawlins (2018) suggested that it is optionally factive, but Anand and Hacquard (2014) took it to be non-veridical and non-factive. For *confess*, Swanson (2012) took it to be factive, Karttunen (2016) only took it to commit the speaker to the subject of the attitude being committed to the CC, and Wyse (2010) listed it under the non-factive predicates. For *establish*, Swanson (2012) took it to be non-factive, but Wyse (2010) listed it under the factive predicates. Finally, we also included *hear* in this class: even though it is often considered a factive sensory predicate (e.g., Beaver 2010, Anand and Hacquard 2014), it has been observed that *hear* also has a non-factive reportative evidential sense (see, e.g., Anderson 1986, Simons 2007), especially when it is combined with complements that describe events that cannot be auditorily perceived, as is the case in our experiments.

2 Experiment 1: Projection

Exps. 1a and 1b were designed to explore the projection of the CC of the 20 clause-embedding predicates in (11). The experiments employed the 'certain that' diagnostic for projection (see also, e.g., Tonhauser 2016, Djärv and Bacovcin 2017, Stevens et al. 2017, Lorson 2018, Tonhauser et al. 2018, Mahler 2019, de Marneffe et al. 2019): on this diagnostic, participants are presented with an utterance in which the content to be investigated occurs in an entailment-canceling environment, like a polar question, as illustrated in (12) for the content of the nominal appositive, that Martha's new car is a BMW.[11]

(12) Patrick asks: *Was Martha's new car, a BMW, expensive?*

To assess projection, that is, whether the speaker presupposes the truth of the content, participants are asked whether the speaker is certain of the content; for instance, in (12), participants are asked whether Patrick is certain that Martha's new car is a BMW. If a listener takes the speaker to be certain of the content, we assume that the content projects, that is, the speaker presupposes its truth; if a listener does not take the speaker to be certain of the content, we assume that the content does not project, that is the speaker does not presuppose its truth.[12]

Following Tonhauser et al. 2018, Exp. 1a implemented the 'certain that' diagnostic with a gradient response scale: participants gave their responses on a slider marked 'no' at one end and 'yes' at the other. We assume that the closer to 'yes' a participant's response is, the more the speaker is certain of the content, that is, the more projective the content is. In other words, we assume that gradient certainty ratings reflect gradience in the speaker's commitment to the truth of the content.[13] As discussed in Tonhauser et al. 2018, a second interpretation of gradient certainty ratings is that they reflect a participant's degree of belief in the speaker's (binary) commitment to the truth of the content. While we adopt the first interpretation in our discussion, we remain agnostic about the interpretation of gradient certainty ratings: either interpretation gives rise to the expectation on the definition of factive predicates in (10a) that certainty ratings for factive predicates are categorically higher than for optionally factive and non-factive ones.

Exp. 1b was identical to Exp. 1a but employed a two-alternative forced choice task where participants categorically responded 'yes' or 'no'.

2.1 Experiment 1a: Gradient projection ratings

2.1.1 Methods

Participants   300 participants with U.S. IP addresses and at least 99% of previously approved HITs were recruited on Amazon's Mechanical Turk platform (ages: 20-71, median: 36; 136 female, 157 male, 2 other, 3 undeclared). They were paid $0.75 for participating.

Materials   Exp. 1a investigated the projection of the CC of the 20 clause-embedding predicates in (11). The clausal complements of the 20 predicates were realized by 20 clauses (provided in Supplement A), for a total of 400 predicate/clause combinations. These 400 predicate/clause combinations were realized as polar questions by combining them with random proper name subjects, as shown in the sample target stimuli in

---

[11]For other diagnostics for projection see, e.g., Smith and Hall 2011, Xue and Onea 2011 and Tonhauser et al. 2013; see also the discussion in Tonhauser et al. 2018.

[12]We assume that projection is an empirical property of utterance content that can be diagnosed with theoretically untrained native speakers. This position may contrast with that of Anand and Hacquard (2014), who suggested that examples in which the speaker is committed to the CC of a non-factive predicate merely give rise to the "illusion of projectivity" (p.76).

[13]Strictly speaking, gradient certainty ratings reflect gradience in the degree to which participants *perceive* the speaker to be committed to the truth of the content. That is, as in any experiment, we are only indirectly measuring the quantity of interest, in this case speaker commitment.

(13), which realize the complement clause *Julian dances salsa*. 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 speaker of the target stimuli was realized by a proper name and displayed in bold-face. The proper names that realized the speakers, the subjects of the clause-embedding predicates and the subjects of the complement clauses were all unique.

(13)   a. **Carol asks:** Did Sandra discover that Julian dances salsa?
        b. **Paul asks:** Does Edward think that Julian dances salsa?

To assess whether participants were attending to the task, the experiment also included 6 polar question control stimuli. As shown in the sample control in (14), the content whose projection was investigated was the main clause content. For instance, in (14), participants were asked whether the speaker was certain that Zack is coming to the meeting tomorrow. We expected participants to rate the projection of these contents as very low because speakers are typically not committed to content they are asking about. For the full set of control stimuli see Supplement B.

(14)   Is Zack coming to the meeting tomorrow?

Each participant saw a random set of 26 stimuli: each set contained one target stimulus for each of the 20 clause-embedding predicates (each with a unique complement clause) and the same 6 control stimuli. Trial order was randomized.

Procedure   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 CC. 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 1.

Figure 1: A sample trial in Experiment 1a

Data exclusion   The data from 34 participants were excluded according to the criteria given in Supplement C. The data from 266 participants (ages 20-71; median: 36; 118 female, 143 male, 2 other, 3 undeclared) were analyzed.

2.1.2   Results and discussion

Figure 2 shows the mean certainty ratings for the target stimuli by predicate as well as for the main clause stimuli (abbreviated 'MC'), in increasing order from left to right. The mean certainty ratings were largely

consistent with impressionistic judgments reported in the literature. First, the ratings for main clause content were lowest overall, as expected for non-projective content. Second, the ratings for factive predicates were among the highest overall, suggesting comparatively high projection of the CC. Third, the mean certainty ratings of many optionally factive predicates were lower than those of many factive predicates and higher than those of main clauses as well as of non-veridical non-factives. However, Figure 2 also shows that the CCs of the 5 predicates assumed to be factive were 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* were at least as projective as the CCs of *reveal* or *discover*. This suggests that projection alone does not categorically distinguish factive predicates from optionally factive and non-factive ones.
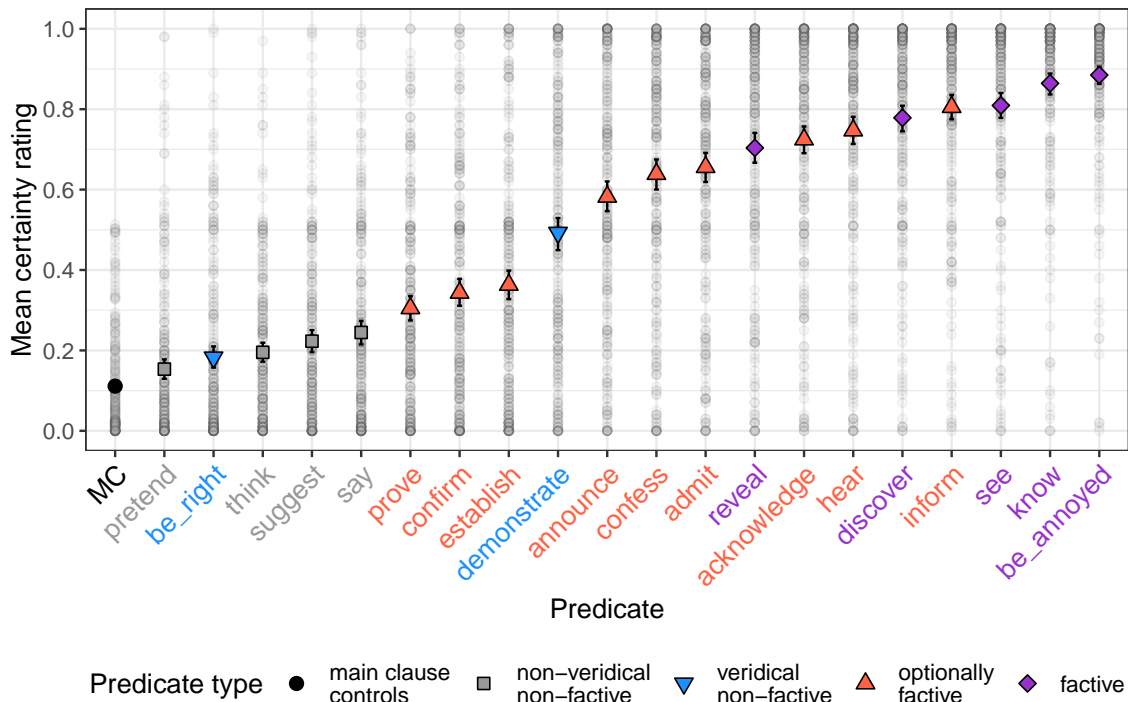

Figure 2: Mean certainty ratings by predicate in Exp. 1a. Error bars indicate 95% bootstrapped confidence intervals. Light gray dots indicate individual participants' responses.

In view of this finding, one might ask whether the categorization of clause-embedding predicates assumed in (11) is correct: perhaps there is a different place in which a line can be drawn between factive and optionally factive predicates based on the projection of the CCs? The mean certainty ratings visualized in Figure 2 suggest that this cannot be done in a non-arbitrary way. For instance, one might consider all predicates factive whose CC is at least as projective as that of *reveal*. This would mean that the predicates typically taken to be factive are factive, as well as *acknowledge, hear* and *inform*. There is, however, no principled reason why the projection of the CC of *reveal* should be the cut-off for factivity. A different approach would be to pick an arbitrary mean certainty rating, say .8, and to categorize predicates whose CC has a mean certainty rating of at least .8 as factive and predicates whose CC has a mean certainty rating below .8 as optionally factive. However, this is as unprincipled an approach as the previous one – one could just as well pick a threshold of .85 or .75, with the result that different predicates would be considered factive. For instance, *discover* (mean: .78) would not count as factive with a threshold of .8, but would with a threshold

of .75, and *see* (mean: .81) would count as factive with a threshold of .8, but not with a threshold of .85.[14]

The possibility of categorizing clause-embedding predicates by the projection of the CC is further called into question by the fact that the CC of all of the 20 predicates investigated is projective, albeit to varying degrees, compared to the non-projective main clause controls. This was established by fitting a Bayesian mixed effects Beta regression model with weakly informative priors using the `brms` (Bürkner 2017) package in R (R Core Team 2016). The model predicted the certainty ratings from a fixed effect of predicate (with treatment coding and 'main clause control' as reference level) and included the maximal random effects structure justified by the design, namely random by-participant and by-item intercepts. This random effects structure captures random variability in projection between participants and between items, where an item is a unique combination of a predicate and a complement clause. A Beta regression model estimates the mean of the outcome distribution (like a linear regression model).[15] We thus obtain, for each predicate, a 95% credible interval for the mean rating for that predicate: this allows us to identify whether the mean ratings of the predicate and of the main clause controls differ. Supplement D 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.

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.[16] 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.

It is possible that asking participants to assess projection with a gradient response scale encouraged them to provide non-categorical answers and that the observed projection variability is an artifact of the task. If so, categorically distinguishing factive predicates from others may be possible if participants were asked to assess projection with a categorical response task. To assess this possibility, we conducted Exp. 1b, which was identical to Exp. 1a but employed a two-alternative forced choice task.

2.2 Experiment 1b: Categorical projection ratings

2.2.1 Methods

Participants 600 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform. They were paid $1.00 for participating.[17]
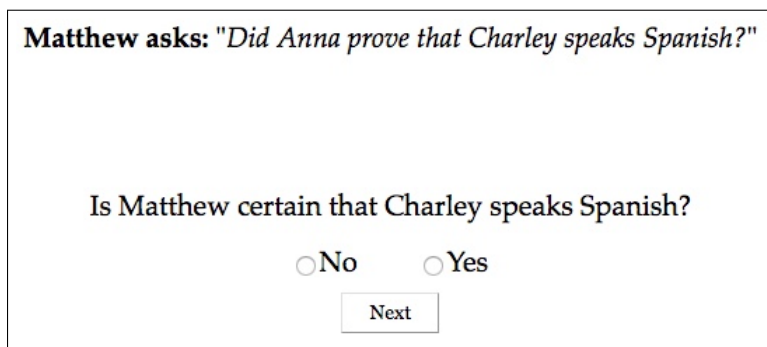
---

[14]Choosing an arbitrary numeric threshold for factivity is also complicated by the fact that there is by-experiment variation in mean certainty ratings. For instance, the mean certainty ratings in Exp. 1a were lower, overall, than the mean certainty ratings in Tonhauser et al.'s (2018) Exps. 1. For example, the CC of *be annoyed*, which was the most projective in both sets of experiments, received mean certainty ratings of .96 and .92 in Tonhauser et al.'s (2018) Exps. 1a and 1b, respectively, but only a .86 in our Exp. 1a. This difference may be due to the fact that Tonhauser et al. (2018) primarily investigated highly projective content whereas our Exp. 1a included a wide range of less projective content: it is possible that participants' certainty ratings were influenced by the overall projection of the contents investigated. We leave this matter to future research.

[15]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 means to identify whether the CCs of the clause-embedding predicates differ from the relevant controls. Both the estimated mean and precision for each predicate are reported in the full model output tables in Supplement D.3.

[16]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 8 for the model code.

[17]Exps. 1a, 2a and 3a (gradient response task) were run in 2017 and 2018, whereas Exps. 1b, 2b and 3b (categorical response task) were run in 2019. Participants in the later experiments were paid more to reflect an increased minimum wage in [the region in which the researchers ran the experiment].

Materials and procedure The materials and procedure were identical to those of Exp. 1a, with the exception of the task: participants responded 'yes' or 'no' to the question of whether the speaker is certain of the relevant content, as shown in Figure 3.

