

# Which predicates are factive? An empirical investigation

Author(s)

February 18, 2020

## Abstract

Properties of the content of the clausal complement have long been assumed to distinguish factive predicates like *know* from non-factive ones like *think* (Kiparsky and Kiparsky 1970, i.a.). There is, however, disagreement about which properties define factive predicates and uncertainty about whether the content of the complement of particular predicates exhibits the properties attributed to factive predicates. This situation has led to a lack of consensus about which predicates are factive, which is troublesome because the distinction between factive and non-factive predicates has played a central role in linguistic theorizing. This paper presents the findings of experiments designed to investigate critical properties of the content of the complement of clause-embedding predicates, with the goal of understanding how such predicates can be categorized. We find that factive predicates are more heterogeneous than previously assumed and that there is little empirical support for the assumed categorical distinction between factive and non-factive predicates. We discuss implications of our findings for one area where the factive/non-factive distinction has played a central role, namely presuppositions: we propose that projectivity is sensitive to more fine-grained meaning distinctions between clause-embedding predicates.

## 1 Introduction

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

- (1) a. Sam { knows / thinks / acknowledges } that it is raining.
- b. Sam doesn’t { know / think / acknowledge } that it is raining.
- c. Does Sam { know / think / acknowledge } that it is raining?

---

<sup>1</sup>Kiparsky and Kiparsky (1970) also took clause-embedding predicates to differ in their syntactic properties. We ignore syntax in this paper because Kiparsky and Kiparsky (1970:fn.3) already pointed out that syntactic properties do not align with projectivity: *know* and *realize*, for instance, were taken to be syntactically non-factive but semantically factive. For more recent discussion see White and Rawlins 2018 and references therein.

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

## 1.1 Disagreement about the definition of factive predicates

One reason for the lack of consensus about which predicates are factive is disagreement about the definition of factive predicates:<sup>2</sup> whereas Kiparsky and Kiparsky 1970 defined factive predicates as those for which the CC is presupposed (see also, e.g., Karttunen 1971a,b), researchers more recently defined factive predicates as those for which the CC is both presupposed and entailed (e.g., Gazdar 1979a:119-123, Chierchia and McConnell-Ginet 1990:355, van der Sandt 1992:345, Abbott 2006:3, Schlenker 2010:139, Anand and Hacquard 2014:77, Spector and Egré 2015:fn.7).<sup>3</sup> For instance, Beaver (2001:66f.) noted that Gazdar (1979a:119-123) “describes the inferences associated with factive verbs...as being indefeasible in simple affirmative sentences”, that is, “entailments”. Chierchia and McConnell-Ginet (1990:355) illustrated with *realize* that “[a] sentence can both entail and presuppose another sentence”: *Joan realizes that syntax deals with sentence structure* “both entails and presupposes” that syntax deals with sentence structure. And Schlenker (2010:139), in discussing CCs, pointed to “the pattern of inference which is characteristic of presuppositions: an entailment of the positive sentence is preserved under negation and in questions”.

These two definitions of factive predicates come apart on several predicates, including *take into account*, *bear in mind* and *make clear*. These predicates were considered factive in Kiparsky and Kiparsky 1970: thus, if Rahim utters one of the affirmative sentences in (2a) or one of the polar questions in (2b), he presupposes the truth of the CC, that the groom is unreliable. But true affirmative sentences with these

---

<sup>2</sup>Karttunen (1971b:56) argued that the CC of factive predicates is only reliably presupposed when the matrix clause occurs in the indicative mood, as in the examples in (i), or, in subjunctive mood, if the complement is finite, as in (iia). In practice, researchers nowadays classify clause-embedding predicates in the indicative mood and with finite complements.

- (i) Examples adapted from Karttunen 1971b:60
  - a. That his [friend] is not a [vegetarian] bothers Harry.
  - b. His [friend]’s not being a [vegetarian] bothers Harry.
- (ii) Examples adapted from Karttunen 1971b:60f.
  - a. That his [friend] is not a [vegetarian] would bother Harry if he knew about it. (\*Luckily she is a [vegetarian].)
  - b. His [friend]’s not being a [vegetarian] would bother Harry, if he knew about it. (Luckily she is a [vegetarian].)

<sup>3</sup>Another definition is found in Simons 2007:1043: a predicate is factive if and only if a sentence in which the predicate occurs unembedded entails the content of the complement. We ignore this definition because Simons acknowledges that it is not standard; as mentioned below, predicates with entailed CCs are typically called ‘veridical’.

predicates do not entail the CC, as shown by the acceptability of (2c). Thus, these predicates are factive on a definition of factive predicates according to which the CC is presupposed, but not on a definition of factive predicates according to which the CC is both presupposed and entailed.

- (2) a. Rahim: “Joan { took into account / bore in mind / made clear } that the groom is unreliable.”
- b. Rahim: “Did Joan { take into account / bear in mind / make clear } that the groom is unreliable?”
- c. Joan { took into account / bore in mind / made clear } that the groom is unreliable, but she was wrong to do so, because the groom is a very reliable person.

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

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

**Projection.** Under both definitions, the CC of factive predicates is presupposed, with projection taken to diagnose presupposed CCs. It is unclear, however, whether projection leads to a categorical distinction between clause-embedding predicates. As noted above, Kiparsky and Kiparsky (1970) and Karttunen (1971a) assumed that projection results in a three-way categorical distinction of clause-embedding predicates: the CC of factive predicates was assumed to be presupposed, that is, to categorically project, that of non-factive predicates was assumed to be not presupposed, that is, to not project, and that of optionally factive predicates was assumed to be sometimes presupposed: optionally factive predicates “can be used with or without presupposition of the truth of their complement sentence” (Karttunen 1971a:340).<sup>4</sup> But Karttunen (1971b) already challenged this classification when he pointed out that there are factive predicates that “lose their factivity” (p.64) in polar questions and the antecedents of conditionals; he dubbed this subclass of factive predicates ‘semi-factive’. To illustrate, consider the negated examples in (3a), the polar questions in (3b) and the conditionals in (3c). Karttunen (1971b) suggested that “all the examples in [(3a)] presuppose that John had not told the truth” (p.63) and that a speaker who utters (3b) or (3c) with *regret* presupposes the CC, that is, the CC projects over negation, the polar question or the antecedent of the conditional. However, he pointed out, a speaker who utters the variants of (3b) and (3c) with *realize* or *discover* does not necessarily presuppose the CC: the predicates “permit both factive and non-factive interpretations” (p.63), that is, the CC may but need not project. Accordingly, *realize* and *discover*, but also *find out*, *see* and *notice*, were taken to be semi-factive, and predicates like *regret*, *forget*, *resent* and *make clear* were taken to be properly factive.

- (3) Karttunen 1971b:63f.
  - a. John didn’t { regret / realize / discover } that he had not told the truth.
  - b. Did you { regret / realize / discover } that you had not told the truth?
  - c. If I { regret / realize / discover } later that I have not told the truth, I will confess it to everyone.

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

<sup>4</sup>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.

- (4) a. [When did you discover that you liked stormwater?]  
I didn't discover that I liked stormwater, I discovered that I liked fly-fishing for trout.<sup>5</sup>
- b. Perhaps she realized that she was unable to write anything better than "Goblin Market," or perhaps her "failure" to surpass herself is explained by her turn away from poetry to children's stories and religious materials. (Beaver 2010:87)

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

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

In sum, projection of the CC is not categorical for many factive predicates, contrary to what was originally assumed in Kiparsky and Kiparsky 1970.

These observations give rise to the question of whether projection categorically distinguishes factive predicates from optionally factive ones (or from non-factive ones; recall from above that researchers nowadays typically only distinguish factive and non-factive predicates). One possibility is that the CC of factive and optionally factive predicates may project, but that, overall, the CC of factive predicates is more likely to project than that of optionally factive ones; in other words, speakers who utter sentences with factive and optionally factive predicates can presuppose the truth of the CC, but are more likely to presuppose the truth of the CC of factive predicates than of optionally factive ones. Evidence for this position initially comes from Tonhauser, Beaver, and Degen 2018, whose Experiment 1b investigated the projectivity of the CC of 12 clause-embedding predicates in polar questions, including the factive predicates *reveal*, *discover*, *notice* and *be annoyed* and the optionally factive predicates *establish* and *confess*.<sup>6</sup> They found that the CC of all predicates investigated was projective: participants were more likely to take speakers to be committed to the truth of the CC than to the truth of non-projective (main clause) content. They also found that the CC of the factive predicates was more projective than that of the optionally factive predicates. However, there also were significant differences between the factive predicates: the CC of *discover* was less projective than that of *be annoyed*, *notice* and *know*, and the CC of *reveal* was less projective than all other factive predicates. Given that there are significant differences in projectivity not just between factive and optionally factive predicates, but also among factive predicates, it is an open question whether projection categorically distinguishes factive predicates from optionally factive and non-factive ones.

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

<sup>5</sup><http://science.uncvtv.org/content/whats-my-story-water-quality-engineer>

<sup>6</sup>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).

to them, *inform* is not factive. Similarly, Swanson (2012) took the CC of *establish* to be entailed, but the naturally occurring example in (7) again suggests that the CC does not necessarily follow from affirmative sentences with *establish*.

- (6) a. Mary has informed her parents that she's pregnant.  
b. Doctors falsely informed my parents that I was dying.<sup>7</sup>
- (7) Diplomacy was repudiated, but with the media's help it was falsely established that the 'villain' was refusing to negotiate.<sup>8</sup>

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

- (8) a. Mary, who was under the illusion that it was Sunday, was glad that she could stay in bed. (Klein 1975, as cited in Gazdar 1979a:122 and Heim 1992:fn37)  
b. Falsely believing that he had inflicted a fatal wound, Oedipus regretted killing the stranger on the road to Thebes (Klein 1975)  
c. John wrongly believes that Mary got married, and he is enraged that she is no longer single. (adapted from Egré 2008, based on Schlenker 2003b)

The existence of examples like (8) has led some authors, including Klein (1975), Giannakidou (1998), Schlenker (2003a) and Egré (2008), to conclude that emotive predicates are not factive: a speaker who uses one of these predicates merely presupposes that the subject of the attitude believes the truth of the CC (Heim 1992). Consistent with this position is a proposal by Karttunen (2016) according to which the CC is not entailed but taken to follow from utterances of affirmative sentences as long as there is no indication that the beliefs of the speaker/writer and the subject of the attitude are not aligned.<sup>9</sup> Other authors maintained that emotive predicates are factive despite examples like (8) (e.g., Gazdar 1979a, Abrusán 2011, Anand and Hacquard 2014): they hypothesized that examples like those in (8) are licensed by a kind of free indirect discourse, that is, do not show that the CC of such predicates is not entailed (or presupposed).

