Are there factive predicates? An empirical investigation

Author(s)

September 30, 2021

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 content of the complement of particular predicates exhibit the properties attributed to the content of the complement of factive predicates. This has led to a lack of consensus about which predicates are factive, a troublesome situation given the central role that factivity plays in linguistic theorizing. 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 identify a class of factive predicates. We find that factive predicates are more heterogeneous than previously assumed and that there is little empirical support from these experiments for the assumed categorical distinction between factive and non-factive predicates. We discuss implications of our results for formal analyses of presuppositions, one area where factivity has played a central role. We propose that projection is sensitive to more fine-grained meaning distinctions between clause-embedding predicates than factivity. (Word count of main text: 18,130)

1 Introduction

Clause-embedding predicates are predicates like *know* and *think* that take a finite clause as an argument, as shown in (1), where the finite clause *(that) it is raining* realizes the direct object. Kiparsky and Kiparsky 1970 proposed that there is a class of clause-embedding predicates called 'factive' predicates, like *know*, that distinguish themselves from other predicates, like *think*, in that the content of their clausal complement is presupposed, that is, "[t]he speaker presupposes that the embedded clause expresses a true proposition" (p.147). Thus, a speaker who utters (1) with *know* is taken to presuppose that it is raining, in contrast to a speaker who utters (1) with *think*.

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

To diagnose whether a predicate is factive, that is, whether the content of its complement (henceforth CC) is presupposed, Kiparsky and Kiparsky 1970 relied on the projection diagnostic (see also, e.g., Langendoen and Savin 1971). Specifically, they assumed that the CC of a clause-embedding predicate is presupposed if the speaker is committed to the truth of the CC not only when uttering an affirmative matrix sentence like (1) but also when uttering negated or polar question variants of the affirmative matrix sentence, like those in (2a) and (2b), respectively. Thus, *know* was categorized in Kiparsky and Kiparsky 1970 as a factive predicate because a speaker is committed to the CC when uttering the affirmative sentence in (1) with *know* and when uttering its negated variant in (2a) or the polar question variant in (2b). By contrast, *think* was categorized as non-factive because the CC is not presupposed, that is, a speaker who utters the variants in (1) and (2a-b) with *think* is not necessarily committed to the truth of the CC.[1]

---

[1]Factive predicates were also taken to be distinguished by syntactic properties in Kiparsky and Kiparsky 1970. We ignore syntax here because Kiparsky and Kiparsky (1970:fn.3) already pointed out that syntactic and semantic factivity are not perfectly aligned:

(2) a. Sam doesn't { know / think } that it is raining.

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

Today, some 50 years after Kiparsky and Kiparsky 1970, properties of the CC are still taken to categorically distinguish factive predicates from others, in English and 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 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 and island effects (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 language acquisition and theory of mind (e.g., de Villiers 2005). This paper reports six experiments designed to investigate critical properties of the CC of clause-embedding predicates, with the goal of investigating whether these properties identify a coherent class of factive predicates. What we find, in a nutshell, is that factive predicates are more heterogeneous than previously assumed and that there is little empirical support from these experiments for the assumed categorical distinction between factive and non-factive 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 their definition. As mentioned above, Kiparsky and Kiparsky 1970 defined factive predicates as those for which the CC is presupposed (see also, e.g., Karttunen 1971a,b), with projection diagnosing presupposed CCs. Thus, on this definition, a predicate is factive if and only if the speaker is committed to the truth of the CC in affirmative matrix sentences as well as in negated and polar question variants. More recent work, however, assumes that the CC of a factive predicate is not only presupposed but also entailed (e.g., Gazdar 1979a:119-123, Schlenker 2010:139, Abrusán 2011, Anand and Hacquard 2014:71, Spector and Egré 2015:fn.7).[2] In other words, for a predicate to be factive, not only does the speaker have to presuppose the CC, but the truth of the CC also has to be indefeasibly inferred from affirmative matrix sentences; that is, the truth of the CC has to follow for any competent language user, not just the speaker, from a true affirmative matrix sentence in which the CC is embedded under the factive predicate. The two definitions are given in (3):

(3) 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)

---

know and *realize*, for instance, were taken to be syntactically non-factive but semantically factive. For recent discussion see White and Rawlins 2018, Djärv 2019 and references therein.

[2]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". 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". Authors are not always clear about which definition they subscribe to. For instance, Chierchia and McConnell-Ginet (1990:355) used *realize* to illustrate 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. It is not clear, however, whether they assumed this pattern for factive predicates more generally or whether they assumed that a predicate is factive as along as its CC is presupposed.

b. presupposed and entailed.   (e.g., Gazdar 1979a, Schlenker 2010, Abrusán 2011
                                                          Anand and Hacquard 2014, Spector and Egré 2015)

These two definitions of factive predicates come apart on some predicates, including *take into account* and *make clear*, shown in (4) and (5). These predicates were considered factive in Kiparsky and Kiparsky 1970, on definition (3a), because the CC projects, i.e., the speaker is committed to the truth of the CC, that the groom is unreliable, not just when uttering one of the affirmative matrix sentences in (4a) but also when uttering one of the polar question variants in (4b). But true affirmative matrix sentences with these predicates do not entail the CC, as shown by the acceptability of (5). In this example, the truth of the CC is not an indefeasible inference of the affirmative matrix sentence with one of these predicates because the affirmative matrix sentence can be followed up, without contradiction, by *because the groom is a very reliable person*. If the CC was entailed by the affirmative matrix sentence, (5) should be contradictory. Thus, *take into account* and *make clear*, which were taken to be factive on definition (3a), are not factive on definition (3b).

(4)   a. Joan { took into account / made clear } that the groom is unreliable.

       b. Did Joan { take into account / make clear } that the groom is unreliable?

(5)   Joan { took into account / 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

Presupposed.   The two definitions in (3) share the assumption that the CC of factive predicates is presupposed. In addition to being shared by the two definitions, the assumption is also shared by different formal analyses of presuppositions: that is, the CC of factive predicates is assumed to be presupposed regardless of whether presuppositions are formally analyzed as conventionally coded constraints on the common ground (e.g., Heim 1983, van der Sandt 1992), as derived from lexical alternatives (e.g., Abusch 2010, Romoli 2015), or as pragmatically derived (e.g., Abrusán 2011, Simons et al. 2017). In other words, the assumption that the CC of factive predicates is presupposed is widely entrenched. It is these analyses that motivated the current investigation: it is crucial to understand which predicates they apply to.

As illustrated, Kiparsky and Kiparsky 1970 characterized presupposed content as content that is not part of the main assertion and diagnosed it using the projection diagnostic. This view of presuppositions has been standard for a long time: for instance, in Beaver and Geurts's 2014 article in the *Stanford Encyclopedia of Philosophy*, presupposed content is characterized as content that is "taken for granted, rather than being part of the main propositional content of a speech act" (abstract) and "[p]rojection from embeddings, especially negation, is standardly used as a diagnostic for presupposition" (§2). But the picture is more complex because presuppositions are not the only backgrounded content that projects: conventional implicatures, like those contributed by appositives, are backgrounded, projective content, too (see, e.g., Chierchia and McConnell-Ginet 1990, Potts 2005). Thus, projection alone does not suffice to identify presupposed content.

There are, however, properties on which presuppositions and conventional implicatures differ and we can use those properties to distinguish the two kinds of projective content. One such property is Potts's 2005 antibackgrounding requirement: conventional implicatures, unlike presuppositions, are unacceptable in a context in which they are part of the common ground. To illustrate, assume that the content of (6), that Lance Armstrong survived cancer, is part of the common ground. The continuation in (6a) is unacceptable because this content is conventionally implicated by the appositive *a cancer survivor*. The continuation in (6b), on the other hand, is acceptable: here, the content is presupposed, as the CC of *know*.

(6)   Lance Armstrong survived cancer.   (Potts 2005:34)

      a. #When reporters interview Lance, a cancer survivor, he often talks about the disease.

      b. And most riders know that Lance Armstrong is a cancer survivor.

Like the CC of *know*, illustrated in (6b), the CCs of clause-embedding predicates more generally are acceptable when their content is already part of the common ground: for instance, the example in (6b) is acceptable when *know* is replaced with *say, believe, take into account* or *confirm*. Thus, if a CC projects, we can assume that it is presupposed rather than conventionally implicated because it does not exhibit the antibackgrounding requirement typical of conventional implicatures. Consequently, even though not every projective content is a presupposition, we can continue to assume that projective CCs are presupposed, as is standard practice in the field.[3]

    There does not, to date, exist systematic empirical evidence that projection distinguishes factive predicates from others. To the contrary: there are hints in the literature that cast doubt on this assumption, which makes an investigation into the role of projection in classifying predicates a pressing issue. As a first hint, note that Kiparsky and Kiparsky 1970 did not only identify factive and non-factive predicates but predicates that are neither factive nor non-factive, including *acknowledge, remember* and *announce*. Such predicates do not specify "whether their complement sentences represent presuppositions by the speaker" (Kiparsky and Kiparsky 1970:164) and they "can be used with or without presupposition of the truth of their complement sentence" (Karttunen 1971a:340). Thus, the CCs of these predicates, which we refer to as 'optionally factive' predicates, can be presupposed and, therefore, can project at least occasionally. It is an open, empirical question whether factive and optionally factive predicates are distinguishable by how regular the projection behavior of their CCs is.[4]

    A second hint that casts doubt on the assumption that projection distinguishes factive predicates from others comes from Karttunen's 1971b observation 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 (7a), the polar questions in (7b) and the conditionals in (7c). Karttunen (1971b) suggested that "all the examples in [(7a)] presuppose that John had not told the truth" (p.63) and that a speaker who utters (7b) or (7c) 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 (7b) and (7c) 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, not properly factive, in contrast to predicates like *regret, forget, resent* and *make clear*.

(7)   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.

---

[3]One exception are Anand and Hacquard (2014), who suggested (pp.74f.) that even though the CC of predicates like *acknowledge, admit* and *confirm* projects, the CC is not presupposed. They did not, however, offer a diagnostic for distinguishing projective CCs that are presupposed from projective CCs that are not presupposed. Ultimately, their argument against classifying such predicates as factive rests on the observation that their CC is not entailed.

[4]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.

The observation that some factive predicates are merely semi-factive again gives rise to the question of whether factive predicates are distinguishable from other predicates by how regular the projection behavior of their CCs is.[5]

As a third hint, note that not even the CCs of predicates assumed to be properly factive are necessarily presupposed. Consider *be aware* and *know*, which Karttunen (1971b) took to be properly factive: their CCs do not project in the naturally occurring examples in (8) from Beaver 2010. The observation that the CCs of factive predicates need not project and that the CCs of optionally factive predicates may project further increases the need to investigate whether projection distinguishes factive predicates from others.

(8) 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)

   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)

Finally, there are some experimental results that suggest that projection does not distinguish factive predicates from others. Tonhauser et al. 2018 investigated the projection of the CC of 12 clause-embedding predicates in polar questions, including 10 factive predicates and 2 optionally factive ones, namely *establish* and *confess*.[6] As expected, Tonhauser and her colleagues found that the CC of the 12 predicates investigated was projective and that the CC of the 10 factive predicates was more projective than that of the 2 optionally factive ones. However, they also observed significant projection differences between the factive predicates, which gives rise to the question of whether projection would continue to distinguish factive from optionally factive predicates, if a greater variety of optionally factive predicates were investigated. The findings of de Marneffe et al. 2019, who investigated the projection of the CC of 45 clause-embedding predicates in naturally occurring data, suggest that this is not the case: although the CCs of factive predicates were among the most projective, the CCs of factive predicates were not categorically more projective than that of other predicates. We present these findings in detail in section 2.3. Hints like these motivate a systematic investigation of whether projection identifies a coherent class of factive predicates.