> **Matthew asks:** *"Did Anna prove that Charley speaks Spanish?"*
>
> Is Matthew certain that Charley speaks Spanish?
>
> ○No   ○Yes
>
> Next

Figure 3: A sample trial in Experiment 1b

Data exclusion The data from 164 participants were excluded according to the criteria given in Supplement C. The data from 436 participants (ages 18-81; median: 37; 207 female, 223 male, 2 other) were analyzed.

2.2.2 Results and discussion

Figure 4 shows the proportion of 'yes' ratings (indicating projection) for the predicates and the main clause stimuli, in increasing order from left to right. Jittered gray dots indicate the individual participants' ratings. The order of predicates according to the projection of their CC in Exp. 1b is strikingly similar to that of Exp. 1a, as evidenced by the very high Spearman rank correlation of .983 (see Supplement E for a visualization and details on this coefficient).

   The critical findings of Exp. 1a were replicated. First, the CCs of the factive predicates were not categorically more projective than the CCs of the optionally factive predicates: the CCs of *acknowledge, hear* and *inform* were again at least as projective as the CCs of *reveal* or *discover*. Second, a non-arbitrary line between factive and optionally factive predicates cannot be drawn: the CCs of all 20 predicates were projective compared to the non-projective main clause controls.[18] This was established by fitting a Bayesian mixed effects logistic regression model with weakly informative priors using the `brms` package in R. The model predicted the log odds of a 'yes' over a 'no' response from a fixed effect of predicate (with treatment coding and 'main clause' as reference level) and included the maximal random effects structure justified by the design, namely random by-participant and by-item intercepts. The estimated log odds of a 'yes' response was greater for each predicate than that of the main clause controls, i.e., the 95% credible intervals for the estimated coefficients did not contain 0 for any predicate. See Supplement D.3 for model details. Thus, like Exp. 1a, the results of Exp. 1b do not support the assumption of the definition of factive predicates in (10a) that factive predicates are categorically distinguished from optionally factive and non-factive ones by the projection of the CC.

---

[18]According to Spector and Egré (2015:1739), it is not clear whether the CC of *say* is projective. The findings of our Exps. 1 suggest that it is weakly projective.
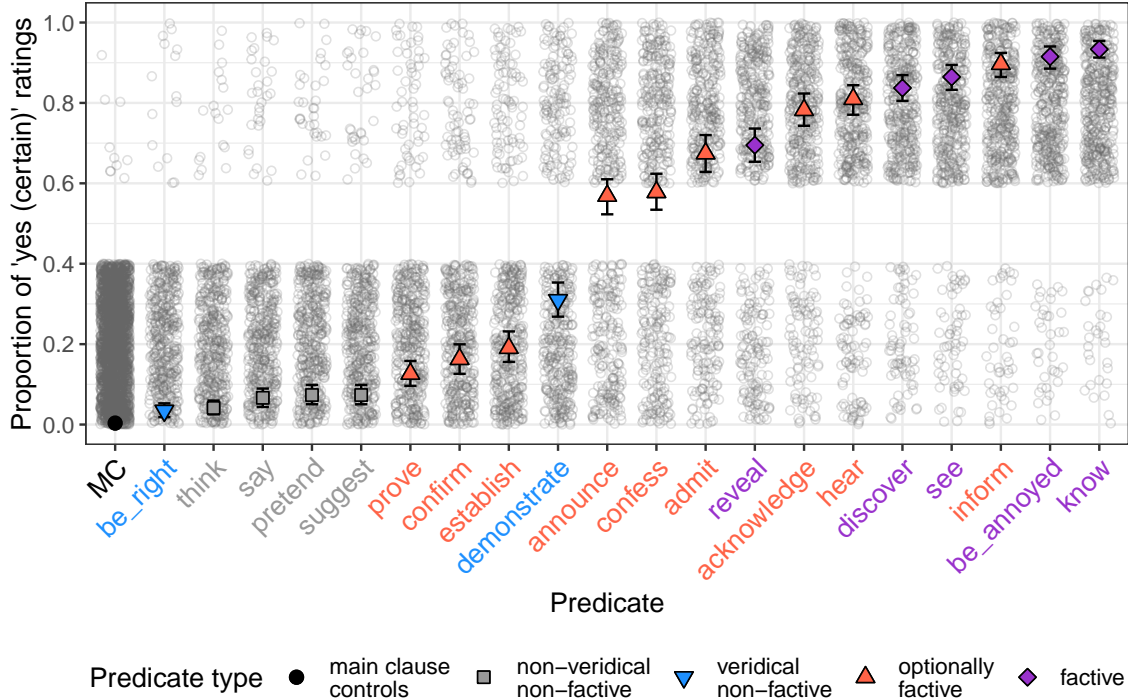
Figure 4: Proportion of 'yes' ratings by predicate in Exp. 1b. Error bars indicate 95% bootstrapped confidence intervals.

2.3 Converging evidence for failure of projection-based definition of factive predicates

The results of Exps. 1a and 1b suggest that the definition of factive predicates in (10a), according to which a predicate is factive if and only if its CC is presupposed, fails to identify a class of factive predicates. Converging evidence comes from three additional datasets: the CommitmentBank (de Marneffe, Simons, and Tonhauser 2019), the MegaVeridicality dataset (White and Rawlins 2018, White et al. 2018) and the VerbVeridicality dataset (Ross and Pavlick 2019).[19]

The CommitmentBank is a collection of 1,200 naturally occurring English discourses from the Wall Street Journal, the British National Corpus and the Switchboard corpus: each discourse consists of a target sentence with a clause-embedding predicate embedded an entailment-canceling operator and up to two preceding context sentences, as illustrated in (15), with *know* in the target sentence embedded under negation:

(15)   What fun to hear Artemis laugh. She's such a serious child. I didn't know she had a sense of humor.
(de Marneffe et al. 2019:109)

Annotators recruited on Amazon's Mechanical Turk platform provided certainty ratings for the CCs of the target sentences on a 7-point Likert scale labeled at three points: speaker/author is certain that the CC is true (coded as 3), speaker/author is not certain whether the CC is true or false (coded as 0), and speaker/author is certain that the CC is false (coded as -3); the lower half of the scale was included to differentiate speaker/author non-commitment to the CC from speaker/author commitment to the falsity of the CC (as may arise in, for instance, *I don't know that I agree.*). de Marneffe et al. (2019) assumed, for

---

[19]We re-plotted the data, obtained at `https://github.com/mcdm/CommitmentBank`, `http://megaattitude.io` and `https://github.com/alexisjihyeross/verb_veridicality`, respectively. For the analysis files, see the GitHub repository mentioned in footnote 8.

certainty ratings above 0, that the higher the certainty rating, the more projective the CC. Figure 5 shows the mean certainty ratings for the CCs in the 982 discourses with 45 non-veridical non-factive, optionally factive and factive clause-embedding predicates (target sentences embedded under non-epistemic modals were excluded; see de Marneffe et al. 2019:§3). In line with the results of Exps. 1, projection of the CC does not categorically distinguish factive predicates from optionally factive and non-factive ones: for instance, the CCs of the optionally factive predicates *foresee, confess, show, tell* and *accept* are at least as projective as those of some factive predicates.
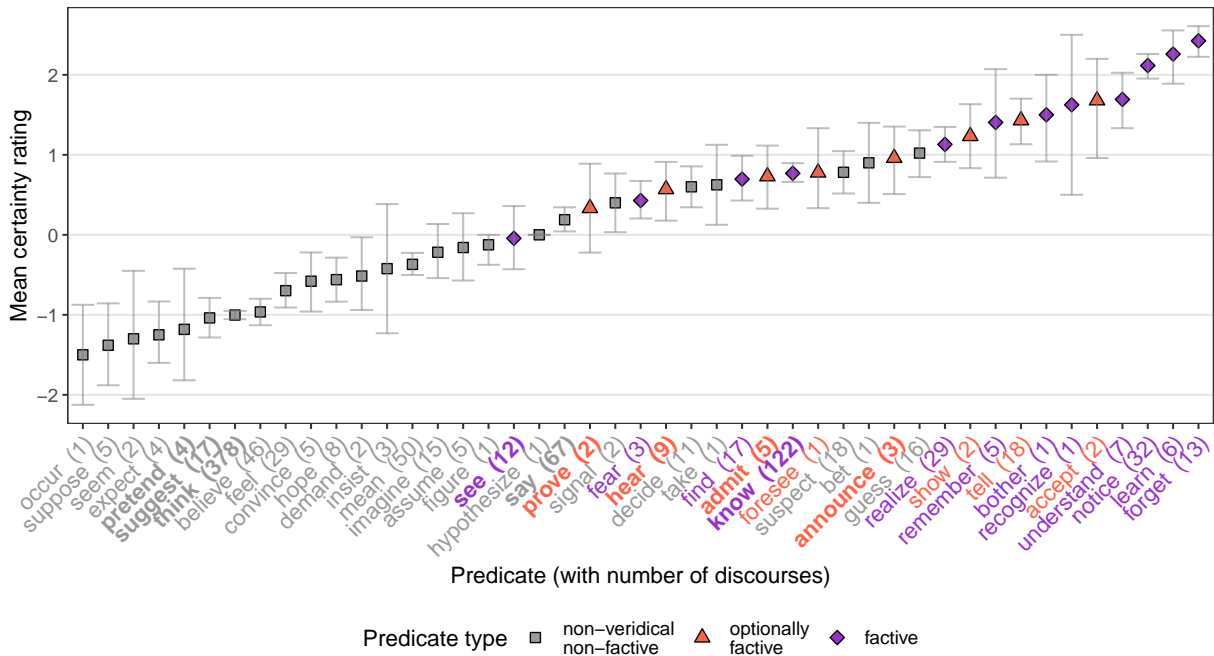

Figure 5: Mean certainty ratings by predicate, with number of discourses in parentheses, for 982 discourses in the CommitmentBank (de Marneffe, Simons, and Tonhauser 2019). Error bars indicate 95% bootstrapped confidence intervals. Predicates included in our Exps. 1 shown in bold.

A second piece of converging evidence that the definition in (10a) fails to identify a class of factive predicates comes from the VerbVeridicality dataset. This dataset includes 859 naturally occurring indicative matrix sentences with a total of 78 clause-embedding predicates and their clausal complements from the MultiNLI (Williams et al. 2018) corpus. To diagnose projection of the CC, negated variants of the 859 matrix sentences were created; an example is given in (16). For each negated variant, three participants rated on a 5-point Likert scale whether the CC is true given the target sentence; one endpoint of the scale was labeled 'definitely true' (coded as 2) and the other 'definitely not true' (coded as -2).

(16)   The GAO has not indicated that it is unwilling to compromise.

Figure 16 plots the mean projection ratings for the CCs of the 78 predicates in the VerbVeridicality dataset, with labels for the 15 predicates from our Exps. 1 that are included in the MegaVeridicality dataset (*see* and *saw*, which are both included in the VerbVeridicality dataset, are plotted separately). As shown, projection of the CC does not categorically distinguish factive predicates from optionally factive and non-factive ones: the CCs of all of our optionally factive predicates are at least as projective as that of some factive predicate.
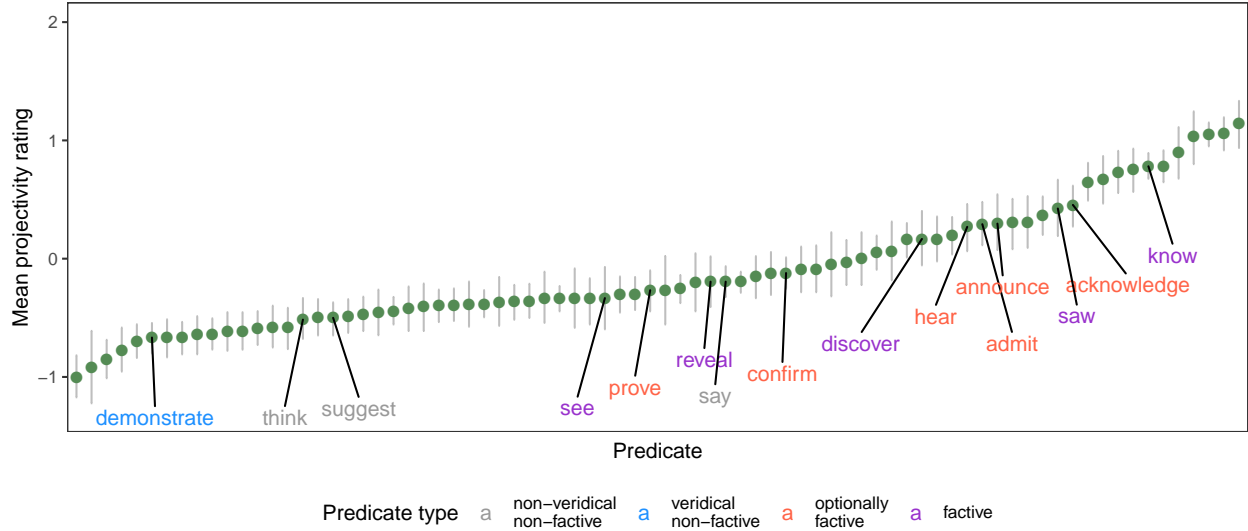
Figure 6: Mean projectivity rating by predicate in green, with 95% bootstrapped confidence intervals, for the 78 predicates in the VerbVeridicality dataset (Ross and Pavlick 2019), with labels for the 15 predicates featured in our experiments.

A third piece of converging evidence comes from the MegaVeridicality dataset, which contains projection ratings for the CCs of 517 English clause-embedding predicates. The stimuli that participants rated consisted of combinations of these predicates with (what is referred to in White and Rawlins 2018 as) low content arguments, as shown in (17) for *know*: the predicates and their clausal complements were embedded under negation in stimuli like (17a) and under negation, in the antecedent of a conditional and under the question operator in stimuli like (17b). To assess projection, participants were asked to respond to the question *Did that thing happen?* for stimuli like (17a) and to respond to the question posed by stimuli like (17b). The response options were 'yes', 'maybe or maybe not' and 'no'.