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

- (9) a. Everyone knew that stress caused ulcers, before two Australian doctors in the early 80s proved that ulcers are actually caused by bacterial infection. (Hazlett 2010:501)  
b. John suffers from paranoia. He falsely believes that the police is spying on him and what is more he knows they are listening to his phone calls. (Abrusán 2011:514)

In short, some clause-embedding predicates may not be factive according to a definition of factive predicates on which the CC is both presupposed and entailed, but they may be according to a definition on which the CC of factive predicates is merely required to be presupposed. There has, to date, not been an investigation of which clause-embedding predicates are such that their CC is entailed.<sup>10</sup>

<sup>7</sup><https://www.quora.com/Has-a-doctor-ever-falsely-informed-you-that-you-re-dying>

<sup>8</sup><https://books.google.de/books?id=BnU1qPRO-34C>, p.66

<sup>9</sup>Karttunen (2016) nevertheless calls emotive predicates factive.

<sup>10</sup>White and Rawlins's 2018 MegaVeridicality data set includes entailment and projection ratings for the CCs of 517 clause-

### 1.3 Goal of this paper and clause-embedding predicates investigated

As illustrated, there is no consensus in the literature about which predicates are factive. This lack of consensus is due, in part, to disagreement about the definition of factive predicates:

(10) Two definitions of factive predicates

A clause-embedding predicate is factive if and only if the content of its clausal complement is

- a. presupposed. (e.g., Kiparsky and Kiparsky 1970, Karttunen 1971a,b)
- b. presupposed and entailed. (e.g., Gazdar 1979a, Chierchia and McConnell-Ginet 1990, van der Sandt 1992, Abbott 2006, Schlenker 2010, Anand and Hacquard 2014, Spector and Egré 2015)

The lack of consensus is compounded by uncertainty about whether projection, which is taken to diagnose whether the CC is presupposed, categorically distinguishes predicates and disagreement about whether the CC of particular predicates is entailed. This paper investigates the projection and entailment of CCs of clause-embedding predicates, with the goal of understanding how such predicates can be categorized as well as assessing the long-standing and widely-adopted classification into factive and non-factive predicates.

To this end, we present the findings of experiments designed to investigate the projection and entailment of the CCs of the 20 clause-embedding predicates that are listed in (11) with the categories they are typically taken to fall into:<sup>11</sup> factive predicates in (11a), non-factive predicates in (11b) and optionally factive predicates in (11c). Specifically, the 5 factive predicates in (11a) include the cognitive predicate *know*, the change-of-states predicates *discover* and *reveal*, the sensory predicate *see*, and the emotive predicate *be annoyed*. The 6 non-factive predicates in (11b) include 4 ‘non-veridical non-factive’ predicates *pretend*, *suggest*, *say* and *think*, whose CC is typically taken to be neither presupposed nor entailed, as well as the 2 ‘veridical non-factive’ predicates *be right* and *demonstrate*, whose CC is typically taken to be entailed but not presupposed.<sup>12</sup> The 9 optionally factive predicates in (11c) include *acknowledge*, *admit* and *announce*, which Kiparsky and Kiparsky 1970 listed as optionally factive, as well as *confirm*, *inform*, *confess*, *establish*, *hear* and *prove*.<sup>13</sup>

(11) 20 clause-embedding predicates

- a. factive: *be annoyed*, *discover*, *know*, *reveal*, *see*
- b. non-factive:
  - i. non-veridical non-factive: *pretend*, *suggest*, *say*, *think*
  - ii. veridical non-factive: *be right*, *demonstrate*

embedding predicates, but the authors did not present a by-predicate analysis. We compare the MegaVeridicality ratings to our data in sections 2.1.2 and 3.6. Djärv et al. (2016) set out to investigate whether emotive and cognitive predicates differ in whether the CC is entailed, but the predicates in their stimuli were realized under an entailment-canceling operator, namely polar questions (e.g., *Is Maria aware that Mike is moving back to Chicago?*). Consequently, the CC is not entailed and observed differences do not directly bear on whether the two types of predicates differ in whether the CC is entailed.

<sup>11</sup>The experiments, data and R code for generating the figures and analyses of the experiments reported on in this paper are available at [redacted for review].

<sup>12</sup>Anand and Hacquard (2014) assumed that *demonstrate* is veridical, in contrast to Anand et al. ms. This latter work also takes *reveal* to be non-factive, in contrast to, for instance, Egré 2008, Wyse 2010 or Tonhauser et al. 2018.

<sup>13</sup>Different categories have been assumed for these 6 predicates. For *confirm* and *inform*, Anand and Hacquard (2014) took them to be optionally factive, but recall from above that Schlenker (2010) took *inform* to be factive. For *prove*, White and Rawlins (2018) suggested that it is optionally factive, but Anand and Hacquard (2014) took it to be non-veridical and non-factive. For *confess*, Swanson (2012) took it to be factive, Karttunen (2016) only took it to commit the speaker to the subject of the attitude being committed to the CC, and Wyse (2010) listed it under the non-factive predicates. For *establish*, Swanson (2012) took it to be non-factive, but Wyse (2010) listed it under the factive predicates. Finally, we also included *hear* in this class: even though it is often considered a factive sensory predicate (e.g., Beaver 2010, Anand and Hacquard 2014), it has been observed that *hear* also has a non-factive reportative evidential sense (see, e.g., Anderson 1986, Simons 2007), especially when it is combined with complements that describe events that cannot be auditorily perceived, as is the case in our experiments.

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

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

## 2 Experiment 1: Projection

Exps. 1a and 1b were designed to explore the projection of the CC of the 20 clause-embedding predicates. The experiments employed the ‘certain that’ diagnostic for projection (see also, e.g., Tonhauser 2016, Djärv and Bacovcin 2017, Stevens et al. 2017, Tonhauser et al. 2018, Mahler 2019, de Marneffe et al. 2019): on this diagnostic, participants are presented with an utterance in which the content to be investigated occurs in an entailment-canceling environment, like a polar question, as illustrated in (12) for the content of the nominal appositive, that Martha’s new car is a BMW.<sup>14</sup>

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

To assess projection, that is, whether the speaker presupposes the truth of the content, participants are asked whether the speaker is certain of the content; for instance, in (12), participants are asked whether Patrick is certain that Martha’s new car is a BMW. If a listener takes the speaker to be certain of the content, we assume that the content projects, that is, the speaker presupposes its truth; if a listener does not take the speaker to be certain of the content, we assume that the content does not project, that is the speaker does not presuppose its truth.<sup>15</sup>

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 (coded as 0) and ‘yes’ at the other (coded as 1). We assume that the higher 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. As discussed in Tonhauser et al. 2018, a second interpretation of gradient certainty ratings is that they reflect a listener’s degree of belief in the speaker’s (binary) commitment to the truth of the content. While we adopt the first interpretation in our discussion, we remain agnostic about the interpretation of gradient certainty ratings: either interpretation gives rise to the expectation on the definition of factive predicates in (10a) that certainty ratings for factive predicates are categorically higher than for optionally factive and non-factive ones.

Exp. 1b was identical to Exp. 1a but employed a two-alternative forced choice (2AFC) task where participants categorically responded ‘yes’ or ‘no’.

<sup>14</sup>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.

<sup>15</sup>We assume that projection is an empirical property of utterance content that can be diagnosed with theoretically untrained native speakers. This position may contrast with that of Anand and Hacquard (2014), who suggested that examples in which the speaker is taken to be committed to the CC of a non-factive predicate merely give rise to the “illusion of projectivity” (p.76).

## 2.1 Experiment 1a: Continuous projection ratings

### 2.1.1 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: 20-71, median: 36; 136 female, 157 male, 2 other, 3 undeclared). They were paid \$0.75 for participating.

**Materials** Exp. 1a investigated the projectivity of the CC of the 20 clause-embedding predicates in (11). The clausal complements of the 20 predicates were realized by 20 clauses (provided in Appendix 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 (13), which realize the complement clause *Julian dances salsa*. Eventive predicates, like *discover* and *hear*, were realized in the past tense and stative predicates, like *know* and *be annoyed*, were realized in the present tense. The direct object of *inform* was realized by the proper name *Sam*. The speaker of the target stimuli was realized by a proper name and displayed in bold-face. The proper names that realized the speakers, the subjects of the clause-embedding predicates and the subjects of the complement clauses were all unique.

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

To assess whether participants were attending to the task, the experiment also included the 6 polar question control stimuli shown in (14). (They were also presented to the participants as utterances by a named speaker.) The content whose projectivity was investigated was the main clause content: for instance, in (14a), participants were asked whether the speaker was certain that Zack is coming to the meeting tomorrow. We expected participants to rate the projectivity of these contents as very low because speakers are typically not taken to be certain about content they are asking about.

- (14) 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?

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

**Procedure** Participants were told to imagine that they are at a party and that, on walking into the kitchen, they overhear somebody ask somebody else a question. Participants were asked to rate whether the speaker was certain of the CC. They gave their responses on a slider marked ‘no’ at one end (coded as 0) and ‘yes’ at the other (coded as 1), as shown in Figure 1.

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



**Helen asks:** "*Did Amanda discover that Danny ate the last cupcake?*"

Is Helen certain that Danny ate the last cupcake?

no  yes

Next

Figure 1: A sample trial in Experiment 1a

**Data exclusion** Prior to analysis, the data from 13 participants who did not self-identify as native speakers of American English were excluded. To assess whether the remaining 287 participants attended to the task, we inspected their responses to the 6 control stimuli, for which we expected low responses. We excluded the data from 16 participants whose response means on the controls were more than 2 standard deviations above the group mean. We furthermore identified 5 participants whose variance in overall response distribution was more than 2 standard deviations below the mean by-participant variance: these participants always selected roughly the same point on the response scale. We excluded the data from these 5 participants, too, leaving data from 266 participants (ages 20-71; median: 36; 118 female, 143 male, 2 other, 3 undeclared).