Entailed.   On the definition in (3b), the CC of factive predicates is presupposed and entailed, that is, for a predicate to be factive, its CC must project and also be an indefeasible inference from an affirmative matrix sentence with the predicate. The literature reveals disagreements for many predicates about whether the CC is entailed. 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 (9a). Anand and Hacquard (2014:76), however, pointed out that *inform* can be modified by *falsely*, as in the naturally occurring example in (9b), which they took to show that the CC does not necessarily follow from a true affirmative matrix 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 (10) again suggests that the CC does not necessarily follow from affirmative matrix sentences with *establish*.

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

   b. Doctors falsely informed my parents that I was dying.[7]

---

[5]Contrary to what Karttunen 1971b assumed, non-projection of the CC of semi-factive predicates is possible not just in polar questions and conditionals (Beaver 2010).

[6]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).

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

(10) Diplomacy was repudiated, but with the media's help it was falsely established that the 'villain' was refusing to negotiate.[8]

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 (11), where the truth of the CC does not follow from the affirmative matrix sentences.

(11)   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 (11) 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 matrix sentences as long as there is no indication that the beliefs of the speaker/writer and the subject of the attitude are not aligned; along these lines see also Djärv 2019. Other works maintained that emotive predicates are factive despite examples like (11) (e.g., Gazdar 1979a, Abrusán 2011, Anand and Hacquard 2014): they hypothesized that examples like those in (11) 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 (12) show that the truth of the CC does not necessarily follow from true affirmative matrix 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.

(12)   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)

   c. Most educated people already have opinions about language. They know that it is man's most important cultural invention [...] I will try to convince you that every one of these common opinions is wrong!   (Pinker 1994:17f.)[9]

In short, there is disagreement for many predicates about whether their CC is entailed. Under the definition in (3b), the uncertainty about whether the CC is entailed means that it is unclear whether these predicates are factive. There has, to date, not been a systematic investigation of which clause-embedding predicates have entailed CCs.[10]

---

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

[9]We thank Manfred Krifka (p.c.) for pointing out this example to us.

[10]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. However, because the predicates in their stimuli were realized in polar questions (e.g., *Is Maria aware that Mike is moving back to Chicago?*), it is not clear that the observed differences bear on whether the CC is entailed.

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 to differences in how factive predicates are defined and it is compounded by uncertainty about whether projection categorically distinguishes factive predicates from others and by disagreement about whether the CC of particular predicates is entailed. This paper investigates the projection and entailment of the CCs of 20 clause-embedding predicates, with the goal of understanding whether these two properties identify a coherent class of factive predicates. The predicates we investigate are listed in (13) with the categories they are typically taken to fall into. Specifically, the 5 factive predicates in (13a) 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 (13b) 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.[11] The 9 optionally factive predicates in (13c) include *acknowledge, admit* and *announce* (see Kiparsky and Kiparsky 1970) as well as *confirm, inform, confess, establish, hear* and *prove*.[12]

(13) 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 (3a), 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 identifies a class of factive predicates. Under the definition in (3b), 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. These results allow us to assess whether projection and entailment jointly identify a class of factive predicates.

---

[11] Anand and Hacquard (2014) assumed that *demonstrate* is veridical, in contrast to Anand et al. 2019. This latter work also took *reveal* to be non-factive, in contrast to, for instance, Egré 2008, Wyse 2010 or Tonhauser et al. 2018.

[12] Different categories have been assumed for these 6 predicates. Anand and Hacquard (2014) took *confirm* and *inform* to not be factive, while Schlenker (2010) took *inform* to be factive. For *prove*, White and Rawlins (2018) suggested that it is veridical and non-factive, but Anand and Hacquard (2014) took it to be non-veridical and non-factive. Swanson (2012) took *confess* to be factive, while 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. Swanson (2012) took *establish* to be non-factive, but Wyse (2010) listed it under the factive predicates. We also included *hear*: even though it is often considered a factive predicate (e.g., Beaver 2010, Anand and Hacquard 2014), *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 in our experiments.

2 Experiment 1: Presupposition

Exps. 1a and 1b were designed to investigate which of the CCs of the 20 clause-embedding predicates in (13) are presupposed.[13] 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, 2020, 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 (14) for the content of the appositive, that Martha's new car is a BMW.[14]

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

To assess projection, participants are asked whether the speaker is certain of the content; for instance, in (14), participants are asked whether Patrick is certain that Martha's new car is a BMW. If a participant takes the speaker to be certain of the content, we assume that the content projects, that is, the content is presupposed by the speaker; if a participant does not take the speaker to be certain of the content, we assume that the content does not project, that is, the content is not presupposed.

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.[15] 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 (3a), 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.

Materials   Exp. 1a investigated the projection of the CC of the 20 clause-embedding predicates in (13). 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 (15), which realize the complement clause *Julian dances salsa*. Eventive predicates, like *discover* and *hear*,

---

[13]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].

[14]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.

[15]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, the quantity of interest, in this case speaker commitment, is only indirectly measured.

were realized in the past tense and stative predicates, like *know* and *be annoyed*, were realized in the present tense. The indirect 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.

(15)   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 control stimuli, whose main clause content was investigated. For instance, in the sample control stimulus in (16), 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.

(16)   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 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, based on self-declared non-native speaker status and other 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 CCs. 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 result suggests that projection does not categorically distinguish factive predicates from optionally factive and non-factive ones.
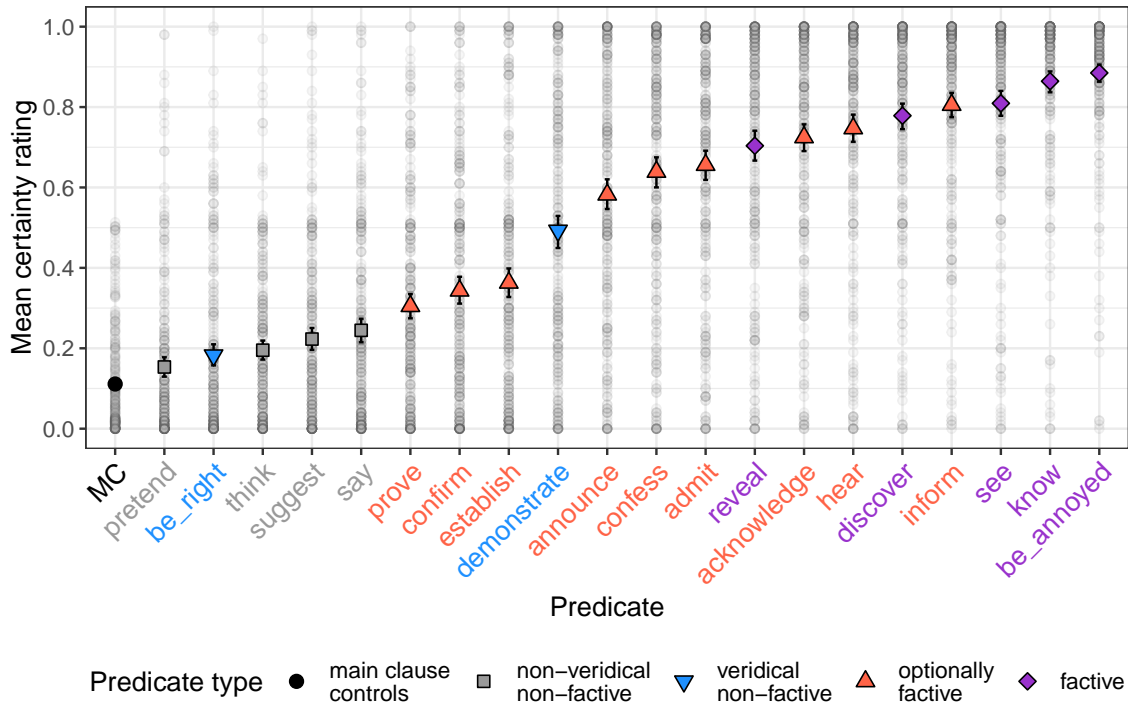

Figure 2: Mean certainty ratings by predicate in Exp. 1a. Error bars indicate 95% bootstrapped confidence intervals. Light gray dots indicate individual participants' responses.

In view of this result, one might ask whether the categorization of clause-embedding predicates assumed in (13) is correct: perhaps there is a different place in which a line can be drawn between factive predicates and others 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.[16]

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).[17] 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 result suggests that the CC of each of the 20 predicates is projective compared to non-projective main clause content, albeit to different degrees.[18] Thus, to distinguish factive predicates from others in this set of 20 predicates, one would need to distinguish one group of projective CCs from another group of projective CCs.

In sum, the results of Exp. 1a suggest that projection does not identify a coherent class of factive predicates. It is possible, however, 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.[19]
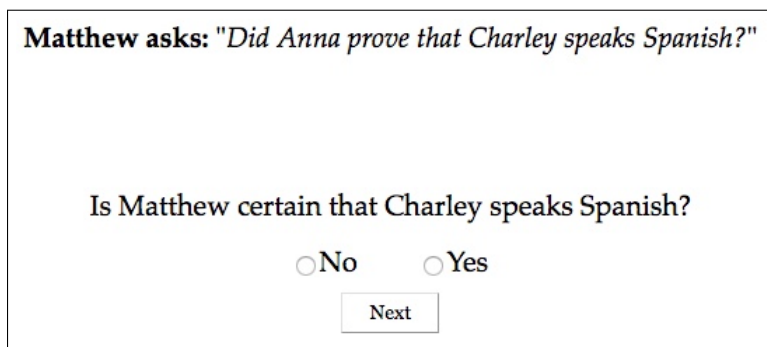
---

[16]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.

[17]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.

[18]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 13 for the model code.

[19]Exps. 1a, 2a and 3a were run in 2017 and 2018, whereas Exps. 1b, 2b and 3b 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' (coded as 1) or 'no' (coded as 0) 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.[20] 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 correlation). The critical results 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.[21] 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 provide support for the definition of factive predicates in (3a), according to which factive predicates are categorically distinguished from optionally factive and non-factive ones by the projection of the CC.

---

[20]Recall that participants gave binary certainty ratings, with 'yes' coded as 1 and 'no' coded as 0. To increase legibility, these ratings are jittered vertically in Figure 4 and other figures like it. Consequently, the individual certainty ratings should not be read against the y-axis.