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

The CCs of each of the 517 predicates in the MegaVeridicality dataset received between 29 and 60 projection ratings (mean: 32). To plot the ratings, we coded a 'yes' response as 1 (the speaker is certain that the thing happened), a 'maybe or maybe not' response as 0 (the speaker is not certain whether the thing happened) and a 'no' response as -1 (the speaker is certain that the thing didn't happen). Figure 7 plots the mean projection ratings for the 517 predicates in the MegaVeridicality dataset, with labels for the 19 predicates from our experiments that are included in the MegaVeridicality dataset (*be right* is not included). As shown, projection of the CC does not categorically distinguish factive predicates from optionally factive and non-factive ones: for instance, the CCs of the optionally factive predicates *confess, acknowledge, admit, announce, hear* and *inform* are at least as projective as those of some factive predicates.
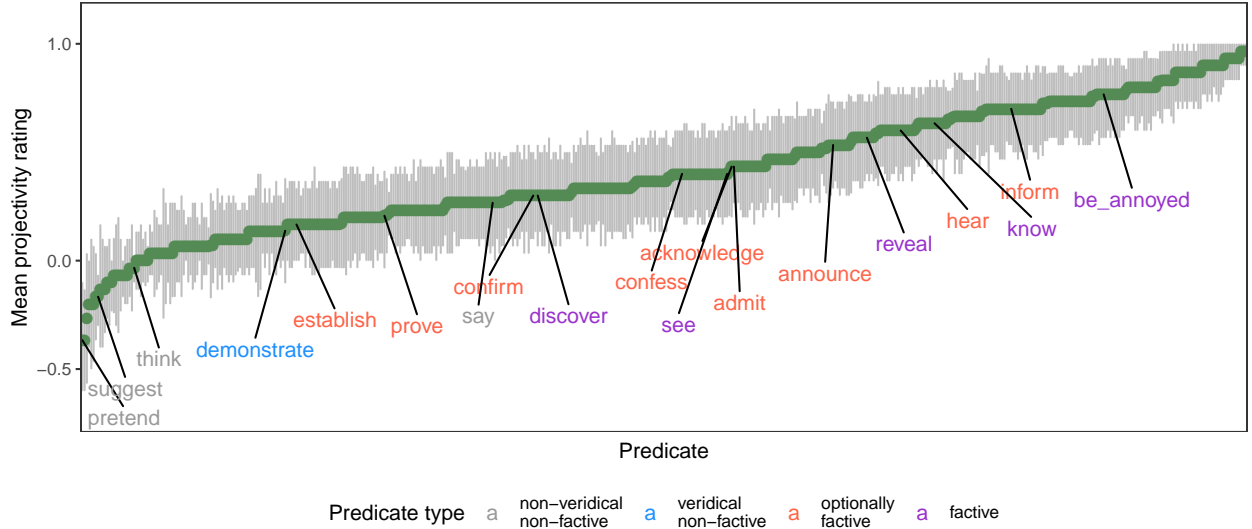
Figure 7: Mean projection rating by predicate in green, with 95% bootstrapped confidence intervals, for the 517 predicates in the MegaVeridicality dataset (White and Rawlins 2018, White et al. 2018), with labels for the 19 predicates featured in our experiments.

In sum, contrary to what is expected under the definition of factive predicates in (10a), the findings of our Exps. 1a and 1b as well as those of the CommitmentBank, the VerbVeridicality dataset, and the MegaVeridicality dataset suggest that projection of the CC does not categorically distinguish factive predicates from optionally factive and non-factive ones. The challenge to the definition in (10a) is compounded by the diversity of the empirical evidence: the evidence comes from constructed and naturally occurring examples, from examples presented to participants/annotators without a context to ones presented in context, from examples with diverse and minimal lexical content, from clause-embedding predicates embedded under a variety of entailment-canceling operators, and from projection ratings collected through different response tasks. We therefore conclude that the definition of factive predicates in (10a), according to which a predicate is factive if and only if the content of the complement is presupposed, fails to identify a class of factive predicates.

The experiments described in the next section were designed to explore whether the definition of factive predicates in (10b), according to which entailment and projection jointly categorize clause-embedding predicates, is empirically supported. To this end, we collected entailment judgments for CCs of the same 20 predicates studied in Exps. 1.

3 Experiments 2 and 3: Entailment

Experiments 2 and 3 explored which of the CCs of the 20 clause-embedding predicates are entailed, to identify which predicates are factive under the definition of factive predicates in (10b). We assume the standard definition of entailment, according to which entailment is a binary, categorical relation between two contents: given sentences ϕ and ψ, the content of ϕ entails the content of ψ if and only if every world in which the content of ϕ is true is also a world in which the content of ψ is true. If there is at least one world in which ϕ is true but ϕ is false, then the content of ϕ does not entail the content of ψ.

To establish whether the content of an indicative matrix sentence with a clause-embedding predicate entails the CC, we used two standard diagnostics for entailment, given in (18). According to the 'inference diagnostic' in (18a), the content of an indicative matrix sentence with a clause-embedding predicate entails

the CC if and only if the CC definitely follows from a true statement of the matrix sentence. According to the 'contradictoriness diagnostic' in (18b), an entailment relationship holds iff an utterance of the indicative matrix sentence that is followed by a denial of the CC is contradictory.

(18) Two diagnostics for entailment   (see, e.g., Chierchia and McConnell-Ginet 1990:§3.1)

  a. Inference diagnostic (Exps. 2a and 2b)
     ϕ entails ψ if and only if, if ϕ is true, then the truth of ψ definitely follows.

  b. Contradictoriness diagnostic (Exps. 3a and 3b)
     ϕ entails ψ if and only if sentences of the form ϕ *and not* ψ are contradictory.

These two diagnostics for entailment were applied to the CCs of the 20 clause-embedding predicates in Exps. 2a/2b and Exps. 3a/3b, respectively: the two versions of each experiment differed in whether participants provided a gradient response (a) or a categorical one (b). Both sets of experiments included control stimuli for which the relevant contents stand in an entailment relation, that is, for which the relevant inference definitely follows (Exps. 2) and that are definitely contradictory (Exps. 3). We assume that the CC of a clause-embedding predicate is not entailed if responses to the CC are significantly different from responses to these control stimuli and that the CC of a clause-embedding predicate is entailed if responses to the CC are not significantly different from responses to these control stimuli. Thus, identifying entailed CCs requires arguing from a null result: given the standard classification of the 20 clause-embedding predicates in (11), we expect the CCs of factive and veridical non-factive predicates to be entailed, i.e., to not be judged differently from the control stimuli.

3.1 Exp. 2a: Gradient ratings using the inference diagnostic for entailment

Exp. 2a investigated which of the CCs of the 20 clause-embedding predicates are entailed based on the inference diagnostic for entailment in (18a). Participants assessed on a gradient scale whether the CC follows from true statements of indicative matrix sentences with the 20 clause-embedding predicates.

3.1.1 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: 19-69, median: 36; 152 female, 148 male). They were paid $0.75 for participating.

Materials   The target stimuli were 400 indicative matrix sentences. They were constructed by combining the 20 clause-embedding predicates with the 20 complement clauses (for 400 predicate/clause combinations, as in Exps. 1) and a proper name subject whose gender differed from the gender of the proper name subject of the embedded clause. As illustrated in the sample stimuli in (19), these matrix sentences were presented to participants as true statements. For each of the 400 target stimuli, the inference that was tested was the inference to the CC: for instance, participants were asked about (19a) whether it follows that Danny ate the last cupcake and about (19b) whether it follows that Emma studied on Saturday morning.

(19) a. **What is true:** Melissa knows that Danny ate the last cupcake.

  b. **What is true:** Jerry pretended that Emma studied on Saturday morning.

The experiment included eight control stimuli that were used to assess whether participants were attending to the task and interpreted the task correctly, as well as to identify inference ratings for entailed content. There were four entailing control stimuli for which the tested inferences are entailments of the true statements: in the sample entailing control in (20), for instance, that Frederick solved the problem is an

entailment of the true statement that Frederick managed to solve the problem. We expected the inference ratings for these four control stimuli to be at ceiling and, as discussed above, we rely on the ratings for these stimuli to identify entailments. In contrast, the inferences tested with the four non-entailing control stimuli, like (21), definitely do not follow from these true statements: for instance, in (21), that Dana wears a wig does not follow from Dana watching a movie last night. We expected the inference ratings for these four control stimuli to be at floor. See Supplement B for the full set of control stimuli.

(20) **What is true:** Frederick managed to solve the problem. (Tested inference: Frederick solved the problem.)

(21) **What is true:** Dana watched a movie last night. (Tested inference: Dana wears a wig.)

Each participant saw a random set of 28 stimuli: each set contained one target stimulus for each of the 20 predicates (each with a unique complement clause) and the same 8 control stimuli. Trial order was randomized.

Procedure   Participants were told that they would read true statements and that they would be asked to assess whether a second statement follows from that true statement. On each trial, participants read the true statement and were asked to respond to the question *Does it follow that...?*, with the complement clause realized by the complement clause of the clause-embedding predicate. Participants responded on a sliding scale from 'definitely doesn't follow' (coded as 0) to 'definitely follows' (coded as 1), as shown in Figure 8.

Figure 8: A sample trial in Experiment 2a

To familiarize participants with the task, they first responded to the two familiarization stimuli in (22), where the inference tested was the CC. Participants who rated (22a) in the lower half of the sliding scale or (22b) in the upper half of the sliding scale were given an explanation for why their answer was wrong. Participants could only advance to the 28 stimuli if they gave a plausible rating to the two familiarization stimuli, that is, a rating in the upper half of the scale for (22a) and in the lower half for (22b).

(22)   a. **What is true:** Drew is correct that Patty lives in Canada.

        b. **What is true**: Drew believes that Patty lives in Canada.

Data exclusion   The data from 41 participants were excluded according to the criteria given in Supplement C. The data from 259 participants (ages 19-69; median: 36; 132 female, 128 male) were analyzed.

3.1.2 Results

Figure 9 shows the mean inference ratings for the target stimuli by predicate, in increasing order from left to right, as well as for the non-entailing and entailing controls. In line with impressionistic judgments reported in the literature, the mean inference ratings for most factive predicates as well as the veridical non-factive predicate *be right* were quite high, suggesting strong inferences to the truth of the CC. The mean inference ratings of the purportedly factive predicate *reveal* and of the purportedly veridical non-factive predicate *demonstrate* were lower, suggesting comparatively weaker inferences to the truth of the CC.
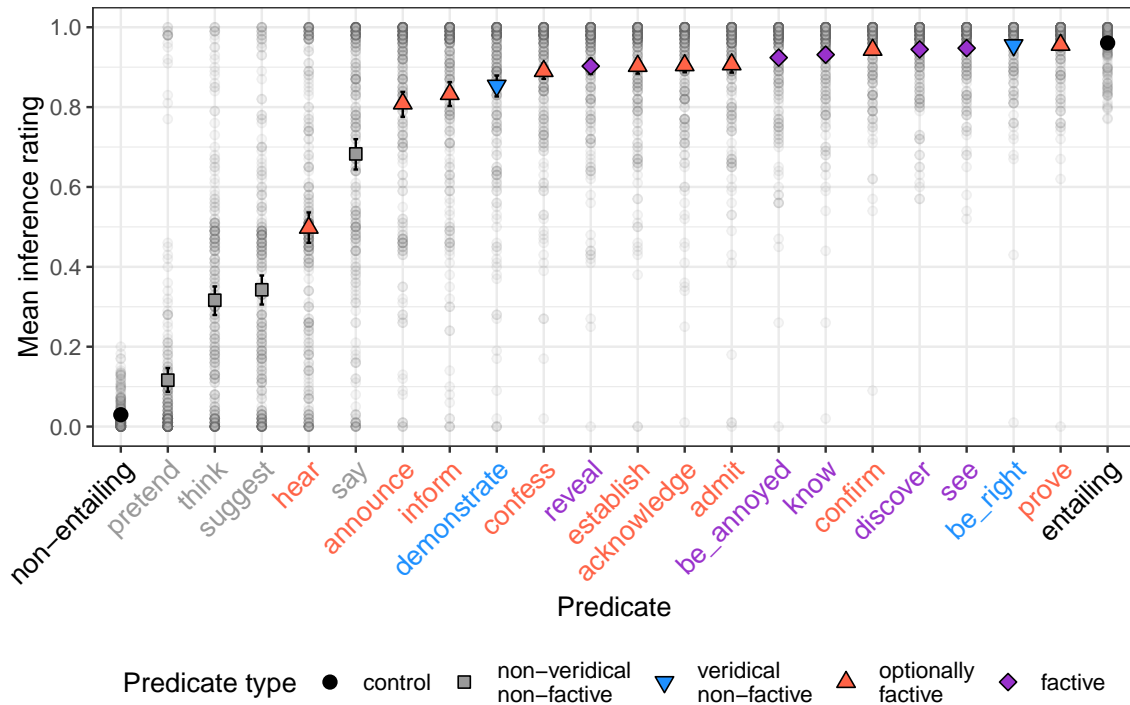

Figure 9: Mean inference rating by predicate in Exp. 2a, including the non-entailing and entailing controls. Error bars indicate bootstrapped 95% confidence intervals. Light gray dots indicate individual participants' ratings.

To assess which predicates have entailed CCs, we fitted a Bayesian mixed effects Beta regression model with weakly informative priors using the `brms` package in R that predicted the inference ratings from a fixed effect of predicate (with treatment coding and 'entailing control' as reference level) and included the maximal random effects structure justified by the design, random by-participant and by-item intercepts. According to this model, the estimated mean for each predicate except *prove* and *be right* was lower than that of the entailing controls, i.e., the 95% credible intervals for the estimated adjustment to the intercept mean did not contain 0 for any predicate other than *prove* and *be right*. See Supplement D.3 for the full model output.[20] Thus, the gradient inference ratings for the CCs of *be right* and *prove* are compatible with the assumption that their CCs are entailed. The inference ratings for the remaining predicates, including the factive predicates *see, discover, know, be annoyed* and *reveal* and the veridical non-factive predicate *demonstrate*, were significantly lower, suggesting that these CCs are not entailed, contrary to assumption.