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

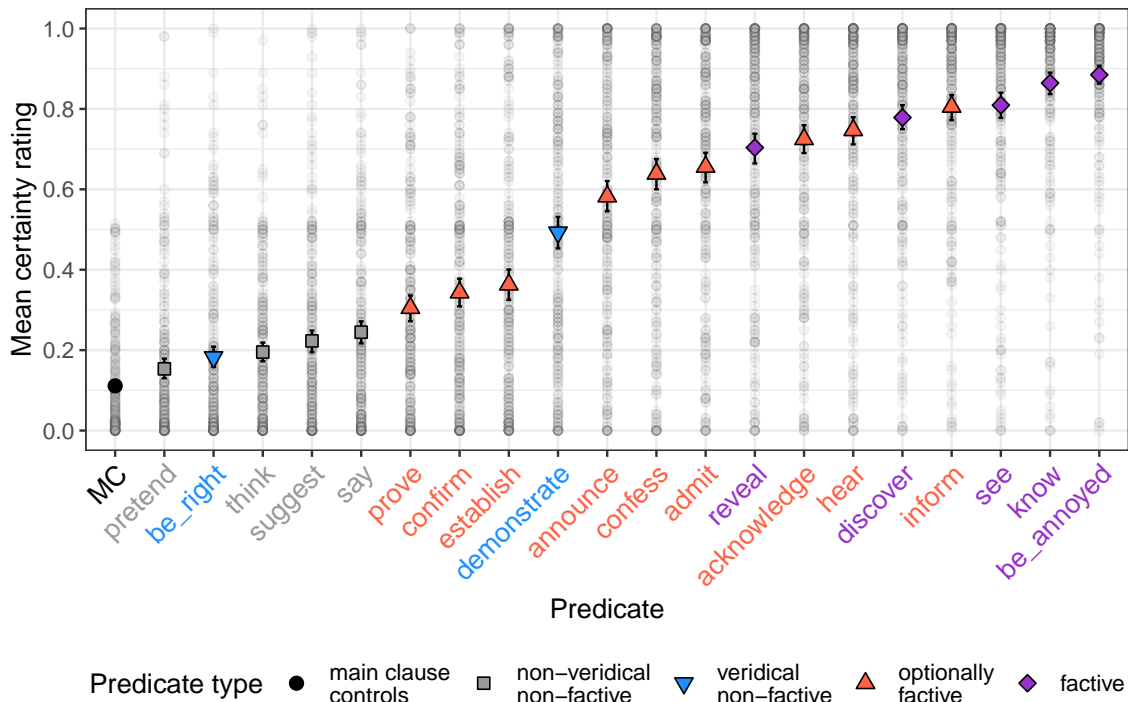


Figure 2: Mean certainty ratings by predicate in Exp. 1a. Error bars indicate 95% bootstrapped confidence intervals. Light gray dots indicate individual participants’ responses.

In view of this finding, one might ask whether the categorization of clause-embedding predicates assumed in (11) is correct: perhaps there is a different place in which a line can be drawn between factive and optionally factive predicates based on the projectivity of the CCs? The mean certainty ratings plotted 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 projectivity 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.<sup>16</sup>

The possibility of categorizing clause-embedding predicates by the projectivity 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

<sup>16</sup>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 projectivity of the contents investigated. We leave this matter to future research.

zero-one-inflated Beta (ZOIB) regression model (Liu and Kong 2015) 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’ as reference level) and included the maximal random effects structure justified by the design: random by-participant and by-content intercepts and random by-participant and by-content slopes for the fixed effect of predicate (capturing random variability in overall projectivity by content and participant as well as random variability in the effect of predicate, which may vary by participant and content).<sup>17</sup>

Mean certainty ratings for all predicates were higher than for the main clause controls [jd: JD use hypothesis() to establish? also: need to adjust if, eg, say doesn’t come out as different. one way of doing this: table of posterior  $p(\text{predicate} > \text{MC})$ , maybe show plots in appendix?], suggesting that the CC of [jd: all?] predicates is at least weakly projective. **DESCRIBE FINDINGS FOR LEAST PROJECTIVE PREDICATES?**

In sum, speakers who utter an interrogative with one of the 20 clause-embedding predicates are taken to be more committed to the CC than to the main clause control content.<sup>18</sup> Thus, to distinguish factive predicates from optionally factive and non-factive ones in this set of 20 predicates, one would need to distinguish one group of projective CCs from another group of projective CCs. The results suggest that there is no principled way of doing so.

#### APPENDIX:

- give more general explanation of why a ZOIB model is appropriate for these data,
- Four chains converged after 2000 iterations each (warmup = 1000,  $\hat{R} = 1$  for all estimated parameters).
- In order to evaluate the evidence for an effect of the at-issueness of the generalization, we report 95% credible intervals and the posterior probability  $P(\beta > 0)$  that the at-issueness coefficient  $\beta$  is greater than zero. A 95% credible interval (CI) demarcates the range of values that comprise 95% of probability mass of our 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  $P(\beta > 0)$  is close to zero or one. Posterior probabilities tell us the probability that the parameter has a certain value, given the data and model (these probabilities are not frequentist  $p$ -values). In order to present statistics as close to widely used frequentist practices, and following Nicenboim and Vasisht 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).
- The model provided evidence for the predicted effect of the at-issueness of the generalization on the at-issueness of the prejacent: the prejacent of EASs for which the truth of the generalizations was more likely to follow from the common ground were more likely to receive a ‘yes’ (at-issue) rating than the prejacent of EAS for which the truth of the generalization was less likely to follow from the common ground (posterior mean  $\beta = 1.29$ , 95% CI=[0.69,1.87],  $P(\beta > 0) = 1$ ). This finding suggests that the prejacent of EASs is more at-issue when the generalization is less at-issue than when the generalization is more at-issue, as predicted by the analysis in section 2.

It is possible that asking participants to assess projection with a gradient response scale encouraged them to provide non-categorical answers and that the observed projection variability is an artifact of the task. If so, categorically distinguishing factive predicates from others may be possible if participants were

<sup>17</sup>For details on the ZOIB model see Appendix B[jd: JD create]

<sup>18</sup>According to Spector and Egré (2015:1739), it is not clear whether the CC of *say* is projective. Our Exp. 1a findings suggest that it is weakly projective.

asked to assess projection with a categorical response task. To assess this possibility, we conducted Exp. 1b, which was identical to Exp. 1a but employed a two-alternative forced choice task.

## 2.2 Experiment 1b: Categorical projection ratings

### 2.2.1 Methods

**Participants** 600 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon’s Mechanical Turk platform. They were paid \$1.00 for participating.<sup>19</sup>

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

Figure 3: A sample trial in Experiment 1b

**Data exclusion** For participants who took the experiment more than once, we only analyzed the data from the first time they took the experiment, leaving data from 525 participants. Prior to analysis, we excluded the data from 43 participants who did not self-identify as native speakers of American English. We also excluded the data from 46 participants who gave a wrong rating (‘yes’) to more than one of the six controls. The data from 436 participants (ages 18-81; median: 37; 207 female, 223 male, 2 other) entered the analysis.

### 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 436 individual participants’ ratings. The order of predicates according to the projectivity 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 Appendix C for a visualization).<sup>20</sup>

<sup>19</sup>Exps. 1a, 2a and 3a (gradient response task) were run in 2017 and 2018, whereas Exps. 1b, 3a and 3b (categorical response task) were run in 2019. Participants in the later experiments were paid more to reflect an increased minimum wage in California.

<sup>20</sup>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 two variables can be described using a monotonic function; if the coefficient is positive, the value of one variable tends to increase with an increase in the other. In the case of our experiments, a coefficient of 1 would mean that there is a perfectly monotone increasing relation between the ranking of the predicates in Exp. 1a and Exp. 1b: for any two predicates  $p_1$  and  $p_2$ , if  $p_1$  ranks below  $p_2$  in Exp. 1a (that is, the mean certainty rating of  $p_1$  is lower than that of  $p_2$ ), then that ranking is preserved in Exp. 1b.

The critical findings of Exp. 1a were replicated. First, the CCs of the factive predicates were not categorically more projective than the CCs of the optionally factive predicates: the CCs of *acknowledge*, *hear* and *inform* were again at least as projective as the CCs of *reveal* or *discover*. Second, a non-arbitrary line between factive and optionally factive predicates cannot be drawn: the CCs of all 20 predicates were projective compared to the non-projective main clause controls [jd: insert model info here] (for details on the model that was fitted, see Appendix C). Thus, like Exp. 1a, the results of Exp. 1b do not support the assumption of the definition of factive predicates in (10a) that factive predicates are categorically distinguished from optionally factive and non-factive ones by the projectivity of the CC.

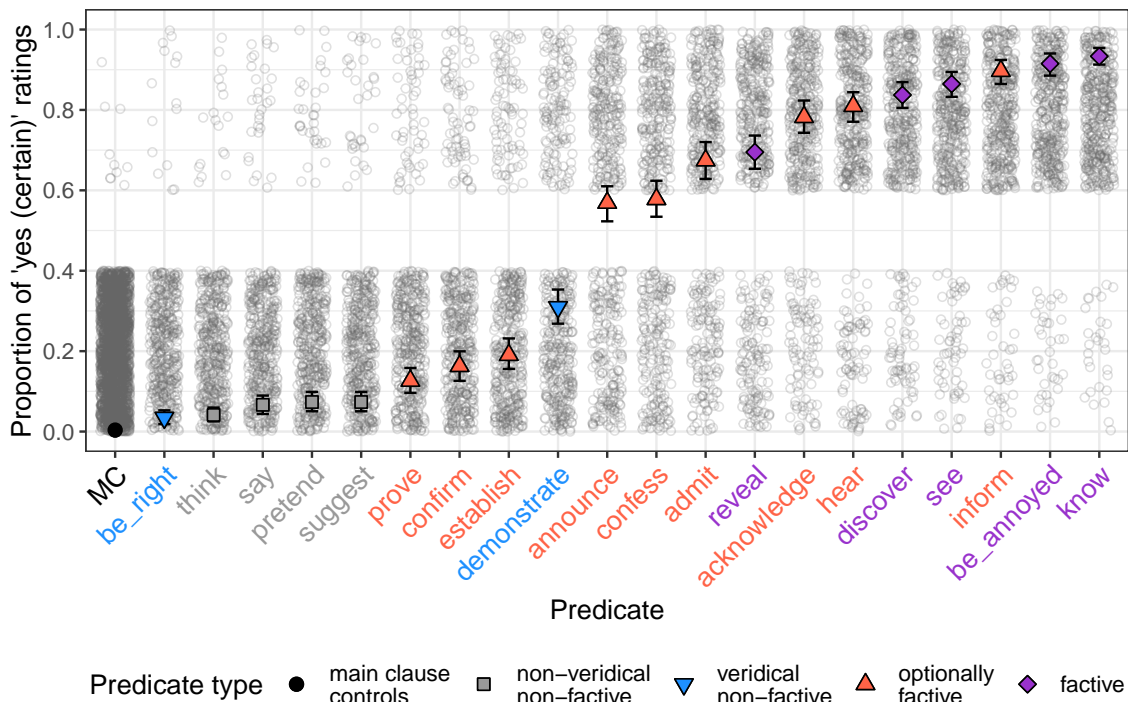


Figure 4: Proportion of ‘yes’ ratings by predicate in Exp. 1b. Error bars indicate 95% bootstrapped confidence intervals.

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

The findings of our Exps. 1a and 1b suggest that the definition of factive predicates in (10a), according to which a predicate is factive if and only if its CC is presupposed, fails to identify a class of factive predicates. Converging evidence comes from two additional datasets: the CommitmentBank (de Marneffe, Simons, and Tonhauser 2019) and the MegaVeridicality dataset (White and Rawlins 2018, White et al. 2018).<sup>21</sup>

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

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

<sup>21</sup>We re-plotted the data, obtained at <https://github.com/mcdm/CommitmentBank> and <http://megaattitude.io>, respectively. For the analysis files, see the aforementioned GitHub repository.

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 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 findings of our Exps. 1, projection of the CC does not categorically distinguish factive predicates from optionally factive and non-factive ones: for instance, the CCs of the optionally factive predicates *foresee*, *confess*, *show*, *tell* and *accept* are at least as projective as those of some factive predicates.