[21]According to Spector and Egré (2015:1739), it is not clear whether the CC of *say* is projective. The results of our Exps. 1 suggest that it is weakly projective.
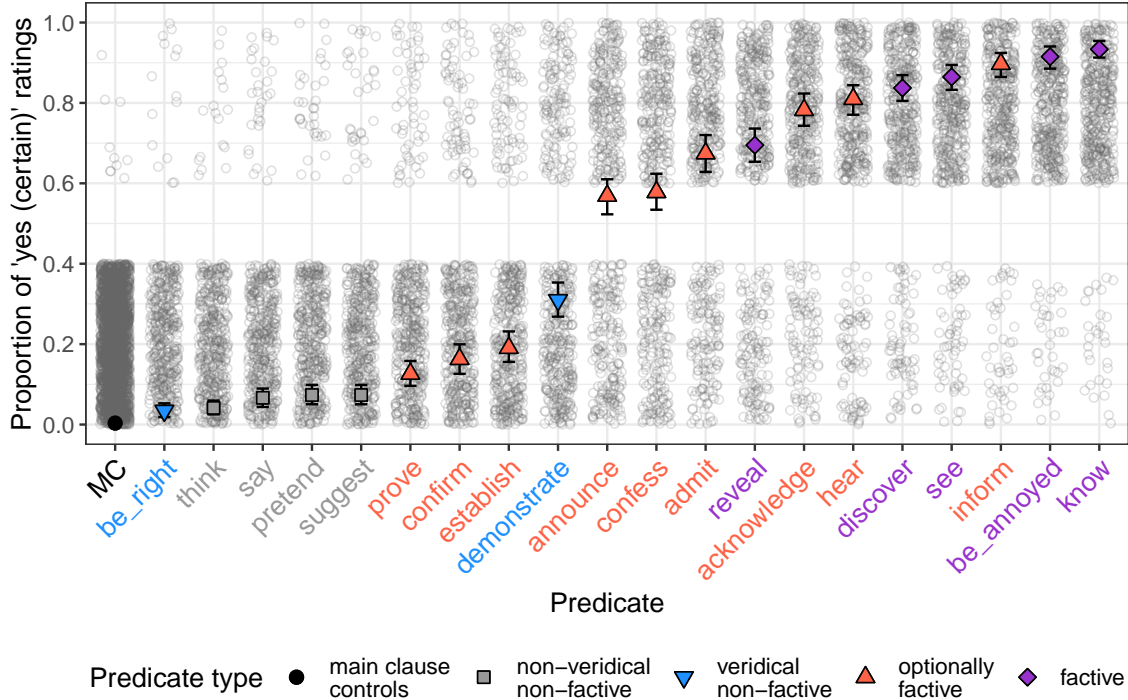
Figure 4: Proportion of 'yes' ratings by predicate in Exp. 1b. Error bars indicate 95% bootstrapped confidence intervals. Gray dots indicate individual participant responses (either 0 or 1, jittered vertically for legibility).

2.3 Interim summary and converging evidence

The results of Exps. 1 suggest that projection does not categorically distinguish factive predicates from others: contrary to what is expected on definition (3a), we observed that the CCs of predicates standardly classified as factive are not categorically more projective than the CCs of all other predicates. We now discuss converging evidence to this effect 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).[22]

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 (17), with *know* in the target sentence embedded under negation:

(17)   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 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.[23]
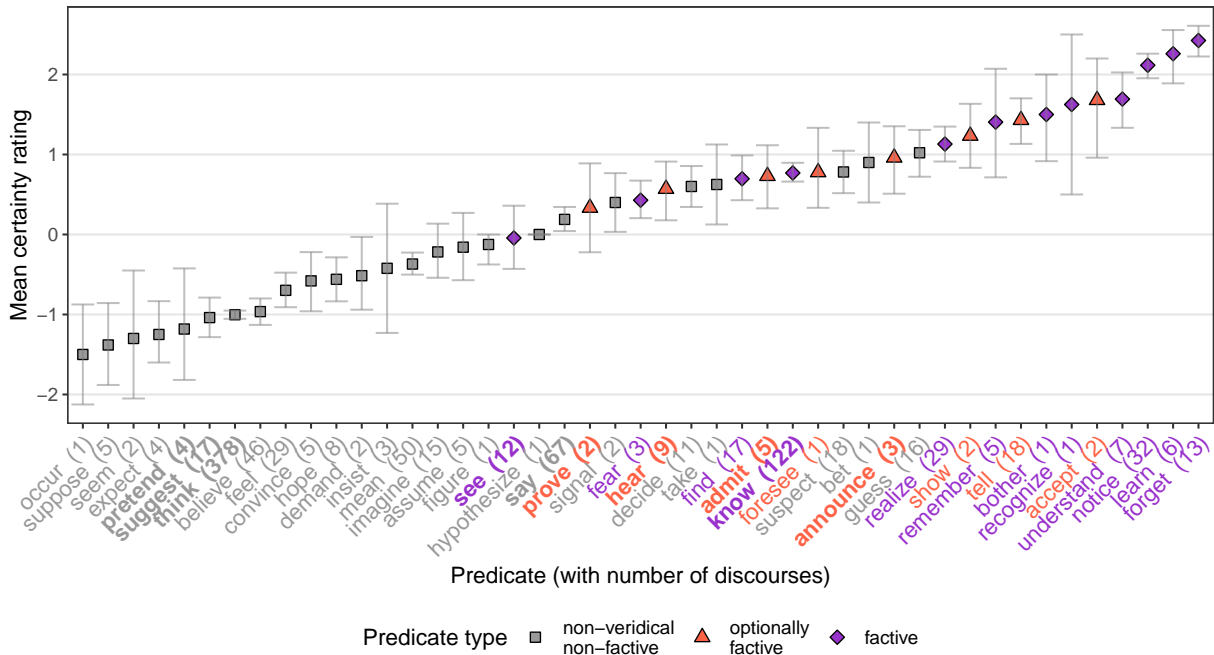

Figure 5: Mean certainty ratings by predicate, with number of discourses in parentheses, for 982 discourses in the CommitmentBank (de Marneffe, Simons, and Tonhauser 2019). Error bars indicate 95% bootstrapped confidence intervals. Predicates included in our Exps. 1 shown in bold.

A second piece of converging evidence that projection does not 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 (18). 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).

(18)   The GAO has not indicated that it is unwilling to compromise.   (Ross and Pavlick 2019:2234)

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 (*see* and *saw*, which are both included in the VerbVeridicality dataset, are plotted separately). As shown, projection does not categorically

---

[23]de Marneffe et al. 2019 found that a model that predicts certainty ratings from factivity captures less of the variance than a model that predicts certainty ratings from predicate lemma. This result is replicated in our Exps. 1, as discussed above.

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.
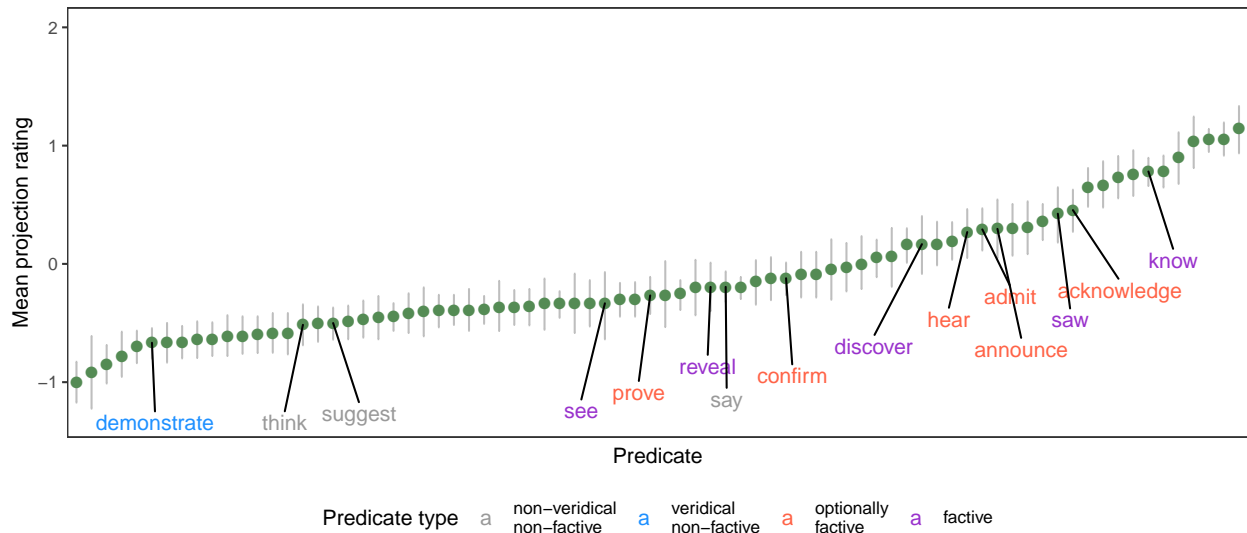

Figure 6: Mean projection rating by predicate in green, with 95% bootstrapped confidence intervals, for the 78 predicates in the VerbVeridicality dataset (Ross and Pavlick 2019), with labels for the 15 predicates featured in our experiments.

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 White and Rawlins (2018) referred to as low content arguments, as shown in (19) for *know*. The predicates were embedded under negation in stimuli like (19a) and under negation, in the antecedent of a conditional and in a question in stimuli like (19b). To assess projection, participants were asked to respond to the question *Did that thing happen?* for stimuli like (19a) and to respond to the question posed by stimuli like (19b). The response options were 'yes', 'maybe or maybe not' and 'no'.

(19) 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. As shown, projection of the CC does not categorically distinguish factive predicates from others: 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.
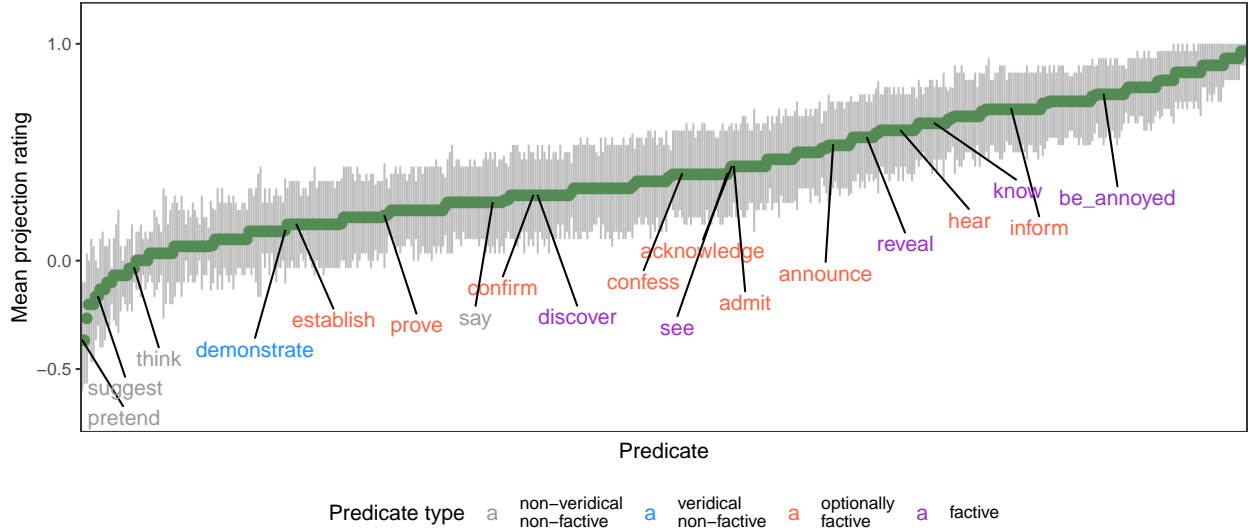
Figure 7: Mean projection rating by predicate in green, with 95% bootstrapped confidence intervals, for the 517 predicates in the MegaVeridicality dataset (White and Rawlins 2018, White et al. 2018), with labels for the 19 predicates featured in our experiments.

In sum, contrary to what is expected under the definition of factive predicates in (3a), the results of our Exps. 1a and 1b as well as our meta-analyses of the CommitmentBank, the VerbVeridicality dataset, and the MegaVeridicality dataset suggest that projection of the CC does not identify a coherent class of factive predicates. The challenge to definition (3a) is compounded by the diversity of the empirical evidence: it comes from constructed and naturally occurring examples, from examples presented with and without a context, from examples with diverse and with minimal lexical content, from clause-embedding predicates embedded under different entailment-canceling operators, and from projection ratings collected through different response tasks.

The experiments described in the next section were designed to explore whether the definition of factive predicates in (3b), according to which presupposition and entailment jointly identify a class of factive 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 whether the definition of factive predicates in (3b) identifies a coherent class of factive predicates. 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 ϕ is true is also a world in which ψ is true. If there is at least one world in which ϕ is true but ψ is false, then ϕ does not entail 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 (20). According to the 'inference diagnostic' in (20a), 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 (20b), an entailment relationship holds if and only if an utterance of the indicative matrix sentence that is followed by a denial of the CC is contradictory.