---

[20]A Bayesian mixed effects linear regression with the same fixed and random effects structure yielded similar results, except that in addition to *prove* and *be right*, there was no evidence for a difference between the entailing controls and *know, confirm, discover* and *see*. See the Github repository mentioned in footnote 8 for the model output.

3.1.3 Discussion

The results of Exp. 2a have two unexpected implications. The first one concerns the predicates for which Exp. 2a found that their CCs are entailed, namely *be right* and *prove*. Given definition (10b), according to which a predicate is factive if and only if its CC is presupposed and entailed, this means that these two predicates are factive if their CCs are presupposed. As discussed above, Exps. 1 found that the CCs of all 20 predicates investigated were at least mildly projective, compared to main clause content. We are therefore led to the conclusion that, according to the definition in (10b), these two predicates are factive. This is a very unsatisfying conclusion indeed, given that the CCs of these two predicates were among the least projective.

The second unexpected implication concerns some of the predicates for which Exp. 2a found that their CCs are not entailed, namely *see, discover, know, be annoyed, reveal* and *demonstrate*. Under the definition of factive predicates in (10b), this result means that, contrary to what is typically assumed, the predicates *see, discover, know, be annoyed* and *reveal* are not factive, nor is the predicate *demonstrate* veridical.

Our way of identifying entailed CCs, namely by comparing the CCs of the 20 predicates to entailed content, was motivated by the standard definition of entailment. One can, of course, ask whether there might be better ways of identifying entailed CCs. For instance, one might consider those CCs entailed whose mean inference rating was at least as high as that of *demonstrate* (.85): this would mean that the CCs of all of the predicates typically taken to be factive or veridical non-factive are entailed, as well as the CCs of some optionally factive predicates. There is, however, no principled reason why the ratings for the CC of *demonstrate* should be the threshold for entailment. One could also pick an arbitrary mean inference rating, say .9, and categorize CCs with mean ratings of at least .9 as entailed, and CCs with mean ratings below .9 as not entailed. Of course, this way of identifying entailed CCs is also not principled. After all, one could just as well pick a threshold of .95 or .85, with the result that the CCs of different predicates would be considered entailed: for instance, the CC of *reveal* (.9) would count as entailed with a threshold of .85 but not with a threshold of .95. Furthermore, neither of these alternative ways of identifying entailed CCs are compatible with the standard definition of entailment because, under both, one would need to allow for CCs to count as entailed even though they received significantly lower ratings than entailed content.

It is also possible, however, that the surprising and theoretically undesirable results of Exp. 2a – that the CCs of *see, discover know, be annoyed, reveal* and *demonstrate* are not entailed – is an artifact of the response task. For instance, the gradient response scale on which participants responded in Exp. 2a may have contributed to lower inference ratings for the CCs of these predicates, resulting in their classification as not entailed. To assess this possibility, we conducted Exp. 2b, which was identical to Exp. 2a but employed a two-alternative forced choice task.

3.2 Exp. 2b: Categorical ratings using the inference diagnostic for entailment

3.2.1 Methods

Participants   600 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform. They were paid $1 for participating.

Materials and procedure   The materials and procedure were identical to those of Exp. 2a, with the exception of the task: participants responded 'yes' or 'no' to the question of whether the CC follows from the true statement, as shown in Figure 10.

Data exclusion   The data from 225 participants were excluded according to the criteria given in Supplement C. The data from 375 participants (ages 18-73; median: 38; 187 female, 187 male, 1 undeclared) were analyzed.

What is true: Brian is annoyed that Zoe calculated the tip.

Does it follow that Zoe calculated the tip?

○No   ○Yes

Next

Figure 10: A sample trial in Experiment 2b

3.2.2 Results and discussion

Figure 11 shows the proportion of 'yes' ratings by predicate, in increasing order from left to right, as well as on non-entailing and entailing controls. The results were strikingly similar to those of Exp. 2a: the proportions of 'yes' responses were quite high for most factive predicates as well as for the veridical non-factive predicate *be right*, but lower for the purportedly factive predicate *reveal* and the purportedly veridical non-factive predicate *demonstrate*. The very high Spearman rank correlation of .989 between Exps. 2a and 2b shows that the ordering of the predicates is almost entirely identical between the gradient and the categorical inference measure (see Supplement E for a visualization).
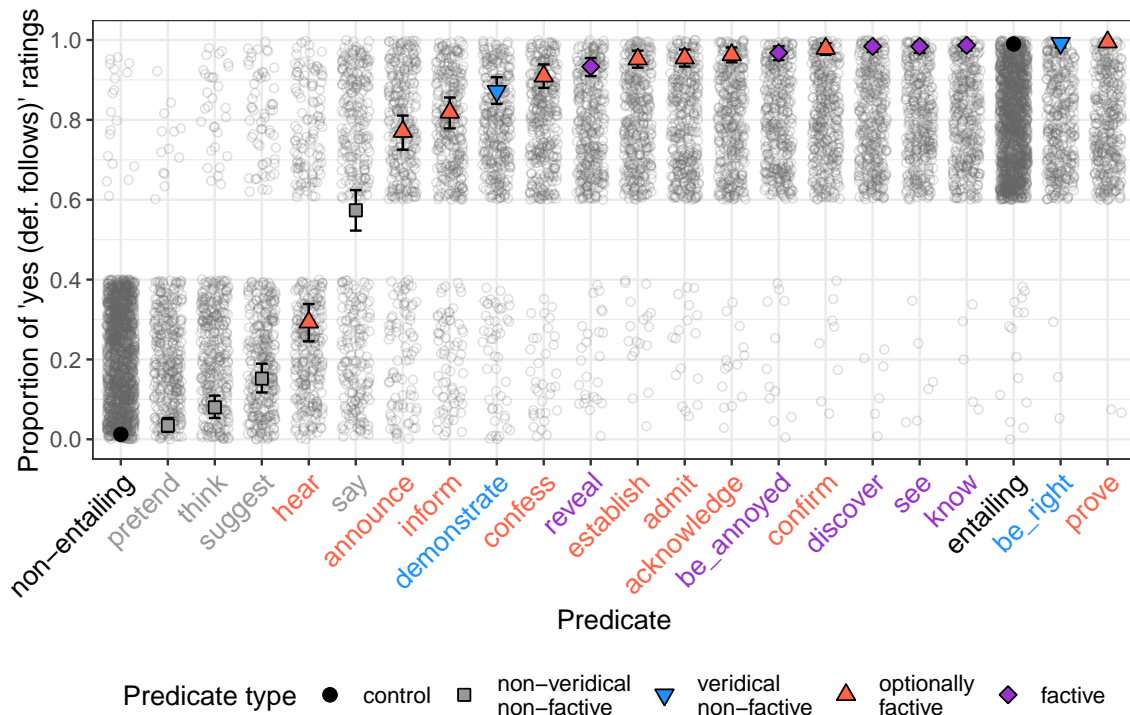

Figure 11: Proportion of 'yes' responses by predicate in Exp. 2b. Error bars indicate 95% bootstrapped confidence intervals. Jittered light gray dots indicate individual participants' responses of 'yes' (coded as 1) and 'no' (coded as 0).

To assess which predicates have entailed CCs, we fitted a Bayesian mixed effects logistic regression model with weakly informative priors using the `brms` package in R that predicted the log odds of a 'yes' over

20

a 'no' response from a fixed effect of predicate (with treatment coding and 'entailing control' as reference level) and included the maximal random effects structure justified by the design, random by-participant and by-item intercepts. The estimated log odds of a 'yes' response was lower for each predicate than for the entailing controls, i.e., the 95% credible intervals for the estimated coefficients did not contain 0 for any predicate, except for *prove, be right, know, see, discover* and *confirm*. Thus, the categorical inference ratings for the CCs of *prove, be right, know, see, discover* and *confirm* are compatible with the assumption that the CCs of these predicates are entailed. The categorical inference ratings for the remaining predicates, including the factive predicates *be annoyed* and *reveal* and the veridical non-factive predicate *demonstrate*, were significantly lower, suggesting that these CCs are not entailed, contrary to assumption.

Changing the response task from a gradient one (Exp. 2a) to a categorical one (Exp. 2b) resulted in slightly different conclusions: whereas Exp. 2a only supports the assumption that the CCs of *be right* and *prove* are entailed, Exp. 2b additionally supports the assumption that the CCs of *know, see, discover* and *confirm* are entailed. However, the implications of Exp. 2b are just as unexpected as those of Exp. 2a. First, because the CCs of *confirm, discover, see, know, be right* and *prove* are entailed and projective, this set of predicates is identified, by the definition of factive predicates in (10b), as factive. This is an unsatisfying conclusion because this set of predicates is very heterogeneous with respect to projection: the CC of *be right* was the least projective of the 20 predicates investigated while that of *know* was the most projective. Thus, this set of predicates should hardly be considered a natural class. Second, the predicates *be annoyed* and *reveal* are surprisingly not factive under the definition of factive predicates in (10b) and the predicate *demonstrate* is surprisingly not veridical.

The findings of both Exps. 2a and 2b have unexpected implications for the identification of factive predicates. One might ask whether these findings and their implications are an artifact of the inference diagnostic for entailment. To assess this possibility, we conducted Exps. 3a and 3b, which were identical to Exps. 2a and 2b, respectively, but employed the contradictoriness diagnostic for entailment.

3.3 Exp. 3a: Gradient ratings using the contradictoriness diagnostic for entailment

Exp. 3a investigated which of the CCs of the 20 clause-embedding predicates are entailed, based on the contradictoriness diagnostic for entailment in (18b). Participants rated the contradictoriness of utterances of English sentences of the form ϕ *but not* ψ, where ϕ is an indicative matrix sentence with a clause-embedding predicate and ψ is its clausal complement, as in (23):

(23)   Mary announced that she is pregnant, but she's not.

Gradient contradictoriness ratings were collected.

3.3.1   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: 18-72, median: 35; 137 female, 162 male, 1 other). They were paid $0.75 for participating.

Materials   On target trials, the 400 predicate/clause combinations were combined with a proper name subject and a *but*-clause that denied the truth of the CC. As shown in (24), stimuli were presented to participants as utterances by named speakers. The proper names that realized the speakers, the subjects of the 20 predicates and the subjects of the complement clauses were all unique. The gender of the proper name subject of the predicate was distinct from the gender of the proper name in the complement clause, to ensure that the pronoun in the elliptical *but*-clause unambiguously referred to the individual referred to in the complement clause of the predicate.

(24)   a. **Christopher:** *"Melissa knows that Danny ate the last cupcake, but he didn't."*
       b. **Susan:** *"Jerry pretended that Emma studied on Saturday morning, but she didn't."*

The experiment included eight control stimuli that were used to assess whether participants were attending to the task, as well as to identify contradictoriness ratings for entailed content. There were four control stimuli, like (25), for which we expected contradictoriness ratings to be at ceiling because the first clause of each of these sentences entails the negation of the second clause. In (25), for instance, the content of *Madison laughed loudly* entails that Madison laughed, that is, the negation of *she didn't laugh*. As discussed above, we relied on the contradictoriness ratings for these control stimuli to identify entailments. By contrast, we expected the contradictoriness ratings for the four non-contradictory control stimuli, like (26), to be at floor, because the content of the second clause does not contradict the content of the first clause. For instance, in (26), the content that Vanessa is good at math does not entail the negation of the content that the speaker is not good at math. For the full set of control stimuli see Supplement B.

(25)   Madison laughed loudly and she didn't laugh.

(26)   Vanessa is really good at math, but I'm not.

Each participant saw a random set of 28 stimuli: each set contained one target stimulus for each of the 20 predicates (each with a unique complement clause) and the same 8 control stimuli. Trial order was randomized.

Procedure.   Participants were told that they would read utterances produced by a speaker and were asked to assess whether the speaker's utterance is contradictory. On each trial, participants read the speaker's utterance and then gave their response on a slider marked 'definitely no' at one end (coded as 0) and 'definitely yes' at the other (coded as 1), as shown in Figure 12. Participants first responded to two familiarization stimuli.

Figure 12: A sample trial in Experiment 3a

Data exclusion   The data from 37 participants were excluded according to the criteria given in Supplement C. The data from 263 participants (ages 18-72; median: 36; 126 female, 136 male, 1 other) were analyzed.