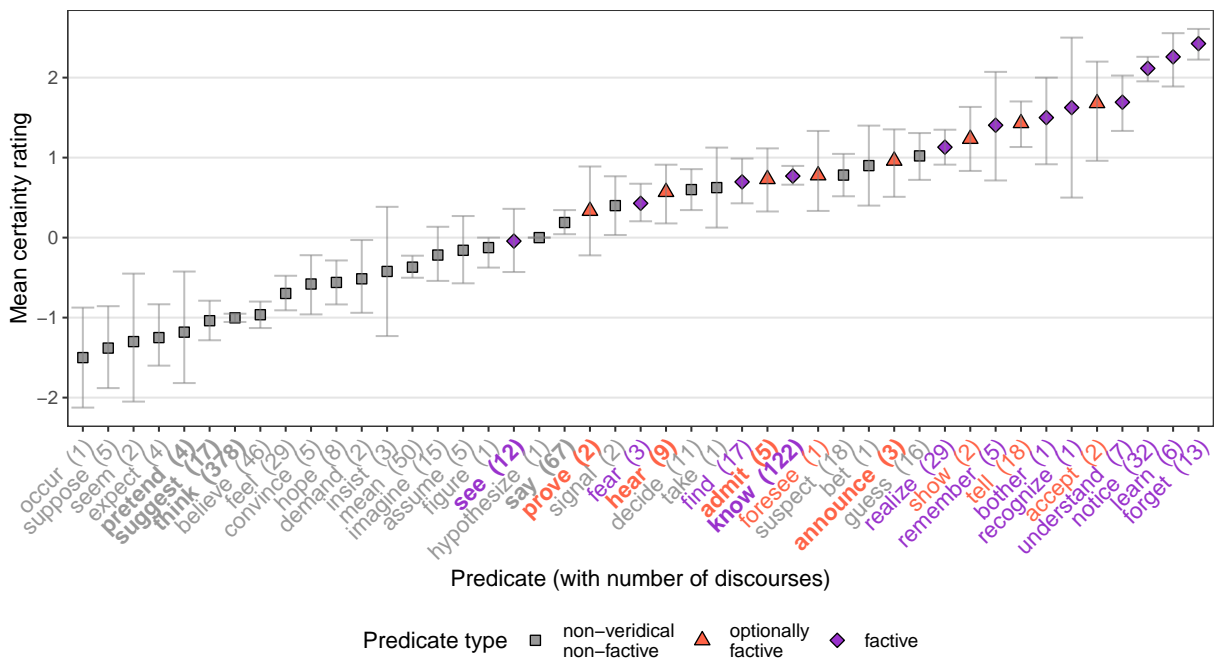


Figure 5: Mean certainty ratings by predicate, with number of discourses in parentheses, for 982 discourses in the CommitmentBank (de Marneffe, Simons, and Tonhauser 2019). Error bars indicate 95% bootstrapped confidence intervals. Predicates included in our Exps. 1 given in bold.

A second 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 low content arguments, as shown in (16) for *know*: the predicates and their clausal complements were embedded under negation in stimuli like (16a) and under negation and the question operator in stimuli like (16b). To assess projection, participants were asked to respond to the question *Did that thing happen?* for stimuli like (16a) and to respond to the question posed by stimuli like (16b). The response options were ‘yes’, ‘maybe or maybe not’ and ‘no’.

- (16) 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?



Each of the 517 predicates in the MegaVeridicality dataset received between 29 and 60 projectivity 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 6 plots the mean projectivity ratings for the 517 predicates in the MegaVeridicality dataset, with labels for the 19 predicates from our experiments that are included in the MegaVeridicality dataset (*be right* is not included). As shown, projection of the CC does not categorically distinguish factive predicates from optionally factive and non-factive ones: for instance, the CCs of the optionally factive predicates *confess*, *acknowledge*, *admit*, *announce*, *hear* and *inform* are at least as projective as those of some factive predicates.

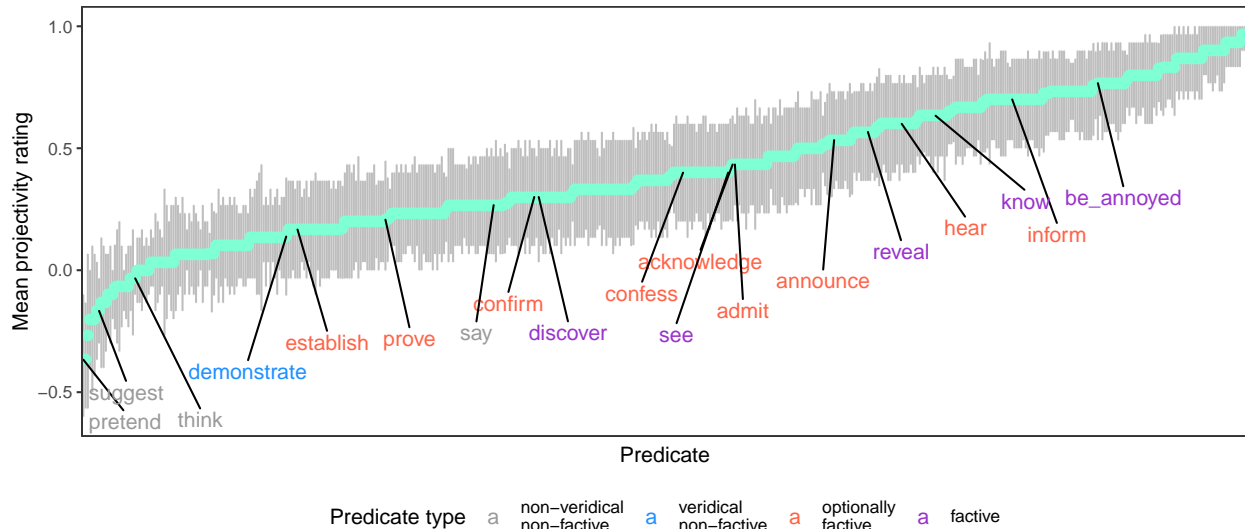


Figure 6: Mean projectivity rating by predicate in light green, with 95% bootstrapped confidence intervals, for the 517 predicates in the MegaVeridicality dataset (White and Rawlins 2018, White et al. 2018), with labels for the 19 predicates featured in our experiments.

In sum, contrary to what is expected under the definition of factive predicates in (10a), the findings of our Exps. 1a and 1b as well as those of the CommitmentBank and the MegaVeridicality datasets suggest that projection of the CC does not categorically distinguish factive predicates from optionally factive and non-factive ones. The challenge to the definition in (10a) is compounded by the diversity of the empirical evidence: the evidence comes from constructed and naturally occurring examples, from examples presented in a minimal context to ones presented in context, from examples with diverse and minimal lexical content, and from projection ratings collected through different response tasks. We therefore conclude that the definition of factive predicates in (10a), according to which a predicate is factive if and only if the content of the complement is presupposed, fails to identify a class of factive predicates.

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

### 3 Experiments 2 and 3: Entailment

Experiments 2 and 3 explored which of the CCs of the 20 clause-embedding predicates are entailed. We assume the standard definition of entailment, according to which entailment is a binary, categorical relation between two contents: given sentences  $\phi$  and  $\psi$ , the content of  $\phi$  entails the content of  $\psi$  if and only if every world in which the content of  $\phi$  is true is also a world in which the content of  $\psi$  is true. Thus, if there is even one world in which  $\phi$  is true but  $\psi$  is false, then the content of  $\phi$  does not entail the content of  $\psi$ . Given this definition, establishing that the contents of two natural language sentences  $\phi$  and  $\psi$  do not stand in an entailment relation involves identifying a world in which the content of  $\phi$  is true and the content of  $\psi$  is false. **[jd: the following sentence is rhetorically a little weird: aren't we precisely interested in identifying when there is an entailment relationship? the way it's framed makes it sound like we're mostly interested in showing non-entailment. – more generally, do we need all the detail of this paragraph or can we just move directly to the next paragraph, which introduces the standard diagnostics for entailment?]** In this paper, however, we are also interested in establishing that the contents of two sentences  $\phi$  and  $\psi$  do stand in an entailment relation. Establishing this is straightforward if the content of some expression in  $\phi$  denotes a subset of the content of some expression in  $\psi$ : for instance, in (17a), the content of  $\phi$  entails the content of  $\psi$  because the denotation of *black cat* in  $\phi$  is a subset of the denotation of *cat* in  $\psi$ . No such structural relation exists, however, between the content of a sentence with a clause-embedding predicate and the CC, as in (17b).

- (17) a.  $\phi$ : Sam owns a black cat.                       $\psi$ : Sam owns a cat.  
       b.  $\phi$ : Sam knows that it's raining.               $\psi$ : It's raining.

To establish whether the content of an unembedded matrix sentence with a clause-embedding predicate entails the CC, we use standard diagnostics that are taken to provide evidence for an entailment relation. Two standard diagnostics are given in (18). According to the 'inference diagnostic' in (18a), the content of an unembedded matrix sentence with a clause-embedding predicate entails the CC if and only if the CC definitely follows from a true statement of the unembedded matrix sentence. According to the 'contradictoriness diagnostic' in (18b), an entailment relationship holds if and only if an utterance of the unembedded matrix sentence that is followed by a denial of the CC is contradictory.

- (18) Two diagnostics for entailment                      (see, e.g., Chierchia and McConnell-Ginet 1990:§3.1)  
       a. Inference diagnostic (Exps. 2a and 2b)  
            $\phi$  entails  $\psi$  if and only if, if  $\phi$  is true, then the truth of  $\psi$  definitely follows.  
       b. Contradictoriness diagnostic (Exps. 3a and 3b)  
            $\phi$  entails  $\psi$  if and only if sentences of the form  $\phi$  *and not*  $\psi$  are contradictory.

These two diagnostics for entailment were applied to the CCs of the 20 clause-embedding predicates in Exps. 2a/2b and Exps. 3a/3b, respectively. The (a) and (b) versions of the experiments differed in whether participants provided a continuous response (a) or a categorical response (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.

We present the findings of the two experiments in sections 3.1 and 3.4, and then discuss the findings jointly in section 3.5 **[jd: revise after restructuring]**: given the standard classification of the 20



clause-embedding predicates in (11), we expect the CCs of factive and veridical non-factive predicates to be entailed.

### 3.1 Experiment 2a: Continuous entailment ratings using inference diagnostic

Exp. 2a was designed to investigate which of the CCs of the 20 clause-embedding predicates are entailed based on the inference diagnostic for entailment in (18a). In contrast to the entailment relationship, which is binary and categorical, the inference relationship between the contents of two sentences  $\phi$  and  $\psi$  is gradient: given a true statement  $\phi$ , the truth of  $\psi$  follows to varying degrees. To illustrate, assume that Chris wears his pajamas when he is at home and that he doesn't wear pajamas outside of his home. Now consider the sentences  $\phi_1$  to  $\phi_5$  in (19). Given the aforementioned assumption, the truth of the content of the statement  $\psi$  *Chris is wearing his pajamas* definitely follows from a true statement of  $\phi_1$  [jd: **is this true? – the way the assumption is formulated above, it's a habitual – ie, he habitually wears pajamas at home and not outside the home; but presumably, there is a point in time at which he changes from pajamas to outside clothes, so he doesn't wear just pajamas at home, so it doesn't follow from  $\phi_5$  that Chris is wearing pajamas**], it follows to increasingly lower degrees from true statements of  $\phi_2$  to  $\phi_4$  and it does not follow from a true statement of  $\phi_5$ .