(20) 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 a sentence of the form ϕ *and not* ψ is 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 (13), 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 (20a). 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.

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 (21), 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 (21a) whether it follows that Danny ate the last cupcake and about (21b) whether it follows that Emma studied on Saturday morning.

(21) 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 (22), 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 (23), definitely do not follow from these true statements: for instance, in (23), 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.

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

(23) **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.

What is true: Edward proved that Grace visited her sister.

Does it follow that Grace visited her sister?

definitely doesn't follow ▭ definitely follows

Next

Figure 8: A sample trial in Experiment 2a

To familiarize participants with the task, they first responded to the two familiarization stimuli in (24), where the inference tested was the CC. Participants who rated (24a) in the lower half of the sliding scale or (24b) 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 (24a) and in the lower half for (24b).

(24)   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

18

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.
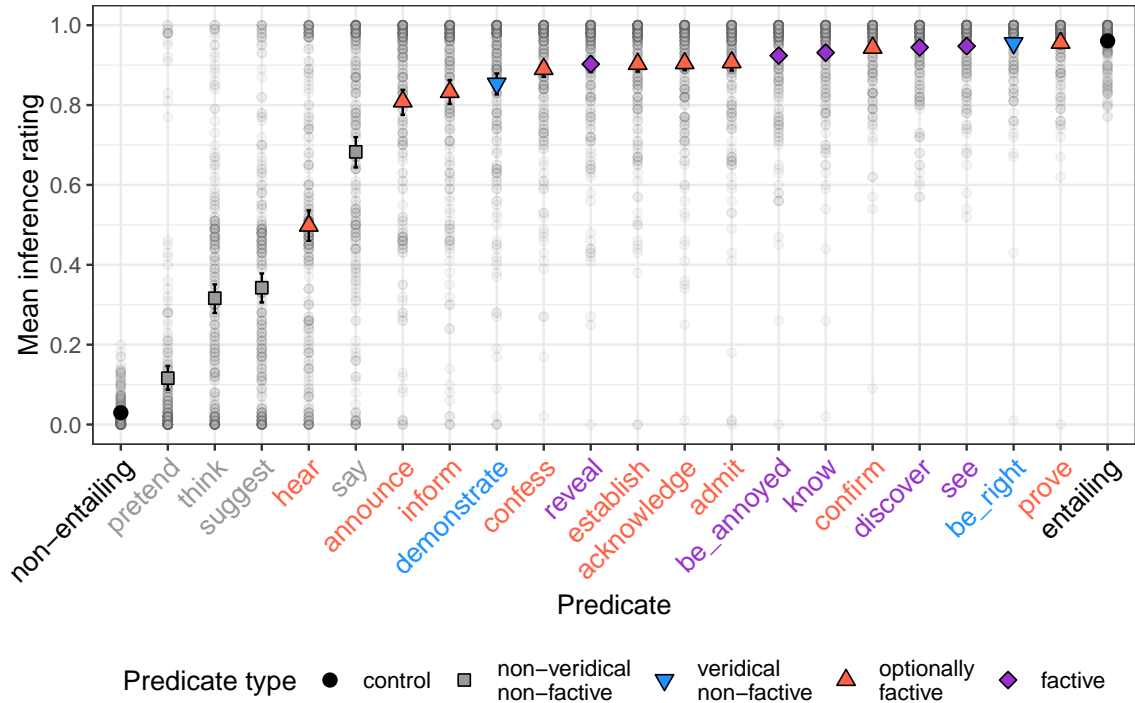

Figure 9: Mean inference rating by predicate in Exp. 2a, including the non-entailing and entailing controls. Error bars indicate bootstrapped 95% confidence intervals. Light gray dots indicate individual participants' ratings.

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.[24] 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.

[24] 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 13 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 (3b), 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 (3b), these two predicates are factive. This is a very unsatisfying conclusion because 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 (3b), 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 rating for *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 result of Exp. 2a—that *see, discover know, be annoyed* and *reveal* are not factive under definition (3b)—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 and, therefore, not factive under definition (3b). 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.

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.

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 ordering of the predicates is almost entirely identical to the gradient measure of Exp. 2a, as evidenced by the very high Spearman rank correlation of .989 between Exps. 2a and 2b (see Supplement E for a visualization).
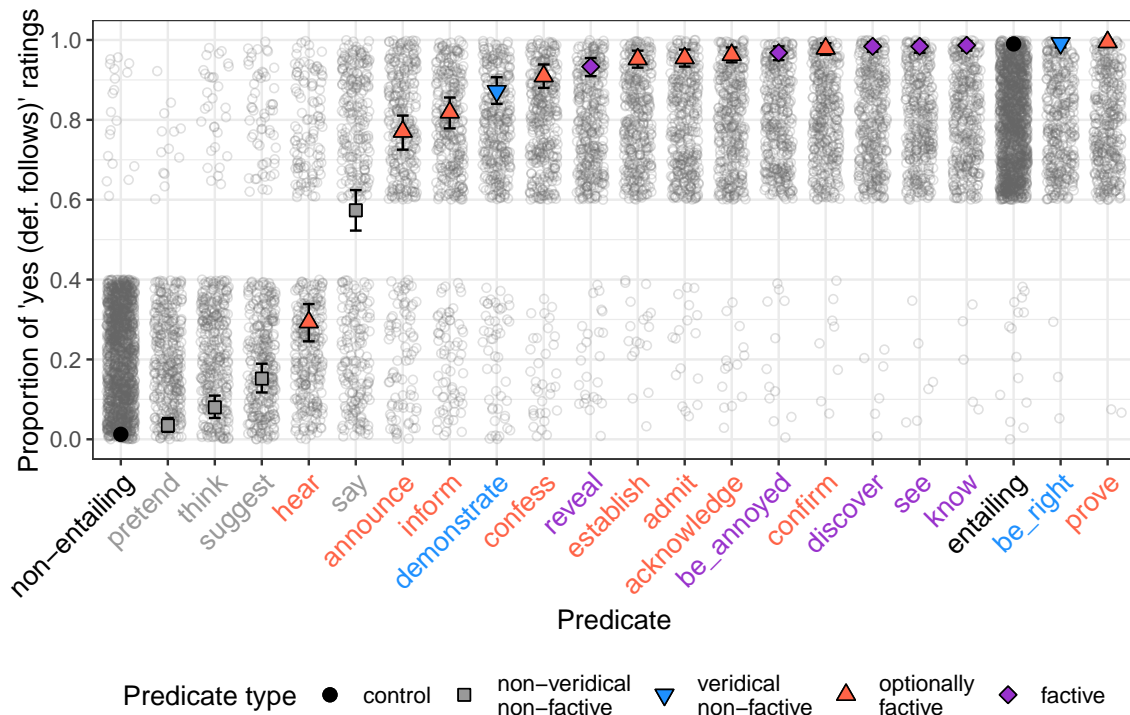
Figure 11: Proportion of 'yes' responses by predicate in Exp. 2b. Error bars indicate 95% bootstrapped confidence intervals. Vertically jittered light gray dots indicate individual participants' responses of 'yes' (coded as 1) and 'no' (coded as 0).

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

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 (3b), 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 among the least projective of the 20 predicates investigated while that of *know* was among the most projective. Thus, this set of predicates can hardly be considered a natural class. Second, the predicates *be annoyed* and *reveal* are not factive under the definition of factive predicates in (3b) and the predicate *demonstrate* is not veridical.

The results of both Exps. 2a and 2b have unexpected implications for the identification of factive predicates. One might ask whether these results 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 (20b). 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 (25):

(25)   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.

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 (26), 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 in the complement clause.

(26)   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 (27), 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 (27), 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 (28), to be at floor, because the content of the second clause does not contradict the content of the first clause. For instance, in (28), 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.

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

(28)   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 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.

         Margaret: *"Edward heard that Mary is pregnant, but she isn't."*

         Is Margaret's utterance contradictory?

     definitely no   **definitely yes**

                        Next

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.
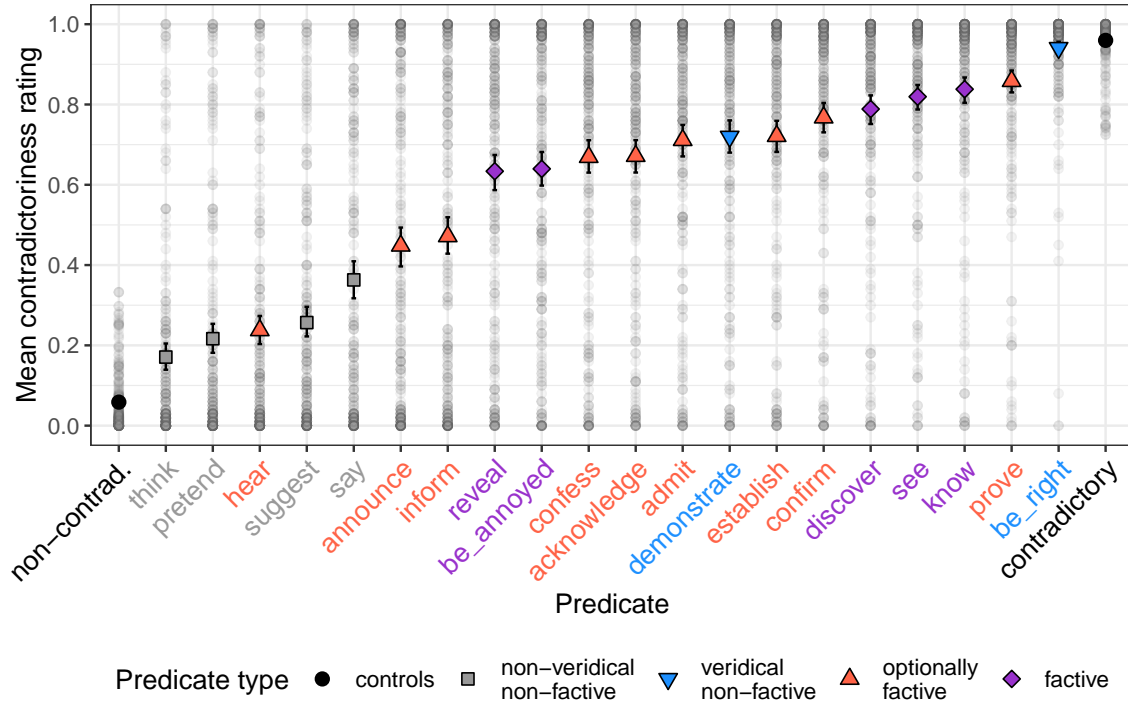

Figure 13: Mean contradictoriness rating by predicate in Exp. 3a, including the non-contradictory and contradictory controls. Error bars indicate bootstrapped 95% confidence intervals. Light gray dots indicate individual participants' ratings.

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.[25] 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 in (3b).

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 result in more CCs being classified as entailed. To assess this possibility, we conducted Exp. 3b, which was

---

[25]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 13 for the model output.
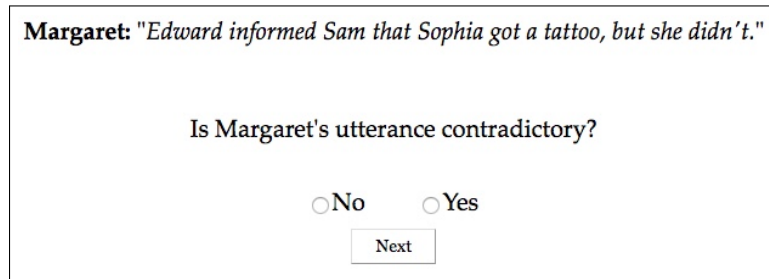
identical to Exp. 3a but employed a two-alternative forced choice task.