3.3.2 Results and discussion

Figure 13 shows the mean contradictoriness ratings for the target and control stimuli. In line with impressionistic judgments reported in the literature, the mean inference ratings for the factive and veridical non-factive predicates were among the highest, and those of the non-factive non-veridical predicates were among the lowest. However, except possibly for *be right*, none of the mean contradictoriness ratings were close to those of the contradictory controls.
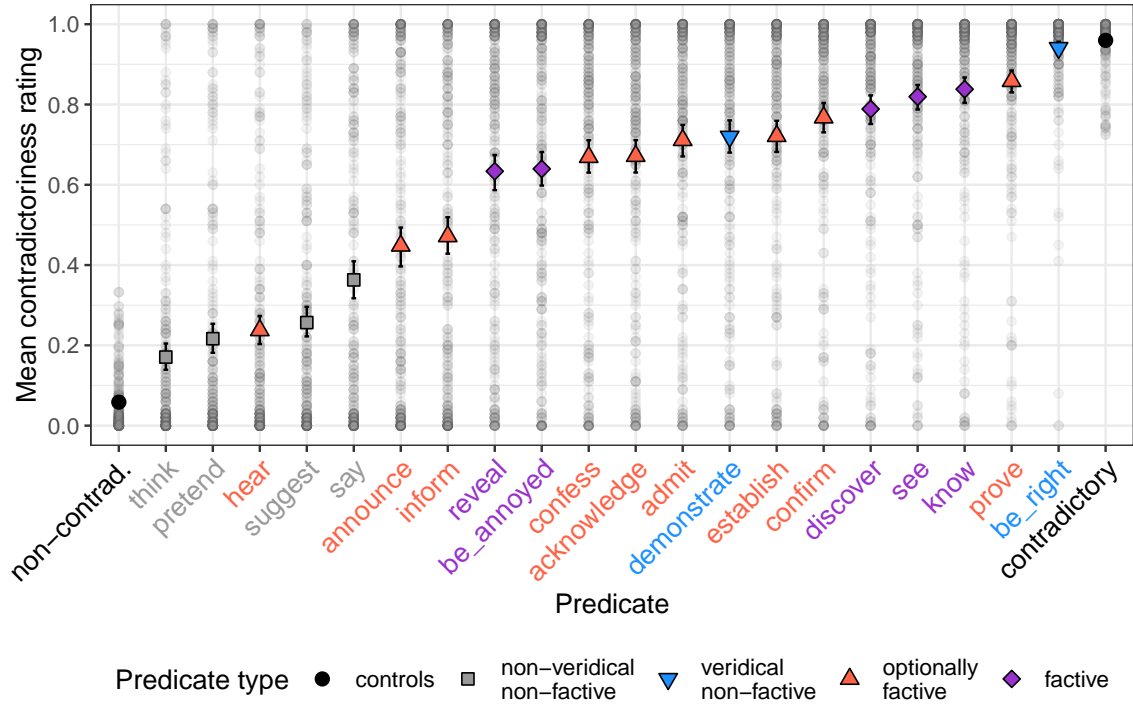

Figure 13: Mean contradictoriness rating by predicate in Exp. 3a, including the non-contradictory and contradictory controls. Error bars indicate bootstrapped 95% confidence intervals. Light gray dots indicate individual participants' ratings.

To assess which predicates have entailed CCs, we fitted a Bayesian mixed effects Beta regression model with weakly informative priors using the `brms` package in R that predicted the contradictoriness ratings from a fixed effect of predicate (with treatment coding and contradictory control as reference level) and included the maximal random effects structure justified by the design, random by-participant and by-item intercepts. According to this model, the estimated mean for all predicates was lower than that of the contradictory controls, i.e., the 95% credible intervals for the estimated adjustment to the intercept mean did not contain 0 for any predicate. See Supplement D.3 for the full model output.[21] Thus, the gradient contradictoriness ratings for none of the predicates are compatible with the assumption that the CCs are entailed and, consequently, none of the 20 predicates are factive under the definition of factive predicates in (10b).

Given the difference in results for Exps. 2a and 2b, one may again ask whether the lack of entailed CCs identified by Exp. 3a is an artifact of the gradient response task, and whether categorical responses might

---

[21]A Bayesian mixed effects linear regression with the same fixed and random effects structure yielded qualitatively identical results, except that the contrast between *be right* and the contradictory controls was not significant. See the Github repository mentioned in footnote 8 for the model output.
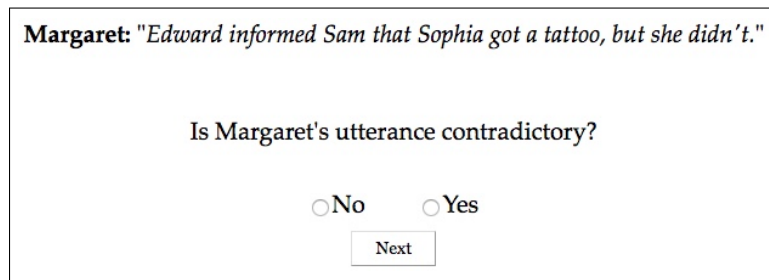
result in more CCs being classified as entailed. To assess this possibility, we conducted Exp. 3b, which was identical to Exp. 3a but employed a two-alternative forced choice task.

3.4 Exp. 3b: Categorical ratings using the contradictoriness diagnostic for entailment

3.4.1 Methods

Participants   600 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform They were paid $1 for participating.

Materials and procedure   The materials and procedure were identical to those of Exp. 3a, with the exception of the task: participants responded 'yes' or 'no' to the question of whether the speaker's utterance is contradictory, as shown in Figure 14.

Figure 14: A sample trial in Experiment 3b

Data exclusion   The data from 247 participants were excluded according to the criteria given in Supplement C. The data from 353 participants (ages 18-73; median: 37; 180 female, 173 male) were analyzed.

3.4.2 Results and discussion

Figure 15 shows the proportion of 'yes' responses on target and control trials. The findings of Exp. 3b were very similar to those of Exp. 3a: except possibly for *be right*, the contradictoriness ratings for the CCs of none of the predicates were close to the contradictory controls. The Spearman rank correlation between Exps. 3a and 3b was very high, at .985 (see Supplement E for a visualization).

   To assess which predicates have entailed CCs, we fitted a Bayesian mixed effects logistic regression model with weakly informative priors using the `brms` package in R that predicted the log odds of a 'yes' over a 'no' response from a fixed effect of predicate (with treatment coding and 'contradictory control' as reference level) and included the maximal random effects structure justified by the design, random by-participant and by-item intercepts. The estimated log odds of a 'yes' response was lower for each predicate than for the contradictory controls, i.e., the 95% credible intervals for the estimated coefficients did not contain 0 for any predicate. Thus, as with Exp. 3a, the results of Exp. 3b suggest that the CCs of none of the 20 predicates are entailed. Consequently, as with Exp. 3a, none of the 20 predicates are factive under the definition of factive predicates in (10b).

3.5 Do projection and entailment jointly identify a class of factive predicates?

According to definition (10a), a predicate is factive if and only if its CC is presupposed. As discussed in section 2, projection alone does not appear to categorically distinguish factive predicates from optionally
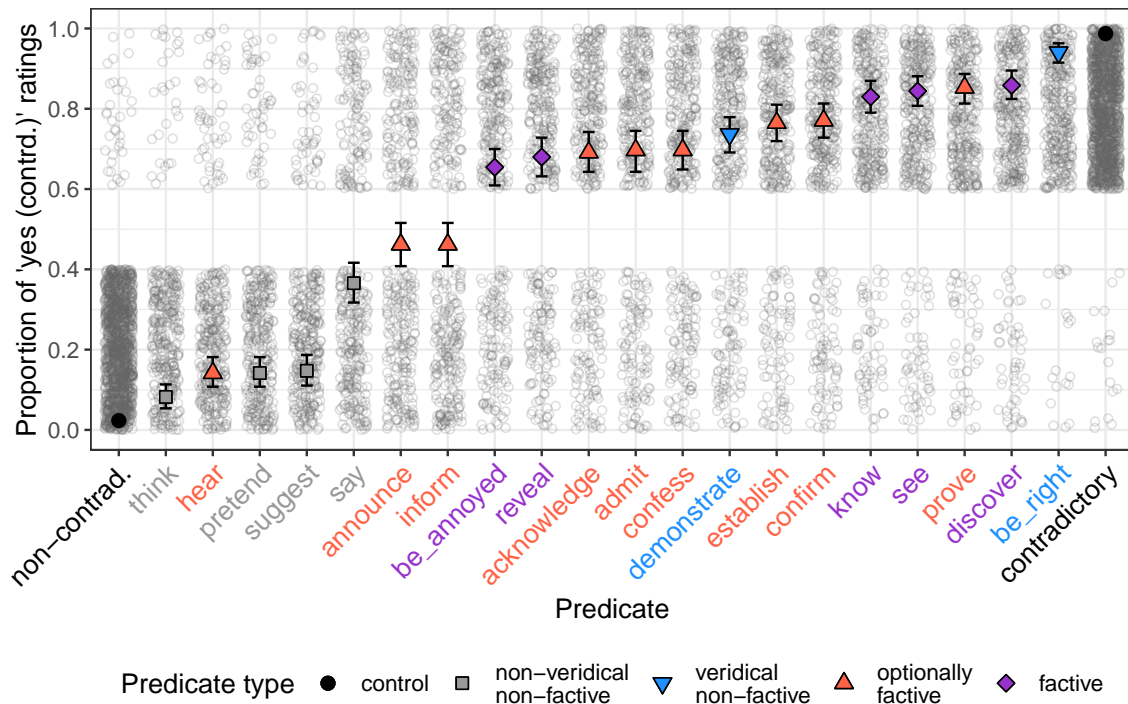
Figure 15: Proportion of 'yes' ratings by predicate in Exp. 3b. Error bars indicate 95% bootstrapped confidence intervals. Jittered light gray dots indicate individual participants' ratings of 'yes' (coded as 1) and 'no' (coded as 0).

factive and non-factive ones and, therefore, fails to identify a class of factive predicates. Definition (10b) differs from definition (10a) in that a predicate is factive if and only if its CC is not just presupposed but also entailed. In Exps. 2 and 3, we applied two standard diagnostics to the CCs of the 20 clause-embedding predicates. To assess whether the CC of a predicate is entailed, we compared the inference ratings (Exps. 2) and the contradictoriness ratings (Exps. 3) for the CCs to the ratings for control stimuli with entailed content. We assumed that the CC of a predicate is entailed if the ratings for that CC are statistically indistinguishable from these control stimuli, and not entailed if the ratings for that CC are significantly lower.

Given these assumptions, we found that the CCs of *prove* and *be right* are entailed based on the gradient and the categorical inference diagnostic (Exps. 2a and 2b), that the CCs of *see, discover* and *confirm* are entailed based on the categorical inference diagnostic (Exp. 2b) but not on the gradient one (Exp. 2a), and that the CCs of none of the 20 predicates are entailed based on the gradient and categorical contradictoriness diagnostics for entailment (Exps. 3). The differences between the findings of Exps. 2 and 3 suggest that the inference and the contradictoriness diagnostic do not both diagnose entailment, contrary to assumption. Given how widely these diagnostics are used, a pressing question for future research is whether one diagnostic is better suited to diagnosing entailment and how to account for the differences.[22]

---

[22]For reasons of space we do not engage with this question in this paper but only make two observations that are relevant to this question. First, the entailed controls received very high ratings across Exps. 2 and 3, suggesting that all four measures are able to diagnose entailed content. de Marneffe et al. (2012:329) suggested that veridicality ratings are influenced by pragmatic reasoning; see also Pavlick and Kwiatkowski 2019. We agree but consider it unlikely that pragmatic reasoning would only influence participants' ratings on the target stimuli, to the exclusion of the control stimuli. Second, even though ratings for the CCs of the 20 predicates were overall lower with the contradictoriness diagnostic than the inference diagnostic, the fact that the Spearman rank correlations of the two diagnostics are very high, namely .954 for the two gradient measures and .934 for the two categorical measures, suggests that both are measuring a similar underlying concept.

Given that the CC of all 20 predicates is at least mildly projective, i.e., presupposed, this means, under the definition in (10b), that the predicates *prove* and *be right* are factive according to the findings of Exp. 2a, that the predicates *prove, be right, know, see, discover* and *confirm* are factive according to the findings of Exp. 2b, and that none of the 20 predicates are factive according to the findings of Exps. 3. The differences between the findings of Exps. 2 and 3 stand in the way of conclusively identifying factive predicates. Moreover, a consequence of the empirical finding that the CCs of more predicates than previously assumed are projective is that relatively more importance falls to entailment in identifying factive predicates. Of all the predicates with projective CCs, those whose CC is entailed are factive and those whose CC is not entailed are not factive. Given that the CC of all 20 predicates we investigated was projective, this has several unfortunate consequences. First, given the results of the gradient inference diagnostic, the predicates *be right* and *prove* are factive, but these both have only weakly projective CCs. Second, given the results of the categorical inference diagnostic, the predicates *prove, be right, see, discover, confirm* and possibly *know* are factive, but this is a heterogeneous set and not a natural class. Finally, given the findings of the contradictoriness diagnostic, none of the 20 predicates is factive. In short, considering entailment in the identification of factive predicates does not have the desired effect: the set of factive predicates is heterogeneous with respect to projection, i.e., not a natural class, or empty. We conclude that the definition of factive predicates in (10b) fails to identify a class of factive predicates.

Converging evidence for this conclusion comes from the VerbVeridicality and MegaVeridicality datasets, which both include entailment ratings in addition to the projection ratings discussed in section 2.3. Consider first the VerbVeridicality dataset: in addition to projection ratings for the CCs of the negated variants of the 859 indicative matrix sentences, it also contains entailment ratings for the 859 indicative matrix sentences; an example is given in (27).

(27) The GAO has indicated that it is unwilling to compromise.

To diagnose entailment, three participants rated on a 5-point Likert scale whether the CC is true given the target sentence; one endpoint of the scale was labeled 'definitely true' (coded as 2) and the other 'definitely not true' (coded as -2). We follow Ross and Pavlick (2019) in referring to these as 'veridicality' ratings.

Figure 16 plots the mean veridicality ratings for the 78 clause-embedding predicates in the VerbVeridicality dataset, with labels for the 15 predicates from our experiments (recall that *see* and *saw* are plotted separately). On the standard definition of entailment, predicates with entailed CCs should have a mean veridicality rating of 2. None of the clause-embedding predicates have a mean veridicality rating of 2: thus, the CCs of none of these 78 clause-embedding predicates are entailed and, consequently, none of the 78 predicates are factive according to the definition in (10b). Relaxing the linking function results in a heterogeneous set of factive predicates: For instance, there are 15 predicates whose CC received a mean veridicality rating at least as high as that of *know* (1.5). If one assumes that these CCs are entailed and that CCs with mean projection ratings of at least .3 are presupposed,[23] then the set of factive predicates is heterogeneous with respect to projection: it includes *notice*, with a mean projection rating of 1.1, *give* (1.1), *realize* (1.1), *explain* (.9) and *know* (.8), but also *acknowledge* (.5), *learn* (.4), *saw* (.4) and *admit* (.3). In short, on the VerbVeridicality set, too, definition (10b) of factive predicates fails to identify a class of factive predicates.