- (19)  $\phi_1$  Chris is at home.  
 $\phi_2$  Chris is likely at home.  
 $\phi_3$  Chris is possibly at home.  
 $\phi_4$  There is a slight possibility that Chris is at home.  
 $\phi_5$  Chris is not at home.

Exp. 2a investigated the strength of the inference to the CCs from true statements of unembedded matrix sentences with the 20 clause-embedding predicates by collecting gradient inference ratings. [jd: **i thought this intro paragraph would tell the reader how the inference diagnostic will be applied to diagnose entailment, but i'm not sure it does that. it introduces notions of gradience but without telling us what for, eg: " In contrast to the entailment relationship, which is binary and categorical, the inference relationship between the contents of two sentences  $\phi$  and  $\psi$  is gradient" – what is the purpose of this? readers are trying to figure out why we're using the inference diagnostic for entailment, but it sounds here like you're telling us that the inference diagnostic doesn't actually capture entailment, so why do this experiment? might be solved by getting rid of the gradience prose? if the purpose of introducing this gradience is to motivate the continuous response task, i think that's better handled in the methods section (or can be discussed later)**]

#### 3.1.1 Methods

**Participants** 300 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon's Mechanical Turk platform (ages: 19-69, median: 36; 152 female, 148 male). They were paid \$0.75 for participating.

**Materials** The target stimuli were 400 unembedded matrix sentences that were constructed by combining the 400 predicate/clause combinations of Exps. 1 with a random 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 (20), 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 (20a) whether it follows that Danny ate the last cupcake and about (20b) whether it follows that Emma studied on Saturday morning.

- (20) 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. For the four entailing control stimuli in (21), the tested inferences are lexical entailment of the true statements: for instance, that Frederick solved the problem is a lexical entailment of the true statement in (21a) 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 in (22) definitely do not follow from these true statements: for instance, in (22a), 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.

(21) Entailing control stimuli

- a. **What is true:** Frederick managed to solve the problem. (Tested inference: Frederick solved the problem.)
- b. **What is true:** Zack bought himself a car this morning. (Tested inference: Zack owns a car.)
- c. **What is true:** Tara broke the window with a bat. (Tested inference: The window broke.)
- d. **What is true:** Vanessa happened to look into the mirror. (Tested inference: Vanessa looked into the mirror.)

(22) Non-entailing control stimuli

- a. **What is true:** Dana watched a movie last night. (Tested inference: Dana wears a wig.)
- b. **What is true:** Hendrick is renting an apartment. (Tested inference: The apartment has a balcony.)
- c. **What is true:** Madison was unsuccessful in closing the window. (Tested inference: Madison closed the window.)
- d. **What is true:** Sebastian failed the exam. (Tested inference: Sebastian did really well on the exam.)

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 7.

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

Does it follow that Grace visited her sister?

**definitely doesn't follow****definitely follows**

Figure 7: A sample trial in Experiment 2a

To familiarize participants with the task, they first responded to the two familiarization stimuli in (23), where the inference tested was the inference to the CC. Participants who rated (23a) in the lower half of the sliding scale or (23b) 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 (23a) and in the lower half for (23b).

- (23) a. **What is true:** Drew is correct that Patty lives in Canada.  
 b. **What is true:** Drew believes that Patty lives in Canada.

After responding to the 28 stimuli, participants filled out a short, optional survey about their age, their gender, their native language(s) and, if English is their native language, whether they are a speaker of American English (as opposed to, e.g., Australian or Indian English). To encourage them to respond truthfully, participants were told that they would be paid no matter what answers they gave in the survey.

**Data exclusion** Prior to analysis, the data from 14 participants who did not self-identify as native speakers of American English were excluded. For the remaining 286 participants, we inspected their responses to the 8 control stimuli. The group mean for the four control stimuli in (21), for which the strength of the inference was expected to be at ceiling, was .94. The group mean for the four control stimuli in (22), for which the strength of the inference was expected to be at floor, was .05. These group means suggest that, overall, participants attended to and understood the task. Importantly, the at-ceiling mean rating for the control stimuli with indisputably entailed content shows that the diagnostic can detect entailed content. The response means of 27 participants were more than 2 standard deviations below the group mean for the control stimuli in (21) or above the group mean for the control stimuli in (22). Closer inspection revealed that these participants' responses to the control stimuli were systematically higher or lower, respectively, suggesting that these participants did not attend to the task or interpreted the task differently. The data from these 27 participants were also excluded. In this experiment, we did not identify any participants who always selected roughly the same point on the response scale. The results presented below are based on data from 259 participants (ages 19-69; median: 36; 132 female, 128 male).

### 3.1.2 Results

Figure 8 plots the mean inference ratings for the target stimuli by predicate, in increasing order from left to right, as well as for the non-entailing controls (abbreviated 'non-ent. C') and the entailing controls (abbreviated 'entailing C'). In line with impressionistic judgments reported in the literature, the mean inference

ratings for most factive predicates as well as the veridical non-factive predicate *be right* are 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* are lower, suggesting comparatively weaker inferences to the truth of the CC.

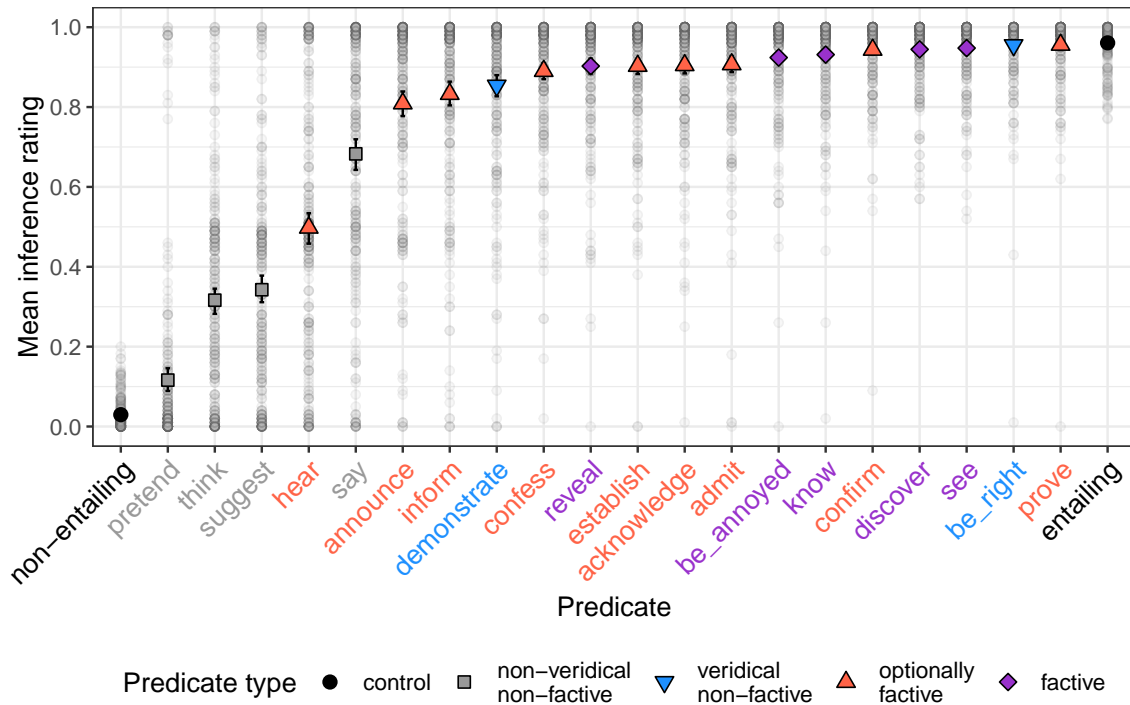


Figure 8: Mean inference rating by predicate in Exp. 2a, including the non-entailing controls (‘non-ent. C’) and the entailing controls (‘entailing C’). Error bars indicate bootstrapped 95% confidence intervals. Light gray dots indicate individual participants’ ratings.

[jd: JD Describe model and update the following if necessary] In other words, only the inference ratings to the CC of *prove* are statistically indistinguishable from those of the controls with indisputably entailed content, suggesting that the CCs of the remaining 19 predicates are not entailed.

### 3.1.3 Discussion

[jd: move up a lot of the discussion from what’s currently sec 3.5, segue to categorical inference diagnostic]

## 3.2 Experiment 2b: Categorical entailment ratings using inference diagnostic

### 3.2.1 Methods

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

**Materials and procedure** The materials and procedure were identical to those of Exp. 2a, with the exception of the task: participants responded ‘yes’ or ‘no’ to the question of whether the CC follows from the true statement in which it was embedded, as shown in Figure 9.

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

Does it follow that Zoe calculated the tip?

☐ No
 ☐ Yes

Next

Figure 9: A sample trial in Experiment 2b

**Data exclusion** For participants who took the experiment more than once, we only analyzed the data from the first time they took the experiment, leaving data from 431 participants. Prior to analysis, we excluded the data from 35 participants who did not self-identify as native speakers of American English. We also excluded the data from 25 participants who 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. The data from 375 participants (ages 18-73; median: 38; 187 female, 187 male) entered the analysis below.

### 3.2.2 Results and discussion

Figure 10 shows the proportion of ‘yes’ ratings on target trials by predicate, in increasing order from left to right, as well as on non-entailing control trials (abbreviated ‘non-ent. C’) and entailing control trials (abbreviated ‘entailing C’). The findings of Exp. 2b are strikingly similar to those of Exp. 2a: the proportions of ‘yes’ ratings are quite high for most factive predicates as well as the veridical non-factive predicate *be right* and the optionally factive predicate *prove*, but the proportions for the purportedly factive predicate *reveal* and the purportedly veridical non-factive predicate *demonstrate* are lower. The order of predicates according to the categorical inference measure in Exp. 2b is almost identical to that obtained via the continuous inference measure in Exp. 2a, as evidenced by the very high Spearman rank correlation of .989 (see Appendix C for a visualization).