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.

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.

> **Margaret:** *"Edward informed Sam that Sophia got a tattoo, but she didn't."*
>
> Is Margaret's utterance contradictory?
>
> ○ No   ○ Yes
>
> Next

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 results 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 (3b).

3.5 Interim summary and converging evidence

According to definition (3b), a predicate is factive if and only if its CC is presupposed and entailed. In Exps. 2 and 3, 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 found that the CCs of *prove* and *be*
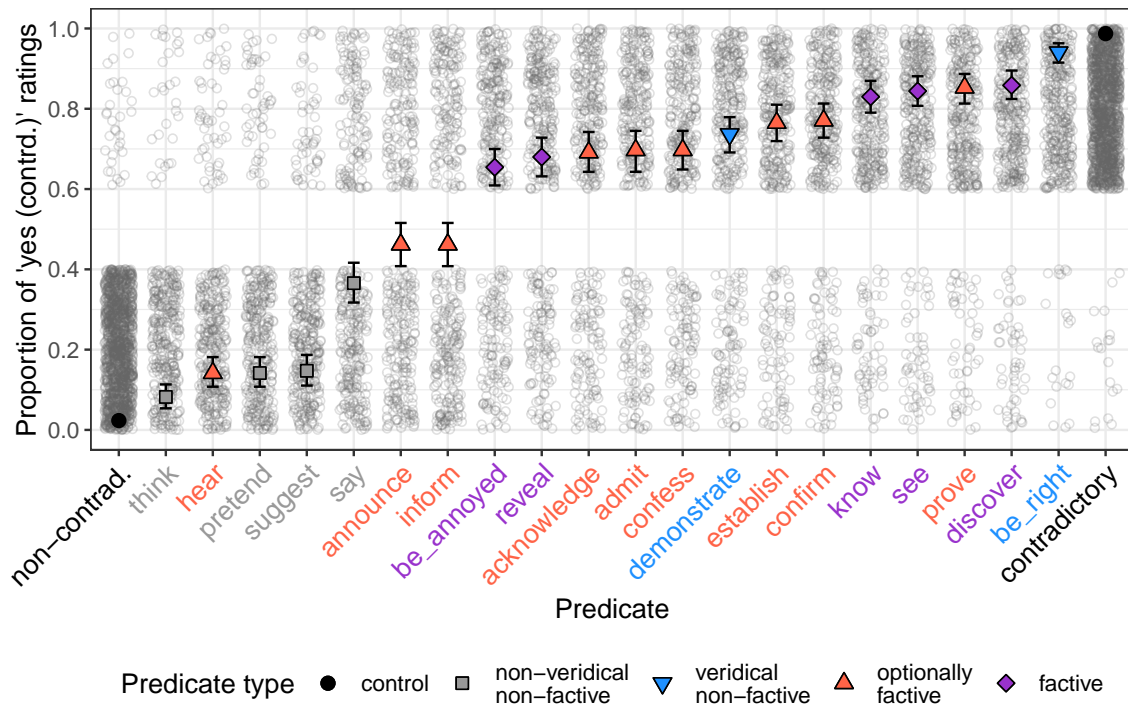
Figure 15: Proportion of 'yes' ratings by predicate in Exp. 3b. Error bars indicate 95% bootstrapped confidence intervals. Vertically jittered light gray dots indicate individual participants' ratings of 'yes' (coded as 1) and 'no' (coded as 0).

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).[26] Given that the CCs of all 20 predicates are at least mildly projective, this means, under definition (3b), that the predicates *prove* and *be right* are factive according to the results of Exp. 2a, that the predicates *prove, be right, know, see, discover* and *confirm* are factive according to the results of Exp. 2b, and that none of the 20 predicates are factive according to the results of Exps. 3.

These results lead us to the conclusion that projection and entailment jointly do not identify a coherent class of factive predicates, contrary to what is expected on definition (3b). First, given the results of the gradient inference diagnostic (Exp. 2a), the predicates *be right* and *prove* are factive, but it is unsatisfying to consider them factive predicates because they both have only weakly projective CCs. Second, given the results of the categorical inference diagnostic (Exp. 2b), the predicates *prove, be right, see, discover, confirm* and possibly *know* are factive, but this is a heterogeneous set of predicates and not a natural class: the CC of *be right* was among the least projective and that of *know* was among the most projective. Finally, given the results of the contradictoriness diagnostic, none of the 20 predicates is factive. In short, the set of predicates

---

[26]The differences in the results of Exps. 2 and 3 highlight the importance for future research to investigate the validity of entailment diagnostics. While the ratings were overall lower with the contradictoriness diagnostic than the inference diagnostic, the high Spearman rank correlations between the ratings from the two diagnostics—.954 for the two gradient measures and .934 for the two categorical measures—suggests that both are measuring a similar underlying concept. Moreover, the entailed controls received very high ratings across Exps. 2 and 3, suggesting that all four measures are able to diagnose entailed content with some reliability. However, de Marneffe et al. (2012:329) have suggested that veridicality ratings (comparable to our inference ratings) are influenced by pragmatic reasoning; see also Pavlick and Kwiatkowski 2019.

identified as factive is either heterogeneous with respect to projection, i.e., not a natural class, or empty.

Converging evidence for our conclusion comes from the VerbVeridicality and MegaVeridicality datasets. 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 (29).

(29) The GAO has indicated that it is unwilling to compromise.   (Ross and Pavlick 2019:2234)

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 (3b). 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,[27] 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, projection and entailment jointly do not identify a coherent class of factive predicates, contrary to what is expected on definition (3b).

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.

---

[27]The VerbVeridicality dataset does not include projection ratings for main clause content to which the projection ratings for the CCs could be compared. The cutoff of a mean projection rating of .3 (on a scale from -2 to 2) is arbitrary but comparable to the mean certainty rating of .15 (on a scale from 0 to 1) for *pretend* in our Exp. 1a.

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 (30) for *know*. To diagnose veridicality, participants were asked to respond to the question *Did that thing happen?*; the response options were 'yes' (which we again coded as 1), 'maybe or maybe not' (0) and 'no' (-1).

(30) 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. Figure 17 plots the mean veridicality ratings, 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 result is compatible with the assumption that the CCs of these 97 predicates are entailed.[28] If we assume, again, that CCs with mean projection ratings of at least .3 are presupposed, then 92 of the 97 predicates are factive. This set is heterogeneous with respect to projection, including predicates like *bother*, with a mean projection rating of .97, *inform* (.7), *know* (.63), *grumble* (.5), *notify* (.4), *gloat* (.4) *convey* (.4) and *confirm* (.3). Thus, on the MegaVeridicality set, too, projection and entailment jointly do not identify a coherent class of factive predicates.
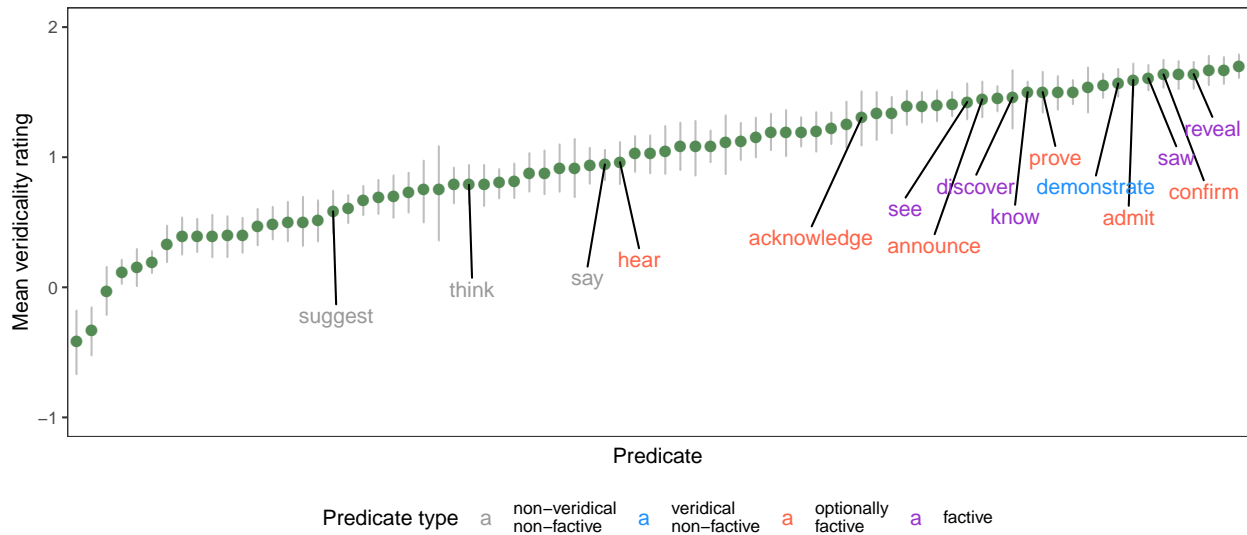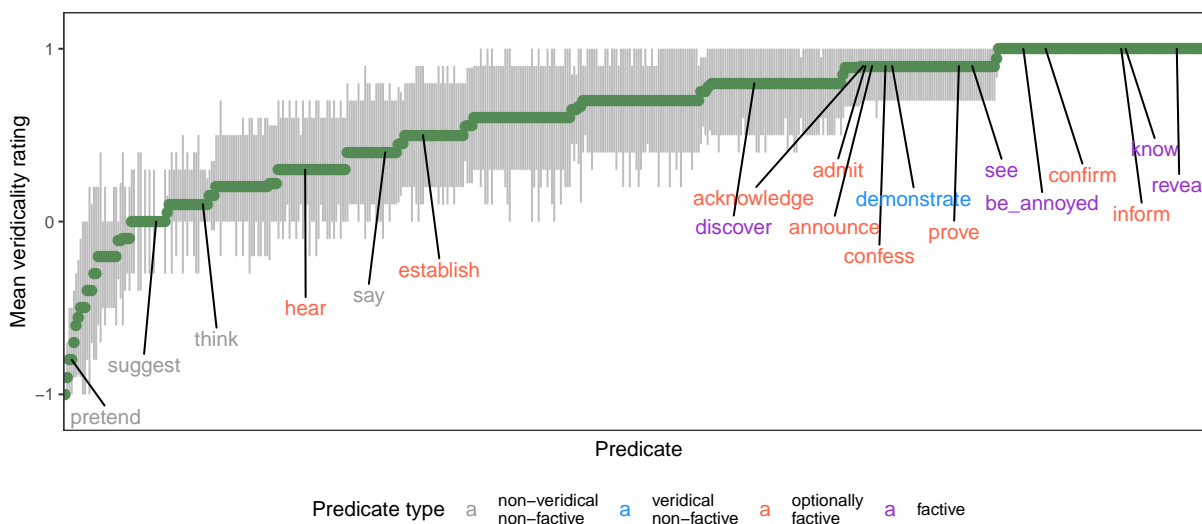

Figure 17: Mean veridicality rating by predicate in green, with 95% bootstrapped confidence intervals, for the 517 predicates in the MegaVeridicality dataset (White and Rawlins 2018, White et al. 2018), with labels for the 19 predicates featured in our experiments.

This conclusion dovetails with that of White and Rawlins (2018), who assumed definition (3b). They pointed out (p.228) that if one were to assume a majority response as threshold for whether the CC is

---

[28]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 indefeasibly 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)

presupposed and entailed, there are 199 factive and 292 non-factive predicates in the MegaVeridicality dataset, but also that "there are not necessarily clear dividing lines between these classes present in the data, suggesting that speaker's inferences about [projection and entailment; the anonymous authors] are generally quite gradient and likely influenced by the fine-grained semantics of particular verbs".