---

[23]The VerbVeridicality dataset does not include projection ratings for non-projective main clause content to which the projection ratings for the CCs could be compared. The cutoff of a mean projection rating of .3 (for ordinal ratings on a 5-point Likert scale coded from -2 to 2) is therefore arbitrary, though plausible: in our Exp. 1a, for instance, the CC of *pretend* was found to be projective compared to the main clause content with a mean certainty rating of .15 (on a scale coded from 0 to 1).
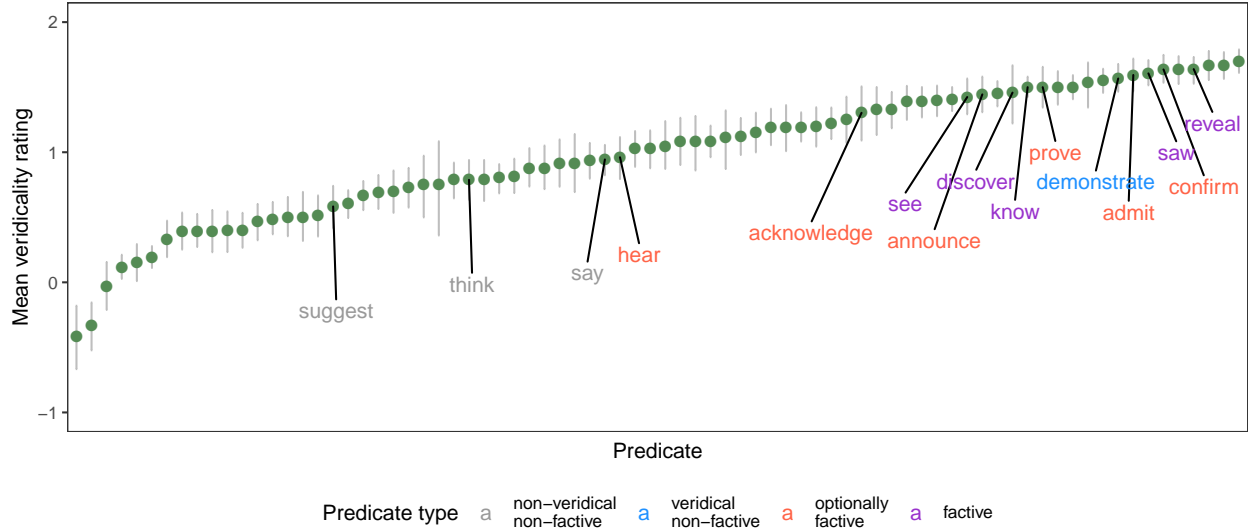
Figure 16: Mean veridicality rating by predicate in green, with 95% bootstrapped confidence intervals, for the 78 predicates in the VerbVeridicality dataset (Ross and Pavlick 2019), with labels for the 15 predicates featured in our experiments.

The MegaVeridicality dataset includes veridicality ratings for the CCs of the same 517 English clause-embedding predicates for which projection ratings were collected. The veridicality ratings were collected for stimuli that combined these predicates with arguments with low lexical content, as shown in (28) for the predicate *know*. To diagnose veridicality, participants were asked to respond to the question *Did that thing happen?*; the response options were 'yes', 'maybe or maybe not' and 'no'.

(28)   Somebody knew that a particular thing happened.

Each of the 517 predicates in the MegaVeridicality dataset received between 9 and 20 entailment ratings (mean: 10) from 159 participants. To plot the ratings, we coded a 'yes' response as 1, a 'maybe or maybe not' response as 0 and a 'no' response as -1. Figure 17 plots the mean veridicality ratings for the 517 predicates in the MegaVeridicality dataset, with labels for the 19 predicates from our experiments. On the standard definition of entailment, one would expect predicates with entailed CCs to have a mean veridicality rating of 1. There are 97 clause-embedding predicates that only received 'yes'/1 ratings, including 3 of the factive predicates we investigated, namely *reveal, know* and *be annoyed*; this finding is compatible with the assumption that the CCs of these 97 predicates are entailed.[24] If we assume, again, that CCs with mean projection ratings of at least .3 are presupposed, then 92 of the 97 predicates are factive,[25] but this set is heterogeneous with respect to projection, including predicates like *bother*, with a mean projection rating of

---

[24]It is not clear that the CCs of these 97 predicates are entailed: this set includes *inform* (see the discussion in section 1) as well as the communication predicates *bitch* and *howl*, whose CCs are not entailed. For instance, it does not follow from the naturally occurring example with *bitch* in (i) that the Democratic party hates all the Jews.

(i)   Rambling to reporters in the Oval Office, Trump bitched that the Democratic party clearly hates all the Jews because of all the support reps Rashida Tlaib and Ilhan Omar received after he strong armed the Israelis into barring them from visiting the country as members of Congress. (`https://www.wonkette.com/08-21-2019`)

White and Rawlins (2018) assumed that the CCs of 376 of the 517 predicates are entailed because they assumed that the CC of a predicate is entailed if and only if the majority of the responses for a predicate was 'yes'. We do not adopt this linking function because it is incompatible with the standard definition of entailment: for instance, given this linking function, the CC of *tweet* is entailed because 11 of the 20 participants responded 'yes', despite the remaining 9 participants responding 'maybe or maybe not'.

[25]White and Rawlins (2018), who assumed the definition of factive predicates in (10b), identified 199 factive predicates in the

.97, *inform* (.7), *know* (.63), *grumble* (.5), *notify* (.4), *gloat* (.4) *convey* (.4) and *confirm* (.3). Thus, on the MegaVeridicality set, too, definition (10b) fails to identify a class of factive predicates.
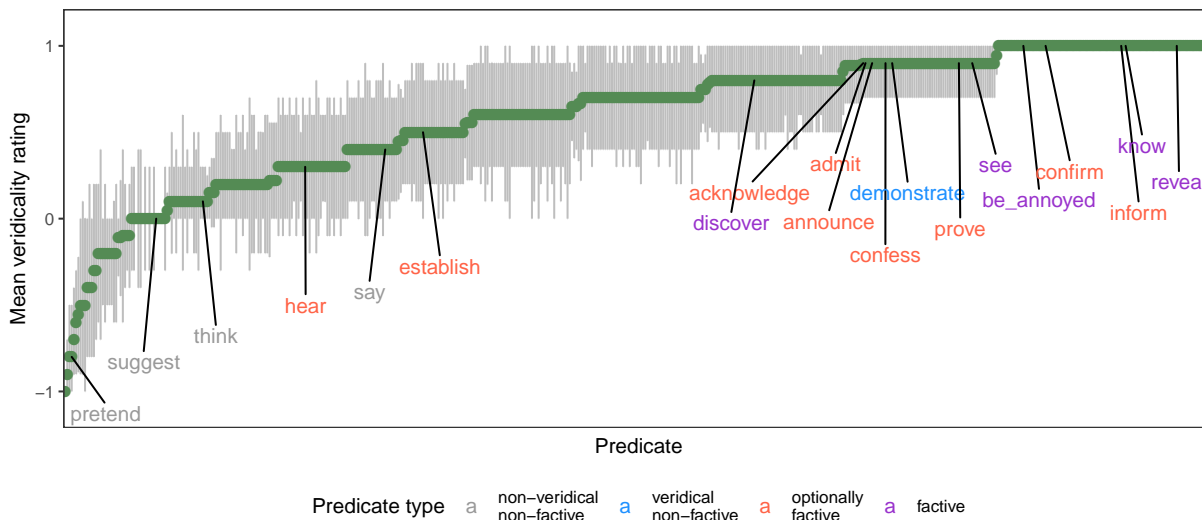

Figure 17: Mean veridicality rating by predicate in green, with 95% bootstrapped confidence intervals, for the 517 predicates in the MegaVeridicality dataset (White and Rawlins 2018, White et al. 2018), with labels for the 19 predicates featured in our experiments.

In sum, contrary to what is expected under the definition of factive predicates in (10b), the findings of our Exps. 1, 2 and 3, as well as those of the VerbVeridicality and MegaVeridicality datasets, suggest that projection and entailment of the CC do not jointly identify a class of factive predicates. Depending on how entailment is diagnosed and which linking function is assumed, the set of predicates identified as factive is either empty (if no predicates have entailed CCs) or heterogeneous (because the CCs of more predicates than previously assumed are at least mildly projective, i.e., presupposed, compared to main clause content).

4 Discussion

Which predicates are factive? Despite the central role that the distinction between factive and non-factive predicates has played in linguistic theorizing, different answers have been given to this question. As discussed in section 1, these different answers are the result of disagreement about how to define factive predicates, uncertainty about whether projection categorically distinguishes factive predicates from optionally factive and non-factive ones, and disagreement about which clause-embedding predicates have entailed CCs. In sections 2 and 3, we reported the results of experiments designed to investigate which predicates are factive based on two common definitions of factive predicates. Our experiments in section 2 investigated the definition in (10a), according to which a predicate is factive if and only if its CC is presupposed. We found that projection alone does not categorically distinguish factive predicates from optionally factive and non-factive ones and, furthermore, that the CCs of all of the predicates we investigated are at least mildly projective. In section 3, we investigated the definition in (10b), according to which a predicate is factive

MegaVeridicality database. However, this finding is due to the authors assuming that a majority of 'yes' responses for a predicate means that the CC is entailed (for positive matrix sentences) or presupposed (for sentences embedded under negation or in a question). As discussed in footnote 24, the linking function for entailed CCs is not compatible with the standard definition of entailment. For presupposed CCs, it is not clear that a majority of 'yes' responses is motivated as a threshold for projection; see the discussion in section 2.1.2.

if and only if its CC is both presupposed and entailed. We found that, depending on how entailment is diagnosed, there are either no factive predicates or that the set of predicates identified as factive is heterogeneous with respect to projection, i.e., not a natural class. We are therefore led to the conclusion that, at present, there is no empirical evidence for a class of factive predicates that is categorically distinguished from optionally factive and non-factive ones by two properties of the CC, namely projection and entailment.

One area in which the distinction between factive and non-factive predicates has played a major role is research on projective content. Formal analyses of the projection of the CC of clause-embedding predicates are generally limited to the CCs of factive predicates. The explanandum is assumed to be that the CCs of factive predicates, but not those of optionally factive or non-factive ones, are presuppositions, which may project. On some analyses (e.g., Heim 1983, van der Sandt 1992), factive predicates lexically specify that the CC is required to be entailed by or satisfied in the relevant context; optionally factive and non-factive predicates do not carry such a lexical specification. On other analyses (e.g., Abrusán 2011, 2016, Romoli 2015, Simons et al. 2017), projection of the CC is derived from the CC being entailed content that is backgrounded; non-entailed CCs of optionally factive and non-factive predicates are again outside of the scope of these analyses. It is within the context of these formal analyses that we were led to investigate which predicates are factive. We wanted to understand which predicates these analyses apply to. Our conclusion that, at present, there is no empirical evidence for a class of factive predicates is unnerving because it means that it is not clear which predicates these analyses apply to and because it challenges long-held and widespread assumptions about clause-embedding predicates.

One way forward is to maintain the assumption that there is a class of factive predicates whose CCs are distinct from those of optionally factive and non-factive predicates and to continue to search for an empirical diagnostic of factive predicates. Another way forward, and the one we advocate for, is to embrace the results reported here: the CCs of many more clause-embedding predicates than previously assumed are projective; there is a lot of variability in how projective the CCs are; and whether the CC is entailed is irrelevant to understanding projection. These findings, we argue, point to several exciting avenues for future research.

First, that the CCs of more clause-embedding predicates than previously assumed are projective means that research on projective content has a much broader empirical scope than previously assumed. Given that research on projective content has focused almost entirely on the CCs of purportedly factive predicates, to the exclusion of the CCs of optionally factive and non-factive ones, our results mean that we are now tasked with investigating and formally analyzing the projection of the CCs of a much broader set of clause-embedding predicates than before. One approach to predicting the projection of the CCs of optionally factive and non-factive predicates has been to assume that such predicates are ambiguous. Spector and Egré (2015), for instance, proposed that predicates like *tell, predict* and *announce* have a factive lexical entry on which the CC is entailed and presupposed and a non-factive one on which it is not. The findings of our Exps. 1 point to some difficulties for this approach. For one, this approach requires the assumption of widespread ambiguity among clause-embedding predicates. It is also not clear which predicates would only have a factive or only a non-factive lexical entry, and which ones would be ambiguous. And, finally, the approach cannot account for the observed by-predicate projection variability. For instance, the assumption that both *discover* and *announce* have a factive and a non-factive lexical entry does not predict that the CC of *discover* should be more projective than the CC of *announce*.

A second avenue for future research emerges from the observation that there is a lot of variability in how projective the CCs of clause-embedding predicates are. Research on projective content has identified a number of factors that influence the projection of utterance content, including context, lexical content, prior content probability, at-issueness, information structure and the perceived degree of reliability and trustworthiness of the attitude holder (e.g., Gazdar 1979a,b, Beaver 2010, Schlenker 2010, Simons et al. 2010, 2017, Abrusán 2011, 2016, Anand and Hacquard 2014, Cummins and Rohde 2015, Djärv and Bacovcin 2017, Tonhauser 2016, in press, Tonhauser et al. 2018, 2019). An exciting question for future research is whether these factors are implicated in the projection of the CCs of all clause-embedding predicates, or only of par-

ticular subsets, and whether they contribute to the observed projection variability between clause-embedding predicates.

One of the factors that clearly plays a role in predicting projection variability is the lexical meaning of the clause embedding predicate. As shown by our Exps. 1, as well as by Tonhauser et al. 2018, a binary distinction between factive and non-factive predicates, or a ternary distinction between factive, optionally factive and non-factive predicates, is too coarse to predict the observed projection variability. We hypothesize that more fine-grained distinctions that are based on the lexical meaning and discourse use of clause-embedding predicates are critical to understanding the projection of the CC. We also hypothesize that the CCs of classes of predicates that share lexical meanings and discourse uses are similar in projection.