The critical findings of Exp. 2a were replicated. [jd: JD insert model info here once issue of exclusions settled]

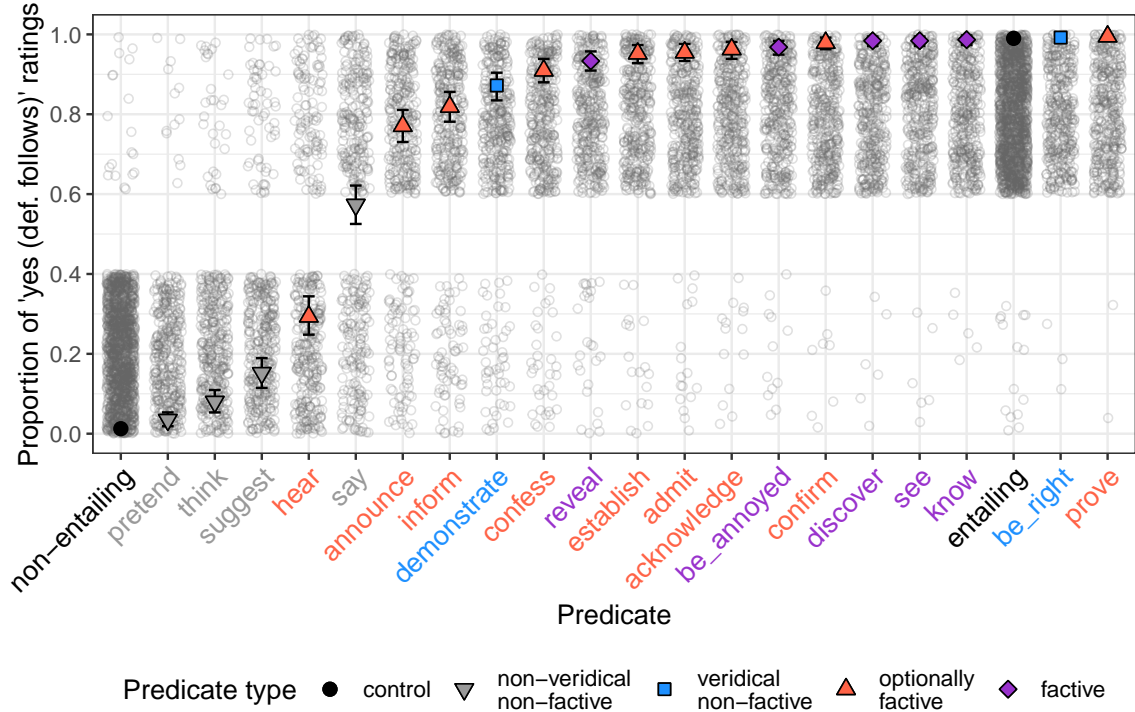


Figure 10: Proportion of ‘yes’ ratings by predicate in Exp. 2b. Error bars indicate 95% bootstrapped confidence intervals. Jittered light gray dots indicate individual participants’ ratings of ‘yes’ (coded as 1) and ‘no’ (coded as 0).

[jd: segue to exps 3]

### 3.3 Experiment 3a: Continuous entailment ratings using contradictoriness diagnostic

This experiment investigated which of the CCs of the 20 clause-embedding predicates are entailed based on the contradictoriness diagnostic for entailment in (18b). Participants rated the contradictoriness of utterances of English sentences of the form  $\phi$  but not  $\psi$ , where  $\phi$  is an unembedded matrix sentence with a clause-embedding predicate and  $\psi$  is its clausal complement, as in (24):

(24) Mary announced that she is pregnant, but she’s not.

Contextual factors may influence contradictoriness ratings: for instance, how contradictory an utterance of (24) is assessed to be may depend on how old or how trustworthy Mary is (see, e.g., Schlenker 2010, de Marneffe et al. 2012). To allow participants’ ratings to reflect degrees of contradictoriness, we collected gradient contradictoriness ratings. We assume that when the content of  $\phi$  entails the content of  $\psi$ , the contradictoriness of an utterance of the form  $\phi$  but not  $\psi$  is not mitigated by contextual factors, and participants’ contradictoriness ratings are at ceiling, that is, indistinguishable from contradictory control stimuli.

#### 3.3.1 Methods

**Participants** 300 participants with U.S. IP addresses and at least 99% of previous HITs approved were recruited on Amazon’s Mechanical Turk platform (ages: 18-72, median: 35; 137 female, 162 male, 1 other). They were paid \$0.75 for participating.

**Materials** On target trials, the same 400 predicate/clause combinations were combined with a random proper name subject and a *but*-clause that denied the truth of the CC. As shown in (25), stimuli were presented to participants as utterances by named speakers. The proper names that realized the speakers, the subjects of the 20 predicates and the subjects of the complement clauses were all unique. The gender of the proper name subject of the predicate was distinct from the gender of the proper name in the complement clause, to ensure that the pronoun in the elliptical *but*-clause unambiguously referred to the individual referred to in the complement clause of the predicate.

- (25) 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. For the four control stimuli in (26), 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 (26a), 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 rely on the contradictoriness ratings for these control stimuli to identify entailments. By contrast, we expect the contradictoriness ratings for the four non-contradictory control stimuli in (27) to be at floor, because the content of the second clause does not contradict the content of the first clause: for instance, in (27a), the content that Vanessa is good at math does not entail the negation of the content that the speaker is not good at math.

- (26) Contradictory control stimuli
- a. Madison laughed loudly and she didn’t laugh.
  - b. Dana has never smoked in her life and she stopped smoking recently.
  - c. Hendrick’s car is completely red and his car is not red.
  - d. Sebastian lives in the USA and has never been to the USA.
- (27) Non-contradictory control stimuli
- a. Vanessa is really good at math, but I’m not.
  - b. Zack believes that I’m married, but I’m actually single.
  - c. Tara wants me to cook for her and I’m a terrific cook.
  - d. Frederick is both smarter and taller than I am.

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 made by a speaker and were asked to assess whether the speaker’s utterance is contradictory. On each trial, participants read the speaker’s utterance and then gave their response on a slider marked ‘definitely no’ at one end (coded as 0) and ‘definitely yes’ at the other (coded as 1), as shown in Figure 11.

To familiarize participants with the task, they first responded to the two familiarization stimuli in (28): participants who rated (28a) in the lower half of the scale or (28b) in the upper half of the 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 training stimuli, that is, a rating in the upper half of the scale for (28b) and a rating in the lower half for (28b).

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

Is Margaret's utterance contradictory?

**definitely no****definitely yes**

Figure 11: A sample trial in Experiment 3a

- (28) a. Drew is aware that Patty lives in Canada, but she doesn't.  
 b. Drew thinks that Patty lives in Canada, but she doesn't.

After responding to the 28 stimuli, participants filled out a short, optional survey about their age, their gender, their native language(s) and, if English is their native language, whether they are a speaker of American English (as opposed to, e.g., Australian or Indian English). To encourage them to respond truthfully, participants were told that they would be paid no matter what answers they gave in the survey.

**Data exclusion** Prior to analysis, the data from 19 participants who did not self-identify as native speakers of American English were excluded. For the remaining 281 participants, we inspected their responses to the 8 control stimuli: as expected, the group mean rating for the non-contradictory control stimuli in (27) was at floor (.08)<sup>22</sup> and the group mean rating for the contradictory control stimuli in (26) was at ceiling (.94). This finding shows that, overall, participants attended to and understood the task. Importantly, the at-ceiling mean rating for the control stimuli with indisputably entailed content shows that the diagnostic can detect entailed content. The mean ratings of 18 participants were more than 2 standard deviations above the group mean for the non-contradictory control stimuli or below the group mean for the contradictory control stimuli. Closer inspection revealed that these participants' responses to the control stimuli were systematically higher or lower, suggesting that they did not attend to the task or interpreted the task differently. The data from these 18 participants were also excluded. In this experiment, we did not identify any participants who always selected roughly the same point on the response scale. The results presented below are based on data from 263 participants (ages 18-72; median: 36; 126 female, 136 male, 1 other).

### 3.3.2 Results

Figure 12 plots 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 are among the highest, and those of the non-factive predicates are among the lowest. However, except possibly for *be right*, none of the mean contradictoriness ratings are at ceiling.

<sup>22</sup>The mean contradictoriness rating of the non-contradictory control stimulus that was of the same form as the target stimuli, namely (27b) *Zack believes that I'm married, but I'm actually single*, was higher, at .17, than the mean ratings of the remaining three non-contradictory control stimuli (means: .05 or .06) and identical to that of the target stimuli with *think*.



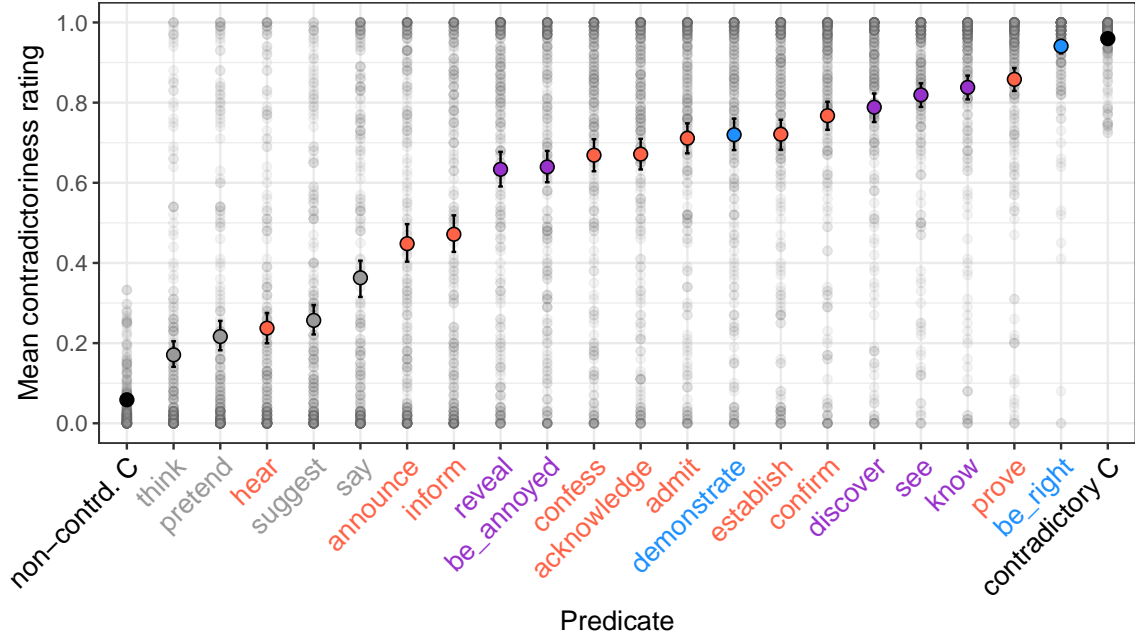


Figure 12: Mean contradictoriness rating by predicate in Exp. 3a, including the non-contradictory controls (‘non-contrd. C’) and the contradictory controls (‘contradictory C’). Error bars indicate bootstrapped 95% confidence intervals. Light gray dots indicate individual participants’ ratings. [jd: add color/shape legend to plot]

[jd: JD Describe models] In other words, none of the contradictoriness ratings for the CCs are statistically indistinguishable from those of the controls with indisputably entailed content, suggesting that the CCs of all 20 predicates are not entailed.