In sum, contrary to what is expected under the definition of factive predicates in (3b), the results 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 or heterogeneous with respect to projection.

4 General discussion

Sections 2 and 3 reported the results of experiments designed to investigate whether two prevalent definitions, repeated below, identify a class of factive predicates.

(3) 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, Schlenker 2010, Abrusán 2011
                                                             Anand and Hacquard 2014, Spector and Egré 2015)

The experiments in section 2 investigated the definition in (3a): Assuming projection as a diagnostic for presupposed CCs, the results of these experiments as well as the meta-analyses of three existing datasets suggest that projection does not identify a class of factive predicates. The experiments in section 3 investigated the definition in (3b): The evidence from our experiments and meta-analyses of existing datasets suggests that depending on how entailment is diagnosed, there either are no factive predicates or that the set of predicates identified as factive is heterogeneous with respect to projection, i.e., not a natural class. In sum, the results of our experiments do not provide empirical evidence for a coherent class of factive predicates under either of the two definitions in (3).

Our main motivation in conducting these investigations was to understand which predicates analyses of the projection of the CC of clause-embedding predicates apply to, as these analyses 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. Given the results of our experiments, we are unfortunately no closer to answering the question of which predicates these analyses apply to.[29]

---

[29]To be clear: Contemporary analyses of the projection of the CCs of factive predicates can predict the non-projection of factive CCs in particular utterances. For instance, in Heim 1983 and van der Sandt 1992, the CC is locally accommodated when its projection would result in a contradictory or uninformative utterance. Although projection of the CC does not result in contradictory or uninformative utterances in the stimuli in our Exps. 1, one might assume, as suggested by an anonymous reviewer, a) that some participants enrich the context in which the polar question stimuli were presented with information that does result in a contradiction or uninformativity, and b) that participants are more likely to enrich the context with such information with some factive predicates than others. Under these assumptions, these analyses may be compatible with the results of our Exps. 1, namely that the projection of the CC of factive predicates is heterogeneous and not categorically higher than that of other predicates. Neither assumption, however, strikes us as independently motivated.

Where do we go from here? There are three broad ways forward. First, one might ask whether there are more appropriate diagnostics for presupposed or entailed content than the ones we have applied, or whether there are more suitable definitions of factive predicates than those in (3). Second, one might challenge our conclusion that the results of our investigations do not identify a class of factive predicates. We engage with several possible objections in section 4.1. Alternatively, one could give up on the assumption that there is a coherent class of factive predicates (see also Karttunen 2016 and Djärv 2019 for such moves). In section 4.2, we discuss what this move would mean for projection analyses.

4.1   Possible objections

The results of Exps. 1 as well as the meta-analyses of the existing datasets in section 2 are strongly suggestive of gradience in the projection of the CCs of clause-embedding predicates. Our discussions in sections 2 and 3 suggested that this gradience precluded identification of a class of factive predicates on both definitions in (3). In the following, we consider several empirical, statistical, and conceptual objections to this conclusion.

4.1.1   Objection 1: Noise might have obscured an underlyingly binary factivity category

Our results in Exps. 1 are contingent on a particular operationalization of projection – via the "certain that" diagnostic – and a particular behavioral measurement – via continuous slider ratings and binary 'yes'/'no' judgments. There are multiple ways in which experimental noise might have been introduced and masked an underlying binary factivity category: i) via the experimental items, which did not consist of the predicates alone, but of sentences in which the predicates were combined with particular subject noun phrases and complement clauses; ii) and which moreover were produced by a named speaker; iii) via the response tasks; and iv) via participant-internal noise. Each of i) to iv) could have introduced systematic noise into the ratings, yielding gradient responses for predicates that may underlyingly be either factive or non-factive.

The experiments reported here were designed to mitigate several possible sources of systematic noise: i) all predicates occurred in the same sentence structures and were combined with the same set of complement clauses across participants; ii) all predicates were presented in utterances situated in the same context; iii) ratings were collected across different tasks; and iv) ratings were obtained from a large number of participants, whose response variability was taken into consideration in the statistical analysis. Reasons to believe that these decisions mitigated noise include the replication in gradient orderings of predicates across continuous (slider ratings) and categorical (selections) response measures (Exp. 1a vs. 1b); the replication of Exp. 1a in Exp. 1 of Degen and Tonhauser 2021 with a Spearman rank correlation of .991; and the replication of gradient orderings of predicates in the meta-analyses of existing datasets, which varied in the experimental tasks and items. We therefore consider it unlikely that experimental noise might have masked an underlying binary factivity category, though there is much need for discussion in the experimental semantics/pragmatics literature about the relation between behavioral data and semantic/pragmatic categories.[30]

4.1.2   Objection 2: The distinction between factive and non-factive predicates might be fuzzy

Another objection to the conclusion that projection does not identify a class of factive predicates is that the expectation that the CCs of factive predicates should be more projective than those of all other predicates is too strong. Specifically, given that the CCs of factive predicates are among the most projective in both Exps. 1a and 1b, one might argue that to motivate a class of factive predicates it suffices to show that assuming such a class explains variability observed in the projection of the CCs of clause-embedding predicates. To engage with this possibility, we analyzed the certainty ratings of Exps. 1 in a Bayesian mixed effects

---

[30]For such a discussion in the experimental syntax literature, see Lau et al. 2014, 2017, Sprouse 2007, Sprouse and Almeida 2013, Hofmeister and Sag 2010, Keller 2000, Sorace and Keller 2005.

Beta regression model (Exp. 1a) and a logistic regression model (Exp. 1b) that predicted projection from a binary factivity predictor, whereby the five predicates in (13a) were classified as factive and the remaining predicates (as well as the main clause controls) as non-factive. We report mean estimates and 95% credible intervals for the effect of this binary factivity predictor. Under such an analysis of the predicates, the CCs of presumed factive predicates were indeed more projective, both in Exp. 1a ($\beta = 1.49, CI = [1.30, 1.67]$) and in Exp. 1b ($\beta = 3.99, CI = [3.40, 4.59]$).

A way to ask statistically whether binary factivity is a useful category to posit is to ask how much variability in the data is captured by the binary factivity models compared to the models that include the individual predicates as predictors, which were reported in sections 2.1 and 2.2. We refer to these latter models as predicate-specific models. Model comparison revealed that predicate-specific models were favored over binary factivity models in both Exp. 1a (WAIC of predicate-specific model: -18971.3 (SE= 359.3); WAIC of binary factivity model: -18643.9 (SE= 355.4)) and Exp. 1b (WAIC of predicate-specific model: 6444.5 (SE= 124.8); WAIC of binary factivity model: 6719.0 (SE= 123.7)).[31] This suggests that the qualitative observation made in sections 2.1 and 2.2, namely that projection variability is better captured by taking into account the individual predicates than a binary factivity classification, is borne out in rigorous model comparison.

4.1.3 Objection 3: There may be uncertainty about which category a predicate belongs to

It has been suggested to us, in personal communication and by a reviewer, that the observed gradience in projection may be compatible with a binary factivity category in combination with two assumptions: first, that predicates may be ambiguous between a factive lexical entry, on which the CC is presupposed and hence projects, and a non-factive lexical entry on which the CC is not presupposed (for such a proposal for clause-embedding predicates see Spector and Egré 2015:1736; for a similar proposal for evaluative adjectives see Karttunen et al. 2014); and, second, that interpreters may be uncertain about which lexical entry a speaker intended in their utterance. Under such a view, a categorical notion of factivity is upheld, and the observed projection gradience can be hypothesized to be due to listeners' uncertainty about the use of any given predicate, which may be resolved preferentially one way or another through pragmatic factors.

The suggestion is that one way to test whether it is possible to maintain a binary factivity category with these two assumptions is to fit so-called finite mixture models to the data (Jacobs et al. 1991, Richardson and Green 1997). These models allow for representing the possibility that multiple components, i.e., multiple sub-datasets, make up an overall dataset. For instance, the data for each predicate from Exp. 1a could be analyzed with mixture models with respect to how many components (one, two, three, four, five, . . .) best account for the data. To illustrate, consider Figure 18, which shows simulated distributions of certainty ratings together with the optimal number of Gaussian components.[32] Given the aforementioned assumptions, one could attribute certainty ratings that are best captured with one Gaussian component with a mean of .4, as in panel (a), to a predicate that only has a non-factive lexical entry. Likewise, one could attribute certainty ratings that are best captured with one Gaussian component with a mean of .85, as in panel (b), to a predicate that only has a factive lexical entry. Certainty ratings like those in panels (c) and (d), which are best captured by two Gaussian components, might then be attributed to predicates that have both a factive and a non-factive lexical entry, with differences between (c) and (d) being due to the predicates interacting differently with pragmatic factors that modulate projection. The projection gradience observed in Exp. 1a could be hypothesized to come about by virtue of one or two components being combined in various ways.

---

[31] The Widely Applicable Information Criterion (WAIC) is a measure of model quality that estimates the pointwise out-of-sample prediction accuracy from a fitted Bayesian model using the log-likelihood evaluated at the posterior simulations of the parameter values and penalizes models with more parameters over models with fewer (Watanabe and Opper 2010). We obtained WAIC values for each model using the `loo` package (Vehtari et al. 2017). Lower WAIC values indicate better models.

[32] Models were fit using the `mclust` and `mixtools` packages (Benaglia et al. 2009, Scrucca et al. 2016) in R.

(a) One Gaussian component, mean = .4.

(b) One Gaussian component, mean = .85.

(c) Two Gaussian components, mean = .6.

(d) Two Gaussian components, mean = .7.

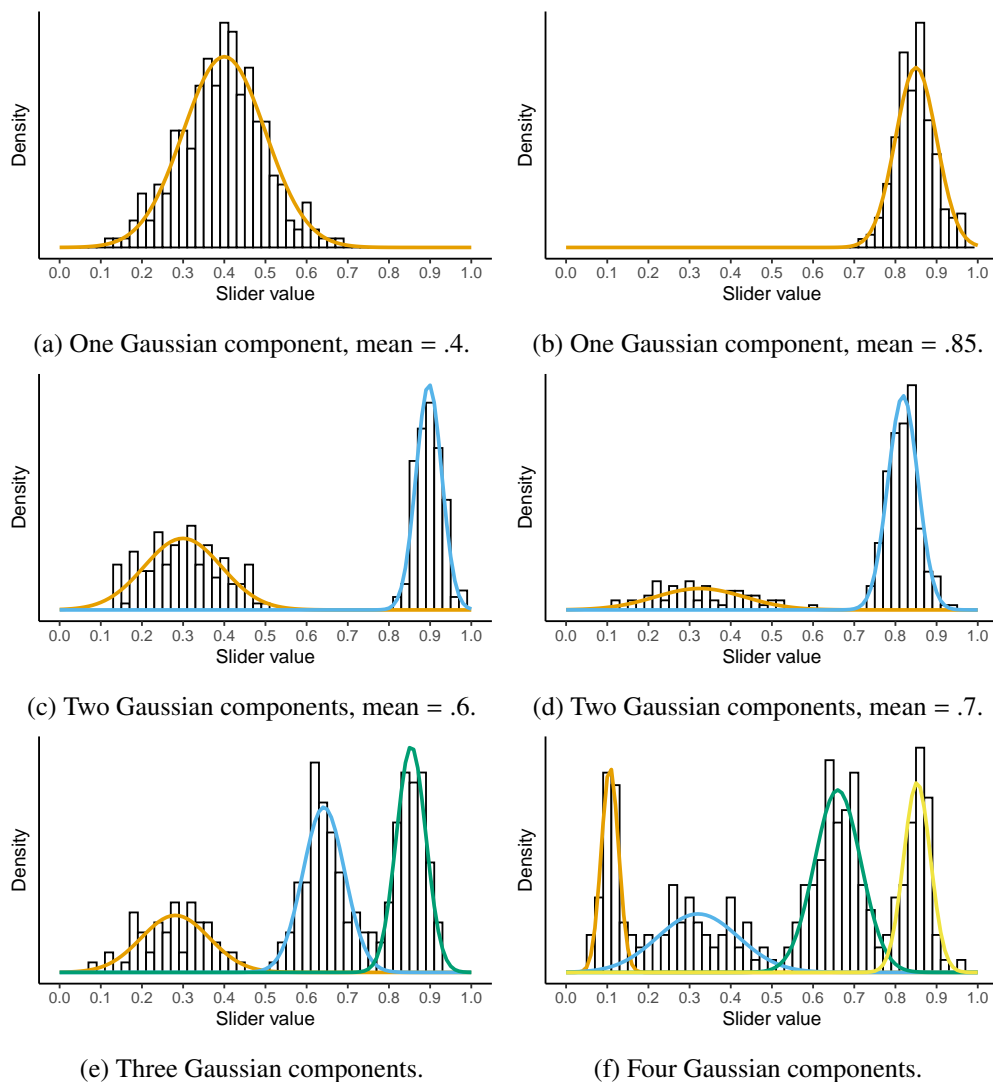(e) Three Gaussian components.