We are aware of two tentative proposals along these lines in the recent literature. Anand and Hacquard (2014) hypothesized that listeners may take speakers to be committed to the CC of assertive predicates, like *acknowledge, admit* and *confirm*, "because of the kind of discourse moves that these predicates report" (p.74). Specifically, utterances of sentences with these predicates can be used to report the success of an uptake in a reported common ground. For example, an utterance of a sentence like *Kim confirmed that Mary is the murderer* can report that the content *p* of the complement, that Mary is the murderer, is accepted "into the common ground of the reported discourse" (*ibid.*). The authors suggested that "[t]his acceptance of *p* can easily bleed into the actual common ground, under the assumption that no subsequent move removed *p* from the common ground" (p.74f). Karttunen (2016) proposed different paths to projection for different classes of clause-embedding predicates. For instance, for predicates with *that*-clause subjects that do not involve an attitude holder, like *be odd, be tragic* or *count*, the projection of the CC is derived from its status as a presupposition; with predicates that express a propositional attitude, like *know* or *forget*, the speaker is only necessarily committed to the attitude holder being committed to the truth of the CC and speaker commitment to the CC is derived as a generalized conversational implicature; for a third subclass of change-of-state predicates, including *discover* and *notice*, the CC is entailed but its projection is derived pragmatically; and for a fourth subclass of communicative predicates, like *acknowledge, admit* and *confess*, the speaker is merely committed to the attitude holder having communicated something that the attitude holder wishes to present as a fact. While such proposals need to be fleshed out in more detail to be predictive, we believe that detailed analyses of the lexical meanings of clause-embedding predicates are a fruitful path to a better understanding of the projection of their CCs.

5   Conclusion

Linguistic research since Kiparsky and Kiparsky 1970 has assumed that properties of the content of the clausal complement distinguish factive predicates from non-factive and (what we call) optionally factive ones. This paper investigated whether the content of the complement of 20 English clause-embedding predicates is projective and entailed, with the goal of identifying factive predicates according to two definitions found in the literature. Our experiments revealed that projection alone does not categorically distinguish factive predicates from optionally factive and non-factive ones, and that a class of factive predicates also does not emerge from the additional consideration of entailment, as assessed by two standard diagnostics. We conclude that there is, at present, no empirical support for the assumed categorical distinction between factive predicates on the one hand and optionally factive and non-factive ones on the other. This calls for a reassessment of projection analyses that are limited to the content of the complement of factive predicates, and also opens up several exciting new avenues for empirical and theoretical research on the projection of the contents of the complements of clause-embedding predicates.

References

Abbott, Barbara. 2006. Where have some of the presuppositions gone? In B. J. Birner and G. Ward, eds., *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn*, pages 1–20. Philadelphia: John Benjamins.

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

Abrusán, Márta. 2014. *Weak island semantics*. New York: Oxford University Press.

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

Anand, Pranav, Jane Grimshaw, and Valentine Hacquard. ms. Sentence embedding predicates, factivity and subjects. In C. Condoravdi, ed., *A Festschrift for Lauri Karttunen*. Stanford: CSLI Publications.

Anand, Pranav and Valentine Hacquard. 2014. Factivity, belief and discourse. In *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.

Aravind, Athulya and Martin Hackl. 2017. Factivity and at-issueness in the acquisition of *forget* and *remember*. In *41st Annual Boston University Conference on Language Development*, pages 46–59.

Beaver, David. 2001. *Presupposition and Assertion in Dynamic Semantics*. Stanford, CA: CSLI Publications.

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.

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

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

Cremers, Alexandre. 2018. Plurality effects in an exhaustification-based theory of embedded questions. *Natural Language Semantics* 26:193–251.

Cummins, Chris and Hannah Rohde. 2015. Evoking context with contrastive stress: Effects on pragmatic enrichment. *Frontiers in Psychology* 6:Article 1779.

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

de Marneffe, Marie-Catherine, Christopher D. Manning, and Christopher Potts. 2012. Did it happen? The pragmatic complexity of veridicality assessment. *Computational Linguistics* 38:301–333.

de Villiers, Jill G. 2005. Can language acquisition give children a point of view. In J. Astington and J. Baird, eds., *Why language acquisition matters for theory of mind*, pages 186–219. Oxford University Press.

Djärv, Kajsa and Hezekiah Akiva Bacovcin. 2017. Prosodic effects on factive presupposition projection. In *Semantics and Linguistic Theory (SALT) XXVII*, pages 116–133. Ithaca, NY: CLC Publications.

Djärv, Kajsa, Jérémy Zehr, and Florian Schwarz. 2016. Cognitive vs. emotive factives: An experimental differentiation. In *Sinn und Bedeutung 21*, pages 367–385.

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

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

Gazdar, Gerald. 1979b. A solution to the projection problem. In C.-K. Oh and D. Dineed, eds., *Syntax and Semantics 11: Presupposition*, pages 57–89. New York: Academic Press.

Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non)veridical dependency*. Dordrecht: John Ben-

jamins.

Giannakidou, Anastasia. 2001. The meaning of free choice. *Linguistics and Philosophy* 24:659–735.

Giannakidou, Anastasia and Alda Mari. 2015. Mixed (non)veridicality and mood choice with emotive verbs. In *51st Meeting of the Chicago Linguistic Society*.

Givón, Talmy. 1995. *Functionalism and grammar*. Amsterdam: John Benjamins.

Guerzoni, Elena and Yael Sharvit. 2007. A question of strength: On NPIs in interrogative clauses. *Linguistics & Philosophy* 30:361–391.

Hazlett, Allan. 2010. The myth of factive verbs. *Philosophy and Phenomenological Research* 80:497–522.

Heim, Irene. 1983. On the projection problem for presuppositions. In M. Barlow, D. Flickinger, and M. Westcoat, eds., *West Coast Conference on Formal Linguistics (WCCFL) 2*, pages 114–125.

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

Heycock, Caroline. 2006. Embedded root phenomena. In M. Everaert and H. van Riemsdijk, eds., *The Blackwell Companion to Syntax*, page xxx. Wiley.

Hintikka, Jaakko. 1975. Different constructions in terms of the basic epistemological verbs: A survey of some problems and proposals. In ??, ed., *The intentions of intentionality and other new models for modalities*. Dordrecht: Reidel.

Hukari, Thomas and Robert Levine. 1995. Adjunct extraction. *Natural Language and Linguistic Theory* 31:195–226.

Jaynes, Edwin and O. 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. 1971a. Implicative verbs. *Language* 47:340–358.

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

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

Karttunen, Lauri and Stanley Peters. 1979. Conventional implicature. In C. Oh and D. Dineen, eds., *Presuppositions*, Syntax and Semantics Vol.11, pages 1–56. New York: Academic Press.

Karttunen, Lauri and Annie Zaenen. 2005. Veridicity. In G. Katz, J. Pustejovsky, and F. Schilder, eds., *Annotating, Extracting and Reasoning about Time and Events*, no. 05151 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Kastner, Itamar. 2015. Factivity mirrors interpretation: The selectional requirements of presuppositional verbs. *Lingua* 164:156–188.

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

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

Landau, Idan. 2001. *Elements of control: Structure and meaning in infinitival constructions*. Springer.

Lasersohn, Peter. 2009. Relative truth, speaker commitment, and control of implicit arguments. *Synthese* 166:359–374.

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

Mahler, Taylor. 2019. Does at-issueness predict projection? It's complicated! In *49th Annual Meeting of the North East Linguistic Society*.

Morey, Richard D., Rink Hoekstra, Jeffrey N. Rouder, Michael D. Lee, and Eric Jan 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.

Pavlick, Ellie and Tom Kwiatkowski. 2019. Inherent disagreements in human textual inferences. In *Transactions of the Association for Computational Linguistics*, pages 677–694. Association for Computational Linguistics.

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.

Rooryck, Johan. 2000. *Configurations of sentential complementation*. London: Routledge.

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

Schlenker, Phillipe. 2003a. The lazy Frenchman's approach to the subjunctive. In *Proceedings of Romance Languages and Linguistic Theory*, pages 269–309.

Schlenker, Philippe. 2003b. A plea for monsters. *Linguistics & Philosophy* 26:29–120.

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, David Beaver, Craige Roberts, and Judith Tonhauser. 2017. The Best Question: Explaining the projection behavior of factive verbs. *Discourse Processes* 3:187–206.

Simons, Mandy, Judith Tonhauser, David Beaver, and Craige Roberts. 2010. What projects and why. In *Semantics and Linguistic Theory (SALT) XXI*, pages 309–327. Ithaca, NY: CLC Publications.

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

Smithson, Michael and Jay Verkuilen. 2006. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods* 11(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. In *39th Annual Meeting of the Cognitive Science Society*, pages 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.

Tonhauser, Judith. 2016. Prosodic cues to speaker commitment. In *Semantics and Linguistic Theory (SALT) XXVI*, pages 934–960. Ithaca, NY: CLC Publications.

Tonhauser, Judith. in press. Projection variability in Paraguayan Guaraní. *Natural Language and Linguistic Theory* .

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

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

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

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

van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins.

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 ??

White, Aaron Steven, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2018. Lexicosyntactic inference in neural models. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4717–4724. Brussels, Belgium: Association for Computational Linguistics.

Williams, Adina, Nikita Nangia, and Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1112–1122. Association for Computational Linguistic.

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

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

Zanuttini, Raffaella and Paul Portner. 2003. Exclamative clauses: At the syntax-semantics interface. *Language* 79:39–81.

Supplemental materials for *Which predicates are factive? An empirical investigation*

A 20 complement clauses

The following clauses realized the complements of the predicates in the three experiments:

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

B Control stimuli in Exps. 1, 2 and 3

(1) Control stimuli in Exps. 1

   a. Is Zack coming to the meeting tomorrow?

   b. Is Mary's aunt sick?

   c. Did Todd play football in high school?

   d. Is Vanessa good at math?

   e. Did Madison have a baby?

   f. Was Hendrick's car expensive?

(2) Control stimuli in Exps. 2

   a. Entailing control stimuli

      i. **What is true:** Frederick managed to solve the problem. (Tested inference: Frederick solved the problem.)

      ii. **What is true:** Zack bought himself a car this morning. (Tested inference: Zack owns a car.)

      iii. **What is true:** Tara broke the window with a bat. (Tested inference: The window broke.)

      iv. **What is true:** Vanessa happened to look into the mirror. (Tested inference: Vanessa looked into the mirror.)

   b. Non-entailing control stimuli

      i. **What is true:** Dana watched a movie last night. (Tested inference: Dana wears a wig.)

      ii. **What is true:** Hendrick is renting an apartment. (Tested inference: The apartment has a balcony.)

      iii. **What is true:** Madison was unsuccessful in closing the window. (Tested inference: Madison closed the window.)

      iv. **What is true:** Sebastian failed the exam. (Tested inference: Sebastian did really well on the exam.)

(3)   Control stimuli in Exps. 3

    a. Contradictory control stimuli

       i. Madison laughed loudly and she didn't laugh.

      ii. Dana has never smoked in her life and she stopped smoking recently.

     iii. Hendrick's car is completely red and his car is not red.

     iv. Sebastian lives in the USA and has never been to the USA.

    b. Non-contradictory control stimuli

       i. Vanessa is really good at math, but I'm not.

      ii. Zack believes that I'm married, but I'm actually single.

     iii. Tara wants me to cook for her and I'm a terrific cook.

     iv. Frederick is both smarter and taller than I am.

C   Data exclusion

Table A1 presents how many participants' data were excluded from the analysis 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 four exclusion criteria:

- 'multiple': Due to an experimental glitch, some participants participated in Exps. 1b, 2b or 3b more than once. Of these participants, we only analyzed the data from the first time they participated.
- 'language': Participants' data were excluded if they did not self-identify as native speakers of American English.
- 'controls': Participants' data were excluded if their response mean on the 6 control items was more than 2 sd above the group mean (Exp. 1a), if they gave a wrong rating ('yes') to more than one of the six controls (Exp. 1b), if their response mean on the entailing or the non-entailing controls was more than 2 sd below or above, respectively, the group mean (Exp. 2a), if they gave more than one wrong rating to one of the eight controls, where a wrong rating is a 'yes' to a non-entailing control and a 'no' to an entailing one (Exp. 2b), if their response means on the contradictory or non-contradictory controls were more than 2 sd below or above, respectively, the group mean (Exp. 3a), and if they gave more than one wrong response to one of the eight control sentences, where a wrong response was a 'yes' to a non-contradictory control or a 'no' to a contradictory one (Exp. 3b).
- 'variance': Participants' data were excluded if they always selected roughly the same point on the response scale (variance on response distribution more than 2 sd below the group mean variance).

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

	recruited	Exclusion criteria				remaining
		multiple	language	controls	variance	
Exp. 1a	300	n.a.	13	16	5	266
Exp. 1b	600	75	43	46	n.a.	436
Exp. 2a	300	n.a.	14	27	0	259
Exp. 2b	600	169	35	21	n.a.	375
Exp. 3a	300	n.a.	19	18	0	263
Exp. 3b	600	170	30	47	n.a.	353

Table A1: Data exclusion in Exps. 1, 2 and 3

D.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).

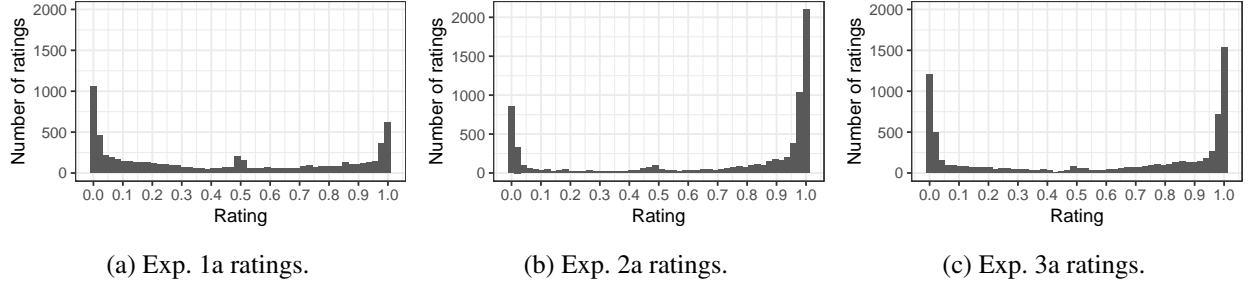| (a) Exp. 1a ratings. | (b) Exp. 2a ratings. | (c) Exp. 3a ratings. |

Figure A1: Histograms of raw slider 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 end-points 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.