### 3.4 Experiment 3b: Categorical entailment ratings using contradictoriness diagnostic

#### 3.4.1 Methods

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

**Materials and procedure** The materials and procedure were identical to those of Exp. 3a, with the exception of the task: participants responded ‘yes’ or ‘no’ to the question of whether the speaker’s utterance is contradictory, as shown in Figure 13.

**Data exclusion** For participants who took the experiment more than once, we only analyzed the data from the first time they took the experiment, leaving data from 430 participants. Prior to analysis, we excluded the data from 30 participants who did not self-identify as native speakers of American English. We also excluded the data from 47 participants who gave more than one wrong rating to one of the eight control sentences, where a wrong ratings was a ‘yes’ to a non-contradictory control and a ‘no’ to a contradictory one. The data from 353 participants (ages 18-73; median: 37; 180 female, 173 male) entered the analysis below.

**Margaret:** "Edward informed Sam that Sophia got a tattoo, but she didn't."

Is Margaret's utterance contradictory?

☐ No
 ☐ Yes

Next

Figure 13: A sample trial in Experiment 3b

### 3.4.2 Results

Figure 14 shows the proportion of ‘yes’ ratings on target and control trials. The findings of Exp. 3b are again very similar to those of Exp. 3a, with a Spearman rank correlation of .985 (see Appendix C for a visualization).

The critical findings of Exp. 3a were replicated. [jd: JD insert model info here once issue of exclusions settled]

## 3.5 Discussion

[jd: much of this discussion still needs to be moved up into the discussion of exp 2a and the results/discussion sections of exps 3a, 3a, 3b]

In Exps. 2 and 3, we applied two standard diagnostics for entailment to the CCs of 20 clause-embedding predicates, with the goal of identifying predicates whose CC is entailed content. To assess whether the CC of a particular predicate is entailed, we compared the inference ratings (Exps. 2) and the contradictoriness ratings (Exps. 3) for the CCs to the ratings for control stimuli with indisputably entailed content: we assumed that the CC of a predicate is entailed if the ratings for that CC are statistically indistinguishable from these control stimuli, and not entailed if the ratings for that CC are significantly lower. [jd: Need to update this: Given these assumptions, we found that the CC of *prove* is entailed according to the inference diagnostic and that the CC of none of the 20 predicates is entailed according to the contradictoriness diagnostic.]

Our way of identifying entailed CCs, namely by comparing them to indisputably 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 or contradictoriness ratings are at least as high as those of *demonstrate*, that is .85 or .72, respectively: this would mean that the CCs of all of the predicates typically taken to be factive or veridical non-factive are entailed, as well as the CCs of some optionally factive predicates. There is, however, no principled reason why the ratings for the CC of *demonstrate* should be the threshold for entailment. One could also pick an arbitrary mean inference or mean contradictoriness rating, say .8, and categorize CCs with mean ratings of at least .8 as entailed, and CCs with mean ratings below .8 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 .85 or .75, with the result that the CCs of different predicates would be considered entailed: for instance, on the contradictoriness diagnostic, the CC of *discover* (mean: .79) would not count as entailed with a threshold of .8, but would with one of .75, and the CC of *know* (mean: .84) would count as entailed with a threshold of .8, but not with one of .85. 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 indisputably entailed content.

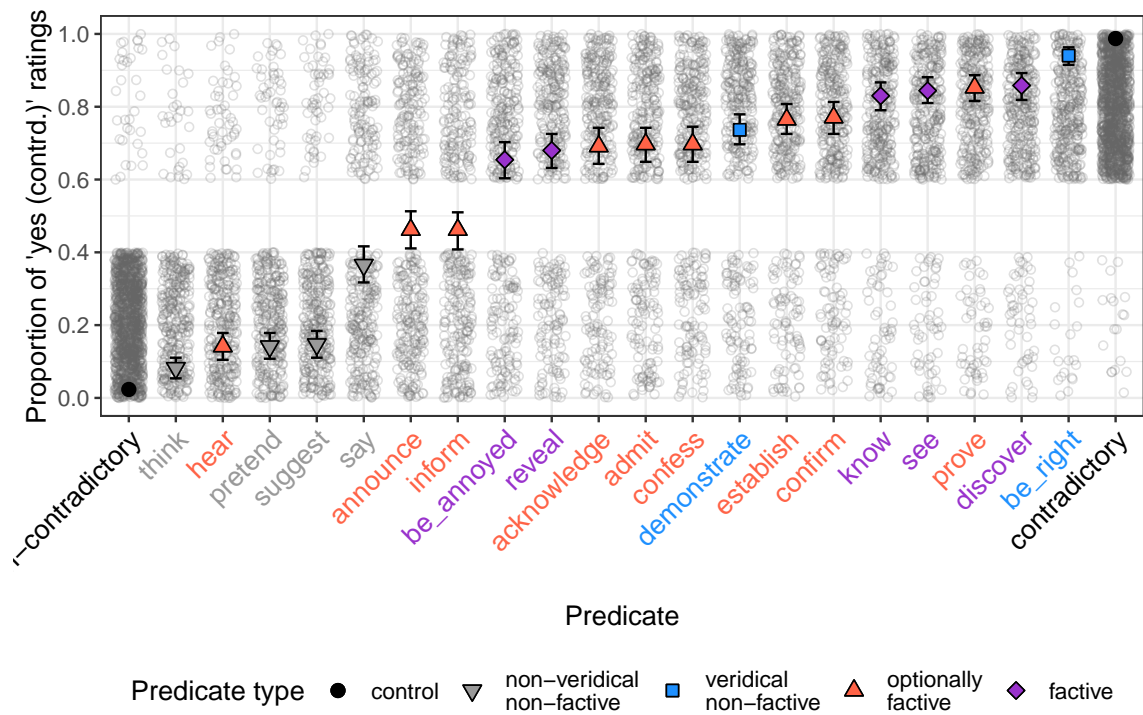


Figure 14: Proportion of 'yes' ratings by predicate in Exp. 3b. Error bars indicate 95% bootstrapped confidence intervals. Jittered light gray dots indicate individual participants' ratings of 'yes' (coded as 1) and 'no' (coded as 0).

One might also wonder whether a different entailment diagnostic might result in the CCs of more clause-embedding predicates being classified as entailed. For instance, one might wonder whether the gradient response scale on which participants responded in Exps. 2 contributed to so few CCs being classified as entailed and whether the finding would be different if participants had been asked to decide in a two-alternative forced choice task whether the content follows from a true statement, on the inference diagnostic, or whether the speaker’s utterance is contradictory, on the contradictoriness diagnostic. To assess this possibility, we ran two follow-up experiments that were identical to Exps. 2 except that we used a two-alternative forced choice task, with ‘yes’ (coded as 1) or ‘no’ (coded as 0) as the response options (see Appendix C for details). We refer to these two experiments as Exp. 2a’ and Exp. 3a’. The two plots in Figure ?? show the proportion of ‘yes’ ratings on the two diagnostics by predicate, again in increasing order from left to right, as well as the ratings for two types of control stimuli. Jittered light gray dots indicate individual participants’ ratings. The findings of Exps. 2’ are strikingly similar to those of Exps. 2: on the inference diagnostic (panel a), the proportions of ‘yes’ ratings are quite high for most factive predicates as well as the veridical non-factive predicate *be right*, but the proportions for the purportedly factive predicate *reveal* and the purportedly veridical non-factive predicate *demonstrate* are lower; on the contradictoriness diagnostic (panel b), the proportion of ‘yes’ ratings for *be right* are quite high, but those of the factive predicates and of the other veridical non-factive predicate are not. The Spearman rank correlations for the two pairs of experiments are very high, providing further evidence for the similarities between the findings: .996 for the two experiments that used the inference diagnostic (Exps. 2a and 2a’) and .985 for the two experiments that used the contradictoriness diagnostic (Exps. 3a and 3b); see Appendix C for figures.

To identify predicates whose CCs are entailed on Exps. 2’, we ideally want to compare the ratings for the CCs to those of the controls with indisputably entailed content, as in Exps. 2. However, in contrast to the CCs of the 20 predicates, to which participants gave ‘no’ ratings (indicating no entailment), there are no ‘no’ ratings for the relevant controls: on the inference diagnostic (panel a) none of the 341 participants whose data are plotted responded ‘no’ to any of the four entailing control stimuli (1,364 data points) and, on the contradictoriness diagnostic (panel b), none of the 302 participants whose data are plotted responded ‘no’ to any of the four contradictory control stimuli (1,208 data points). This lack of variability in the ratings of the relevant control stimuli means that we cannot use the control conditions as reference levels. **Explain why**

Because we were not able to use the controls with indisputably entailed content to identify entailed CCs, we made do by assuming that the CCs of the predicates that received the highest proportion of ‘yes’ ratings are entailed and used those predicates as reference levels: this was *prove* with 340 out of 341 ‘yes’ ratings on the inference diagnostic and *be right* with 291 out of 302 ‘yes’ ratings on the contradictoriness diagnostic. **Discuss models** In other words, on the inference diagnostic, the ratings for *be right*, *discover*, *know* and *see* are indistinguishable from those of *prove* and, on the contradictoriness diagnostic, the ratings for none of the predicates are indistinguishable from those of *be right*. Given the assumption that the CCs of *prove* and *be right* are entailed content, the findings of the inference diagnostic follow-up experiment suggest that the CCs of *be right*, *discover*, *know* and *see* are entailed. In sum, we have only found partial confirmation for the hypothesis that it was due to the gradient response scale of Exps. 2 that hardly any of the 20 clause-embedding predicates were found to have entailed CCs: the CCs of more clause-embedding predicates were classified as entailed by Exp. 2a’ than by Exp. 2a, but not by Exp. 3a’ compared to Exp. 3a (recall that the CC of *be right* is entailed merely by assumption).<sup>23</sup>

This means that we still cannot conclude with any confidence that the CC of any of the 20 predicates we investigated is entailed. The CC of *prove* is entailed according to Exp. 2a, but not according to Exps. 3a or 3b. (We assumed that it was entailed in Exp. 2a’.) And although Exp. 2a’ suggested that the CCs

<sup>23</sup>To be able to compare the findings from the two experiments, we fit a model to the Exp. 2a data with *prove* as the reference level. **Describe model** In other words, on that model the ratings for all predicates are lower than those of *prove*.

of *be right*, *discover*, *know* and *see* are entailed, we have no reason to assume that Exp. 2a' is better at diagnosing entailment than Exps. 2a, 3a and 3b, all of which provided evidence that the CCs of these predicates are not entailed. It is possible, for instance, that Exp. 2a' resulted in more entailed CCs because participants chose 'yes' even if they would have given an inference rating below ceiling on the gradient scale of Exp. 2a. It is also not clear that the inference diagnostic is more suitable to diagnosing entailment than the contradictoriness diagnostic: it is possible that even non-entailed content receives high inference ratings and that participants find it easier to identify non-entailed CCs when judging a potentially contradictory utterance. We therefore conclude that we do not have conclusive evidence that the CC of any of the 20 predicates we investigated is entailed.