(f) Four Gaussian components.

Figure 18: Different simulated rating distributions with overlaid optimal number of Gaussian components.

We caution against this approach on conceptual grounds. What mixture models can do is supply us with an answer to the question of how many components optimally characterize the data. Mixture models are not, however, a theory of factivity or projection, and, as such, they do not specify what the generative model underlying the components is. Thus, while mixture models can be used to test the predictions of theories of factivity or projection that provide us with such a generative model, we caution against their application in the absence of such a theory. And, in fact, we do not currently have such a theory. The account sketched above, on which predicates may have a factive lexical entry, a non-factive one, or both, does not lead us to expect projection data distributions like those in panels (e) and (f) of Figure 18, that are best captured by three or four components, let alone distributions that are best captured by five or six components. Worse yet, in the absence of such a theory, we wouldn't know how to interpret data that are best captured by more than two components.[33] In short, we currently lack a principled reason for applying mixture models to the

---

[33]An additional problem: for simplicity we introduced mixture models with Gaussian mixtures. A preliminary investigation of the data from Exp. 1a via mixture models suggests that 15 of the 20 predicates are best characterized by 3-6 components; none are best characterized by just 1 component (see Supplement F). However, the assumption of normality made by Gaussian mixture

data reported here.

4.1.4 Summary

We have argued that the results of our experiments and the meta-analyses of existing datasets did not find empirical support for a class of factive predicates. Our discussion in sections 4.1.1 to 4.1.3 engaged with several objections to this conclusion. In doing so, the discussion highlighted several empirical, statistical, and conceptual considerations for future research on the classification of clause-embedding predicates.

4.2 Implications for projection analyses

The assumption that there is a class of factive clause-embedding predicates has played a central role in linguistic theorizing on a broad range of topics (see the list in section 1). In this section, we discuss the implications of our findings for one of these topics, namely projection analyses. As mentioned above, the explanandum of such analyses 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. If future research provides empirical support for a class of factive predicates (see section 4.1), then it is the predicates in that class that these analyses may apply to.

But even in the absence of a conclusive list of factive predicates, the results of our Exps. 1 as well as the meta-analyses of the three existing datasets point to exciting avenues for future research on projective content. First, our results show that the CCs of many more clause-embedding predicates than previously assumed are projective. This suggests 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 CC of factive predicates, to the exclusion of the CC of other clause-embedding predicates, our results mean that we are now tasked with investigating and formally analyzing the projection of the CC of a much broader set of clause-embedding predicates than before. Research on projective content has identified several factors that influence the projection of utterance content, including context, lexical content, prior beliefs, 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, Mahler et al. 2020, Mahler 2020, Tonhauser 2016, 2020, Tonhauser et al. 2018, 2019, Degen and Tonhauser 2021). 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 particular subsets, and how they contribute to the observed projection variability between clause-embedding predicates. A framework that offers a principled mechanism for modeling the probabilistic interaction of factors like lexical content, speaker reliability, and prior beliefs about likely meanings is the Rational Speech Act (RSA) framework (Bergen et al. 2016). We believe that implementing the interpretation of utterances with clause-embedding predicates in RSA offers a promising way forward.

Second, our results show that there is a lot of variability in how projective the CCs of the different clause-embedding predicates are. One of the factors that clearly plays a role in predicting projection variability is the lexical meaning of the clause embedding predicate. Our Exps. 1, as well as the results presented in Tonhauser et al. 2018, de Marneffe et al. 2019 and Degen and Tonhauser 2021, suggest that a binary distinction between factive and non-factive predicates is too coarse to adequately predict projection. We hypothesize that more fine-grained distinctions that are based on the lexical meaning and discourse use of clause-embedding predicates are critical; specifically, we hypothesize that the CCs of classes of predicates that share lexical meanings and discourse uses are similar in projection.[34]

---

models does not hold in the bounded slider ratings data of Exp. 1a. One would need to fit mixtures of Beta distributions, for which statistical inference about optimal cluster number is much more complex.

[34]Similarly, White 2021 recently proposed that lexical semantic properties other than factivity may be implicated in another

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 others. This paper investigated whether the content of the complement of 20 English clause-embedding predicates is projective and entailed, with the goal of identifying whether there is a class of factive predicates according to two definitions found in the literature. Our experiments revealed that projection alone does not categorically distinguish factive predicates from others, and that a coherent class of factive predicates also does not emerge from the additional consideration of entailment, as assessed by two standard diagnostics. In addition to considering how the search for a class of factive predicates might continue, we also suggested that our results open up several exciting new avenues for empirical and theoretical research on the projection of the contents of the complements of clause-embedding predicates.

References

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.
Abusch, Dorit. 2010. Presupposition triggering from alternatives. *Journal of Semantics* 27:37–80.
Anand, Pranav, Jane Grimshaw, and Valentine Hacquard. 2019. Sentence embedding predicates, factivity and subjects. In C. Condoravdi and T. Holloway King, eds., *Tokens of meaning: Papers in honor of Lauri Karttunen*. Stanford: CSLI Publications.
Anand, Pranav and Valentine Hacquard. 2014. Factivity, belief and discourse. In L. Crnič and U. Sauerland,

---

domain where factivity has been taken to play a role, namely whether a predicate is responsive.

eds., *The Art and Craft of Semantics: A Festschrift for Irene Heim*, pages 69–90. MIT Working Papers in Linguistics.

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

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.

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

Benaglia, Tatiana, Didier Chauveau, David R. Hunter, and Derek Young. 2009. mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software* 32(6):1–29.

Bergen, Leon, Roger Levy, and Noah Goodman. 2016. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics* 9.

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. *Sinn und Bedeutung* 23:107–124.

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.

Degen, Judith and Judith Tonhauser. 2021. Prior Beliefs Modulate Projection. *Open Mind* 5:59–70.

Djärv, Kajsa. 2019. *Factive and assertive attitude reports*. Ph.D. thesis, University of Pennsylvania.

Djärv, Kajsa and Hezekiah Akiva Bacovcin. 2017. Prosodic effects on factive presupposition projection. *Semantics and Linguistic Theory* 27:116–133.

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. Dinneen, eds., *Syntax and Semantics 11: Presupposition*, pages 57–89. New York: Academic Press.

Gelman, Andrew and Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

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

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. *Chicago Linguistic Society* 51:181–195.

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. *West Coast Conference on Formal Linguistics* 2:114–125.

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

Heycock, Caroline. 2006. Embedded root phenomena. In M. Everaert and H. van Riemsdijk, eds., *The Blackwell Companion to Syntax*, pages 174–209. Wiley.

Hintikka, Jaakko. 1975. Different constructions in terms of the basic epistemological verbs: A survey of some problems and proposals. In J. Hintikka, ed., *The intentions of intentionality and other new models for modalities*, pages 1–25. Dordrecht: Reidel.

Hofmeister, Philip and Ivan A Sag. 2010. Cognitive constraints and island effects. *Language* 86(2):366.

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

Jacobs, Robert a., Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. 1991. Adaptive Mixtures of Local Experts. *Neural Computation* 3(1):79–87.

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

Karttunen, Lauri. 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, Stanley Peters, Annie Zaenen, and Cleo Condoravdi. 2014. The chameleon-like nature of evaluative adjectives. *Empirical Issues in Syntax and Semantics* 10:233–250.

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.

Keller, Frank. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.D. thesis.

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.

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

Lasersohn, Peter. 2009. Relative truth, speaker commitment, and control of implicit arguments. *Synthese*

166:359–374.

Lau, Jey Han, Alexander Clark, and Shalom Lappin. 2014. Measuring gradience in speakers grammaticality judgements. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 36.

Lau, Jey Han, Alexander Clark, and Shalom Lappin. 2017. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive science* 41(5):1202–1241.

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! *North East Linguistic Society* 49:245–255.

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

Mahler, Taylor, Marie-Catherine de Marneffe, and Catherine Lai. 2020. The prosody of presupposition projection in naturally-occurring utterances. *Sinn und Bedeutung* 24:20–37.

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

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

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.

Pinker, Steven. 1994. *The Language Instinct*. New York: HarperPerennial. First published in 1994 by William Morrow and Co. (New York).

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

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

Richardson, Sylvia and Peter J. Green. 1997. On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society* 59(4):731–792.

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.

Scrucca, Luca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal* 8(1):289–317.

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. *Semantics and Linguistic Theory* 11:309–327.

Smith, E. Allyn and Kathleen Currie Hall. 2011. Projection diversity: Experimental evidence. *2011 ESSLLI*

Workshop on Projective Content. pages 156–170.

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

Sorace, Antonella and Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115(11):1497–1524.

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

Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1:123–134.

Sprouse, Jon and Diogo Almeida. 2013. The role of experimental syntax in an integrated cognitive science of language. *The Cambridge handbook of biolinguistics* pages 181–202.

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

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

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

Tonhauser, Judith. 2020. Projection variability in Paraguayan Guaraní. *Natural Language and Linguistic Theory* 38:1263–1302.

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

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

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

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

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

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27(5):1413–1432.

Watanabe, Sumio and Manfred Opper. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11(12):3571–3594.

White, Aaron Steven. 2021. On believing and hoping whether. *Semantics & Pragmatics* 14(6):1–18.

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

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 empir-

ical study. *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 *Are there factive predicates? An empirical investigation*

A 20 complement clauses

The following clauses realized the complements of the predicates in Exps. 1, 2 and 3:

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, that is, if the variance of their response distribution was 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 *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, Gelman and Hill 2006). 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).

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 common practice is to report 95% credible intervals and the posterior probability that the predictor coefficient β is either lower or greater than zero, $P(\beta < 0)$ or $P(\beta > 0)$, 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).