D.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.

D.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 A2 for Exps. 1a and 1b, Table A3 for Exps. 2a and 2b, and Table A4 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.

Table A2: Maximum a posteriori (MAP) model estimates for projection ratings from Exp. 1a (left and middle column, mean μ and precision ϕ) and Exp. 1b (right column, β) with 95% credible intervals. Contrast of each predicate is with main clause control reference level. MAP estimates for which there is no evidence that they are different from 0 are italicized.

| Predictor | Exp. 1a: Beta regression | | Exp. 1b: logistic regression |
	Estimated μ	Estimated ϕ	Estimated β
Intercept	−1.88 [−1.98; −1.78]	1.16 [1.02; 1.29]	−6.37 [−7.10; −5.72]
acknowledge	2.62 [2.45; 2.80]	−0.67 [−0.86; −0.49]	8.00 [7.31; 8.78]
admit	2.45 [2.29; 2.61]	−0.65 [−0.83; −0.48]	7.29 [6.60; 8.04]
announce	2.14 [1.98; 2.30]	−0.74 [−0.92; −0.57]	6.70 [6.04; 7.45]
be annoyed	3.56 [3.37; 3.75]	−0.30 [−0.51; −0.10]	9.41 [8.64; 10.22]
be right	0.52 [0.36; 0.68]	−0.21 [−0.40; −0.01]	2.33 [1.50; 3.21]
confess	2.29 [2.13; 2.45]	−0.71 [−0.88; −0.54]	6.75 [6.07; 7.50]
confirm	1.31 [1.15; 1.46]	−0.46 [−0.64; −0.28]	4.27 [3.60; 5.01]
demonstrate	1.78 [1.63; 1.93]	−0.59 [−0.76; −0.41]	5.31 [4.64; 6.06]
discover	2.90 [2.72; 3.07]	−0.67 [−0.85; −0.48]	8.46 [7.75; 9.25]
establish	1.42 [1.26; 1.58]	−0.55 [−0.73; −0.38]	4.51 [3.83; 5.26]
hear	2.71 [2.53; 2.88]	−0.79 [−0.98; −0.61]	8.22 [7.50; 9.01]
inform	3.00 [2.82; 3.18]	−0.60 [−0.79; −0.41]	9.13 [8.38; 9.94]
know	3.37 [3.18; 3.56]	−0.49 [−0.70; −0.29]	9.72 [8.94; 10.57]
pretend	0.32 [0.15; 0.49]	−0.19 [−0.38; −0.00]	3.22 [2.50; 4.03]
prove	1.16 [1.01; 1.31]	−0.21 [−0.40; −0.03]	3.92 [3.21; 4.67]
reveal	2.59 [2.42; 2.76]	−0.72 [−0.90; −0.55]	7.41 [6.72; 8.16]
say	0.78 [0.63; 0.93]	*-0.12* [−0.30; 0.06]	3.10 [2.35; 3.93]
see	3.01 [2.83; 3.19]	−0.70 [−0.90; −0.51]	8.74 [8.03; 9.55]
suggest	0.79 [0.63; 0.94]	*-0.17* [−0.36; 0.02]	3.22 [2.49; 3.99]
think	0.62 [0.47; 0.78]	*0.01* [−0.17; 0.20]	2.54 [1.75; 3.40]

Table A3: Maximum a posteriori (MAP) model estimates for projection ratings from Exp. 2a (left and middle column, mean μ and precision ϕ) and Exp. 2b (right column, β) with 95% credible intervals. Contrast of each predicate is with entailing control reference level. MAP estimates for which there is no evidence that they are different from 0 are italicized. Predicates are marked for which there is no evidence that they differ from entailing controls (bold: no difference on either categorical or continuous measure; italics: no difference on at least one measure).

	Exp. 2a: Beta regression		Exp. 2b: logistic regression
Predictor	Estimated μ	Estimated ϕ	Estimated β
Intercept	3.00 [2.87; 3.15]	2.24 [2.05; 2.42]	6.20 [5.52; 6.98]
acknowledge	−0.74 [−0.94; −0.54]	−0.64 [−0.89; −0.39]	−1.58 [−2.47; −0.70]
admit	−0.67 [−0.87; −0.48]	−0.52 [−0.77; −0.27]	−1.83 [−2.72; −0.97]
announce	−1.51 [−1.71; −1.31]	−1.35 [−1.59; −1.10]	−4.39 [−5.17; −3.68]
be annoyed	−0.55 [−0.76; −0.35]	−0.50 [−0.75; −0.23]	−1.36 [−2.31; −0.46]
be right	*-0.03* [−0.22; 0.16]	*0.09* [−0.16; 0.34]	*0.36* [−0.93; 1.89]
confess	−0.85 [−1.05; −0.66]	−0.64 [−0.90; −0.38]	−2.79 [−3.60; −2.04]
confirm	−0.22 [−0.41; −0.03]	*-0.00* [−0.26; 0.26]	*-0.85* [−1.84; 0.14]
demonstrate	−1.23 [−1.44; −1.02]	−1.18 [−1.44; −0.93]	−3.35 [−4.15; −2.63]
discover	−0.27 [−0.46; −0.08]	*-0.10* [−0.35; 0.14]	*-0.50* [−1.57; 0.64]
establish	−0.78 [−0.98; −0.58]	−0.68 [−0.92; −0.42]	−1.92 [−2.76; −1.12]
hear	−2.92 [−3.12; −2.72]	−2.05 [−2.27; −1.83]	−7.57 [−8.41; −6.84]
inform	−1.37 [−1.57; −1.16]	−1.29 [−1.54; −1.05]	−3.93 [−4.71; −3.22]
know	−0.40 [−0.60; −0.21]	−0.28 [−0.54; −0.03]	*-0.27* [−1.39; 0.96]
nonMent.C	−6.22 [−6.42; −6.02]	−0.27 [−0.52; −0.04]	−11.92 [−12.94; −11.01]
pretend	−4.47 [−4.72; −4.24]	−2.13 [−2.38; −1.87]	−10.78 [−11.83; −9.85]
prove	*-0.08* [−0.29; 0.14]	*-0.13* [−0.40; 0.14]	*0.92* [−0.57; 2.91]
reveal	−0.80 [−1.00; −0.60]	−0.67 [−0.92; −0.41]	−2.37 [−3.21; −1.54]
say	−2.20 [−2.41; −2.00]	−1.92 [−2.15; −1.69]	−5.79 [−6.59; −5.09]
see	−0.23 [−0.43; −0.02]	*-0.16* [−0.41; 0.09]	*-0.49* [−1.51; 0.62]
suggest	−3.47 [−3.67; −3.27]	−1.90 [−2.12; −1.67]	−8.75 [−9.62; −7.95]
think	−3.64 [−3.84; −3.44]	−1.85 [−2.08; −1.62]	−9.71 [−10.65; −8.87]

Table A4: Maximum a posteriori (MAP) model estimates for projection ratings from Exp. 3a (left and middle column, mean μ and precision ϕ) and Exp. 3b (right column, β) with 95% credible intervals. Contrast of each predicate is with contradictory control reference level. MAP estimates for which there is no evidence that they are different from 0 are italicized.

| Predictor | Exp. 3a: Beta regression | | Exp. 3b: logistic regression |
	Estimated μ	Estimated ϕ	Estimated β
Intercept	2.74	1.46	6.17
	[2.54; 2.93]	[1.23; 1.66]	[5.55; 6.88]
acknowledge	−2.15	−1.04	−4.89
	[−2.39; −1.90]	[−1.29; −0.79]	[−5.65; −4.21]
admit	−2.03	−1.05	−4.86
	[−2.26; −1.77]	[−1.30; −0.78]	[−5.59; −4.19]
announce	−2.90	−1.56	−6.46
	[−3.14; −2.64]	[−1.80; −1.32]	[−7.20; −5.80]
be annoyed	−2.24	−1.28	−5.18
	[−2.47; −1.98]	[−1.53; −1.03]	[−5.91; −4.51]
be right	−0.40	*-0.18*	−1.90
	[−0.67; −0.12]	[−0.48; 0.12]	[−2.70; −1.12]
confess	−2.15	−1.12	−4.85
	[−2.39; −1.89]	[−1.37; −0.86]	[−5.59; −4.18]
confirm	−1.74	−0.99	−4.25
	[−1.99; −1.48]	[−1.24; −0.73]	[−4.99; −3.57]
demonstrate	−1.94	−1.05	−4.53
	[−2.18; −1.69]	[−1.30; −0.79]	[−5.25; −3.84]
discover	−1.63	−1.02	−3.30
	[−1.90; −1.38]	[−1.27; −0.75]	[−4.06; −2.61]
establish	−1.94	−1.00	−4.29
	[−2.19; −1.70]	[−1.27; −0.75]	[−5.06; −3.62]
hear	−3.72	−1.29	−8.98
	[−3.97; −3.45]	[−1.54; −1.01]	[−9.79; −8.27]
inform	−2.78	−1.51	−6.46
	[−3.03; −2.52]	[−1.75; −1.25]	[−7.20; −5.81]
know	−1.37	−0.90	−3.64
	[−1.63; −1.11]	[−1.17; −0.64]	[−4.39; −2.95]
non-contra. control	−5.26	*-0.24*	−11.51
	[−5.54; −4.98]	[−0.51; 0.04]	[−12.43; −10.71]
pretend	−3.72	−1.35	−8.97
	[−3.98; −3.46]	[−1.61; −1.10]	[−9.78; −8.26]
prove	−1.18	−0.71	−3.36
	[−1.44; −0.92]	[−0.98; −0.45]	[−4.10; −2.65]
reveal	−2.27	−1.24	−4.99
	[−2.52; −2.02]	[−1.50; −0.98]	[−5.73; −4.32]
say	−3.17	−1.51	−7.11
	[−3.42; −2.91]	[−1.75; −1.25]	[−7.87; −6.42]
see	−1.43	−0.82	−3.48
	[−1.68; −1.18]	[−1.08; −0.56]	[−4.21; −2.78]
suggest	−3.58	−1.17	−8.92
	[−3.83; −3.31]	[−1.43; −0.92]	[−9.71; −8.21]
think	−3.93	−1.20	−9.81
	[−4.19; −3.66]	[−1.47; −0.94]	[−10.67; −9.06]

E   Comparisons of gradient and categorical ratings

The Spearman rank correlation coefficient, a value between -1 and 1, is a nonparametric measure of rank correlation: the higher the coefficient, the more the relation between the the two variables can be described using a monotonic function; if the coefficient is positive, the value of one variable tends to increase with an increase in the other. In the case of our experiments, a coefficient of 1 would mean that there is a perfectly monotone increasing relation between the ranking of the predicates in Exp. 1a and Exp. 1b: for any two predicates p_1 and p_2, if p_1 ranks below p_2 in Exp. 1a (that is, the mean certainty rating of p_1 is lower than that of p_2), then that ranking is preserved in Exp. 1b.

The three panels of Figure A2 visualize the strong correlation between by-predicate mean ratings in Exps. 1a, 2a and 3a and by-predicate proportion of 'yes' responses in Exps. 1b, 2b, and 3b, respectively.
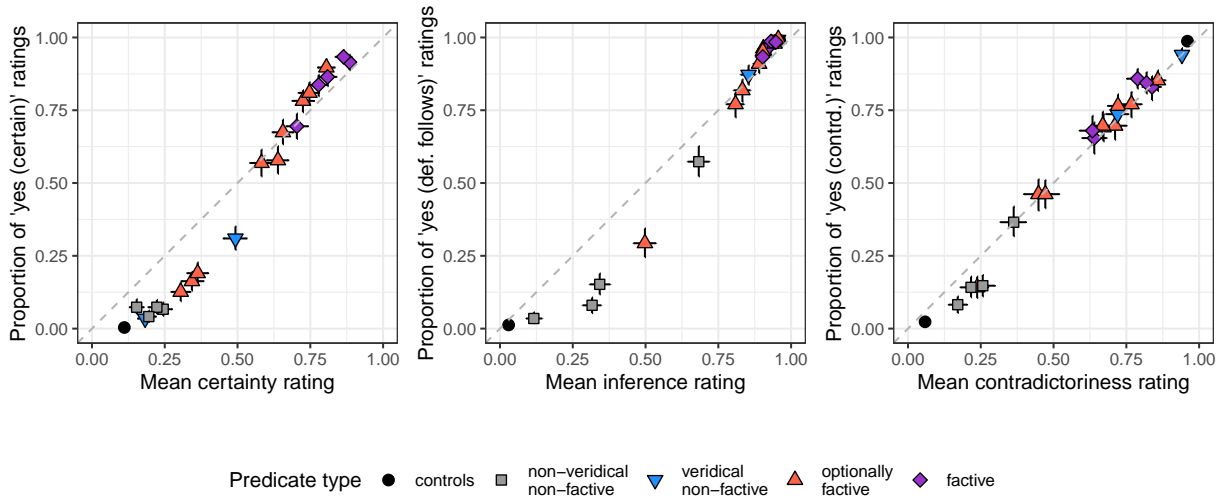

Figure A2: By-predicate proportion of 'yes' responses in two-alternative forced choice task against mean slider ratings in Exps. 1 (left, $r = .98$), Exps. 2 (middle, $r = .99$), and Exps. 3 (right, $r = .99$). Error bars indicate 95% bootstrapped confidence intervals.