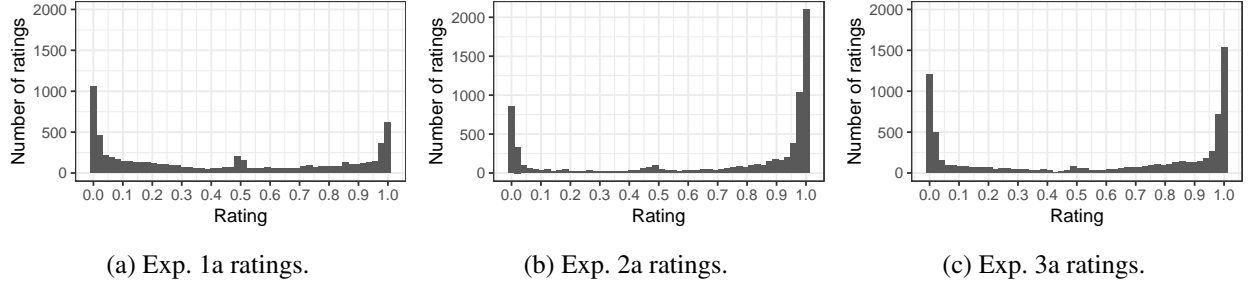(a) Exp. 1a ratings. (b) Exp. 2a ratings. (c) Exp. 3a ratings.

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 endpoints necessarily have a compressed variance; consequently, a condition's mean and its variance are not independent. Beta regression is useful here because it allows for modeling an arbitrarily distributed outcome variable in the [0,1] interval. The Beta distribution is characterized by two parameters, one capturing the mean μ of the distribution and one capturing its precision ϕ, a measure of dispersion. The greater the precision, the more concentrated the values are around the mean, i.e., the lower the variance of the distribution. We follow Smithson and Verkuilen (2006) in modeling μ and ϕ separately for each predictor. That is, we allow each predictor to affect both the mean and the precision of the outcome variable's distribution.

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 certainty 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.

4

Table A2: Maximum a posteriori (MAP) model estimates for certainty 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 inference 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).

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

Table A4: Maximum a posteriori (MAP) model estimates for contradictoriness 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 [2.54; 2.93]	1.46 [1.23; 1.66]	6.17 [5.55; 6.88]
acknowledge	−2.15 [−2.39; −1.90]	−1.04 [−1.29; −0.79]	−4.89 [−5.65; −4.21]
admit	−2.03 [−2.26; −1.77]	−1.05 [−1.30; −0.78]	−4.86 [−5.59; −4.19]
announce	−2.90 [−3.14; −2.64]	−1.56 [−1.80; −1.32]	−6.46 [−7.20; −5.80]
be annoyed	−2.24 [−2.47; −1.98]	−1.28 [−1.53; −1.03]	−5.18 [−5.91; −4.51]
be right	−0.40 [−0.67; −0.12]	*-0.18* [−0.48; 0.12]	−1.90 [−2.70; −1.12]
confess	−2.15 [−2.39; −1.89]	−1.12 [−1.37; −0.86]	−4.85 [−5.59; −4.18]
confirm	−1.74 [−1.99; −1.48]	−0.99 [−1.24; −0.73]	−4.25 [−4.99; −3.57]
demonstrate	−1.94 [−2.18; −1.69]	−1.05 [−1.30; −0.79]	−4.53 [−5.25; −3.84]
discover	−1.63 [−1.90; −1.38]	−1.02 [−1.27; −0.75]	−3.30 [−4.06; −2.61]
establish	−1.94 [−2.19; −1.70]	−1.00 [−1.27; −0.75]	−4.29 [−5.06; −3.62]
hear	−3.72 [−3.97; −3.45]	−1.29 [−1.54; −1.01]	−8.98 [−9.79; −8.27]
inform	−2.78 [−3.03; −2.52]	−1.51 [−1.75; −1.25]	−6.46 [−7.20; −5.81]
know	−1.37 [−1.63; −1.11]	−0.90 [−1.17; −0.64]	−3.64 [−4.39; −2.95]
non-contra. control	−5.26 [−5.54; −4.98]	*-0.24* [−0.51; 0.04]	−11.51 [−12.43; −10.71]
pretend	−3.72 [−3.98; −3.46]	−1.35 [−1.61; −1.10]	−8.97 [−9.78; −8.26]
prove	−1.18 [−1.44; −0.92]	−0.71 [−0.98; −0.45]	−3.36 [−4.10; −2.65]
reveal	−2.27 [−2.52; −2.02]	−1.24 [−1.50; −0.98]	−4.99 [−5.73; −4.32]
say	−3.17 [−3.42; −2.91]	−1.51 [−1.75; −1.25]	−7.11 [−7.87; −6.42]
see	−1.43 [−1.68; −1.18]	−0.82 [−1.08; −0.56]	−3.48 [−4.21; −2.78]
suggest	−3.58 [−3.83; −3.31]	−1.17 [−1.43; −0.92]	−8.92 [−9.71; −8.21]
think	−3.93 [−4.19; −3.66]	−1.20 [−1.47; −0.94]	−9.81 [−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. For instance, in the case of our Exps. 1, 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.
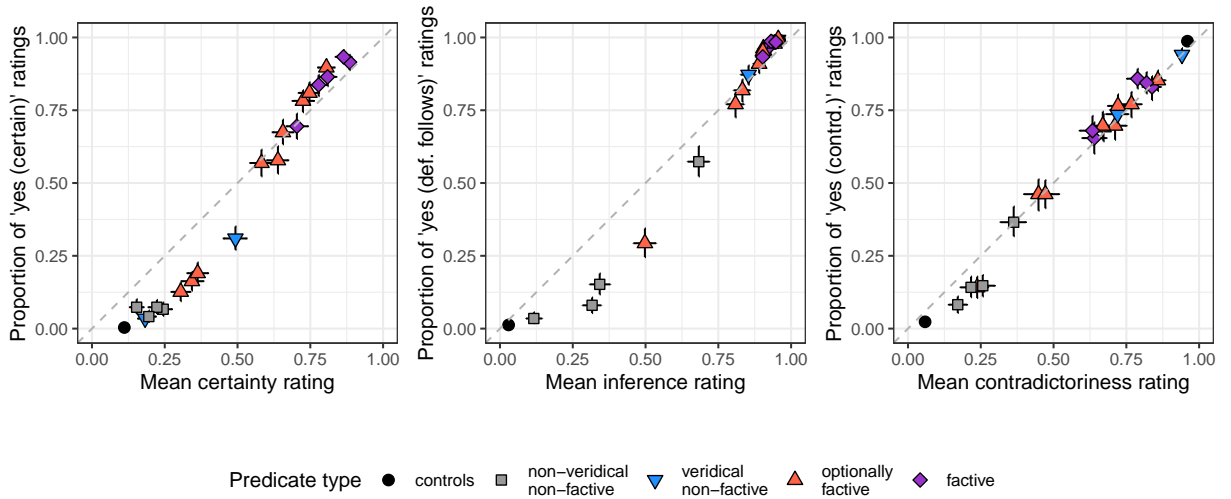

Figure A2: By-predicate proportion of 'yes' responses in two-alternative forced choice task against mean slider ratings in Exps. 1 (left, $r = .98$), Exps. 2 (middle, $r = .99$), and Exps. 3 (right, $r = .99$). Error bars indicate 95% bootstrapped confidence intervals.

F   Mixture models applied to data from Exp. 1a

Figure A3 plots the density of certainty ratings for the 20 predicates from Exp. 1a with the number of Gaussian components overlaid.
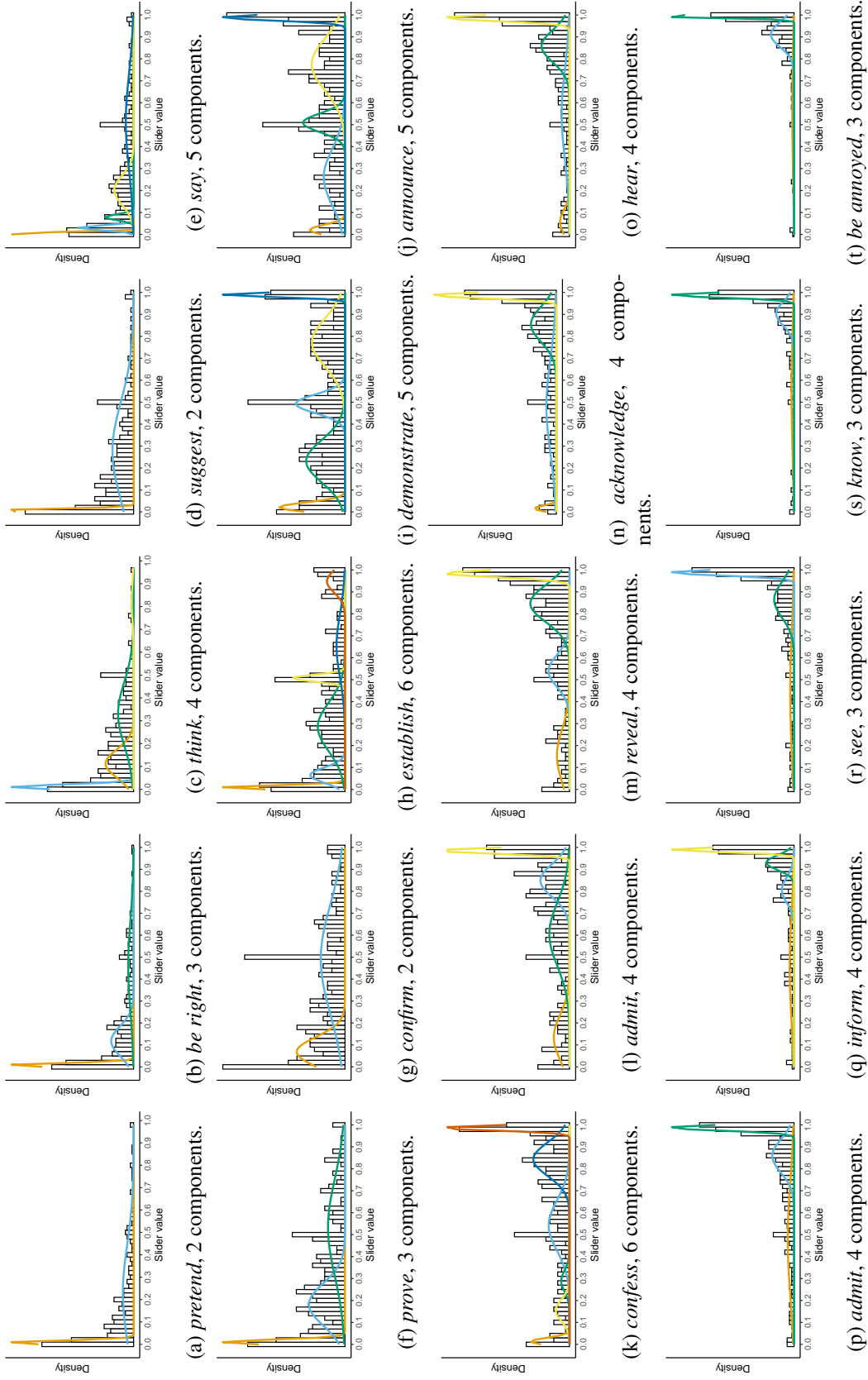
Figure A3: Histograms of certainty ratings from Exp. 1a with overlaid optimal number of Gaussian components.

(a) *pretend*, 2 components.

(b) *be right*, 3 components.

(c) *think*, 4 components.

(d) *suggest*, 2 components.

(e) *say*, 5 components.

(f) *prove*, 3 components.

(g) *confirm*, 2 components.

(h) *establish*, 6 components.

(i) *demonstrate*, 5 components.

(j) *announce*, 5 components.

(k) *confess*, 6 components.

(l) *admit*, 4 components.

(m) *reveal*, 4 components.

(n) *acknowledge*, 4 components.

(o) *hear*, 4 components.

(p) *admit*, 4 components.

(q) *inform*, 4 components.

(r) *see*, 3 components.

(s) *know*, 3 components.

(t) *be annoyed*, 3 components.