This conclusion, we argue, is not due to the inference or the contradictoriness diagnostics not being able to diagnose entailed content: after all, control stimuli with indisputably entailed content received at-ceiling ratings in all four experiments.<sup>24</sup> Further evidence for our conclusion comes from the MegaVeridicality dataset, introduced in section 2.1.2, which also contains entailment ratings for the CCs of the 517 English clause-embedding predicates; these ratings were collected for stimuli that combined these predicates with arguments with low lexical content, as shown in (29) for the predicate *know*. To assess entailment, participants were asked to respond to the question *Did that thing happen?*; the response options were 'yes', 'maybe or maybe not' and 'no'.

(29) Somebody knew that a particular thing happened.

Each of the 517 predicates in the MegaVeridicality dataset received between 9 and 20 entailment ratings (mean: 10) from 159 participants. To plot the ratings, we coded a 'yes' response as 1, a 'maybe or maybe not' response as 0 and a 'no' response as -1. Figure 15 plots the mean entailment ratings for the 517 predicates in the MegaVeridicality dataset, with x-axis labels for 19 of the 20 predicates that we investigated.<sup>25</sup> There are 97 clause-embedding predicates that only received 'yes' ratings, including 3 of the factive predicates we investigated, namely *reveal*, *know* and *be annoyed*; this finding is compatible with the assumption that the CCs of these predicates are entailed.

<sup>24</sup>de Marneffe et al. (2012:329) suggested that veridicality ratings are influenced by pragmatic reasoning. We agree with this possibility but consider it unlikely that pragmatic reasoning would only influence participants' ratings on the target stimuli, to the exclusion of the control stimuli.

<sup>25</sup>The x-axis labels for the following predicates overlap: *acknowledge/announce/admit/confess/demonstrate* and *inform/know*.

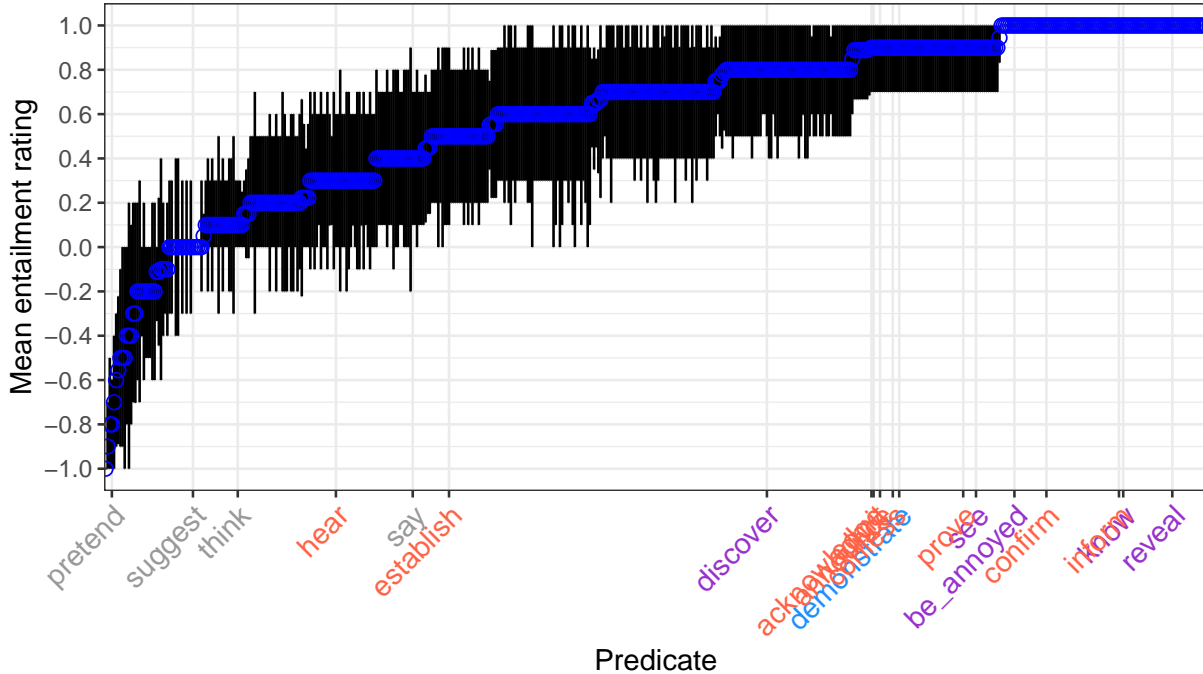


Figure 15: Mean entailment rating in blue, by predicate, with 95% bootstrapped confidence intervals, for the 517 predicates in MegaVeridicality dataset. OUR LABELS

However, as with our Exps. 2 and 2', not all of the factive and veridical non-factive predicates we investigated received only 'yes' ratings: the ratings for the purportedly factive predicates *see* and *discover* and for the purportedly veridical non-factive predicate *demonstrate* were not at ceiling, suggesting that their CCs are not entailed.<sup>26</sup> Furthermore, it is not clear that the CCs of the 97 predicates that only received 'yes' ratings are entailed: this set of 97 predicates includes *inform* (see the discussion in section 1) as well as the communication predicates *bitch* and *howl*, whose CCs are not entailed. For instance, it does not follow from the naturally occurring example with *bitch* in (30) that the Democratic party hates all the Jews.

- (30) 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.<sup>27</sup>

In sum, even though the MegaVeridicality database used a different entailment diagnostic and different response options than our experiments, it supports our conclusion that, of the 20 clause-embedding predicates we investigated, fewer than expected have entailed CCs.

### 3.6 Do projectivity and entailment jointly identify a class of factive predicates?

As discussed in section 1, a predicate is factive on the second definition if and only if its CC is presupposed and entailed. With data on projectivity and entailment in hand, we now return to the question of whether

<sup>26</sup>White and Rawlins (2018) came to a different conclusion because they assumed that the CC of a predicate is entailed if and only if the majority of the responses for a predicate was 'yes'. According to this linking function, the CCs of 376 of the 517 predicates in the MegaVeridicality dataset are entailed. We do not adopt this linking function because it is incompatible with the standard definition of entailment: for instance, given this linking function, *tweet* is a veridical predicate because 11 of the 20 participants responded 'yes', despite the fact that the remaining 9 participants responded 'maybe or maybe not'.

<sup>27</sup><https://www.wonkette.com/08-21-2019>

these two properties jointly distinguish a class of factive predicates from optionally factive and non-factive ones. We first consider the five purportedly factive predicates we investigated, namely *be annoyed*, *discover*, *know*, *reveal* and *see*. Regarding projectivity, Exp. 1a showed that the CC of each of these predicates is quite highly projective, in support of their categorization. Regarding entailment, however, Exps. 2 provided little empirical support for the assumption that the CCs of these predicates is entailed: although Exp. 2a' provided support that the CCs of *discover*, *know* and *see* are entailed, Exps. 2a, 3a and 3b provided evidence that the CCs of none of the five are entailed. Thus, projectivity and entailment do not jointly identify these five predicates as factive.

Is there a different class of predicates that projectivity and entailment jointly identify as factive? One might consider the set of predicates consisting of *be annoyed*, *know*, *see*, *inform* and *discover*: regarding projectivity, the CCs of each of these predicates are the most projective of the 20 predicates investigated. A problem, however, with assuming that this set of predicates is factive according to the second definition is that one would need to disregard the evidence from Exps. 2a, 3a and 3b that the CCs of these predicates are not entailed, even if Exp. 3a' provided evidence for entailment for *know*, *see* and *discover*. Another set one might consider consists of *prove*, *be right*, *discover*, *know* and *see*: regarding projectivity, the CCs of each of these predicates are projective, albeit only very weakly for some; regarding entailment, the CCs of each of these predicates were identified as entailed by at least one of Exps. 2 or 2'. One problem with assuming that this set of predicates is factive according to the second definition is that one would, again, need to disregard the evidence from Exps. 2 and 2' that the CCs of these predicates are not entailed. Another problem is that this set of predicates does not form a natural class, given how much the CCs of these predicates differ in projectivity: for instance, assuming that these five predicates are factive would not be helpful in developing a theory of presupposition projection. We conclude that the second definition of factive predicates is no more successful than the first in distinguishing factive predicates from optionally factive and non-factive ones.<sup>28</sup>

## 4 Discussion

The question addressed in this paper is: Which predicates are factive? Despite the central role that the distinction between factive and non-factive predicates has played in linguistic theorizing, different answers have been given to this question because, as noted in section 1, there is disagreement about how to define factive predicates, uncertainty about whether projection categorically distinguishes factive predicates from optionally factive and non-factive ones, and disagreement about which clause-embedding predicates have entailed CCs. In sections 2 and 3, we presented the findings of experiments designed to investigate which predicates are factive based on two definitions of factive predicates. Our experiments in section 2 investigated the first definition, according to which a predicate is factive if and only if its CC is presupposed: we found that projection alone does not categorically distinguish factive predicates from optionally factive and non-factive ones and, furthermore, that the CCs of all of the predicates we investigated are at least mildly projective. In section 3, we investigated the second definition, according to which a predicate is factive if and only if its CC is both presupposed and entailed: our experiments did not provide conclusive evidence that the CCs of any of the 20 predicates we investigated are entailed and so we concluded that the second definition also does not identify a class of factive predicates. We are therefore led to the conclusion that, at present, there is no empirical evidence for a class of factive predicates that is categorically distinguished from optionally factive and non-factive ones by two properties of the CC, namely projection and entailment.

<sup>28</sup>White and Rawlins (2018), who assumed the second definition of factive predicates, identified 199 factive predicates in the MegaVeridicality database. However, this finding is due to the authors assuming that a majority of 'yes' responses for a predicate means that the CC is entailed (for positive matrix sentences) or presupposed (for sentences embedded under negation or in a question). As discussed in footnote 26, the linking function for entailed CCs is not compatible with the standard definition of entailment. For presupposed CCs, it is not clear that a majority of 'yes' responses is motivated as a threshold for projection; see the discussion in section 2.1.2.

This does not, of course, mean that no empirical evidence for a class of factive predicates could ever be provided. One might, for instance, maintain the second definition of factive predicates and explore other ways of diagnosing entailed CCs. As discussed in de Marneffe et al. 2012 and Pavlick and Kwiatkowski 2019, participants' veridicality ratings appear to be influenced by a variety of pragmatic factors; it is an intriguing question for future research whether one can sufficiently control these pragmatic factors to elicit purely semantic judgments about veridicality to support claims about entailment. Another option is to define factive predicates differently, for instance on the basis of some syntactic property or as predicates for which the majority of participants judge the CC to be projective and entailed (see, e.g., White and Rawlins 2018). Such definitions will have to be assessed by the explanatory power of the resulting categorizations of clause-embedding predicates for the theoretical phenomenon under investigation.

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

Even though our investigation has not yielded a class of factive predicates, it has revealed that the CCs of more clause-embedding predicates than previously assumed are projective, albeit some are only mildly so. This finding means that research on projective content has a much broader empirical scope than previously assumed. Given that research on projective content has focused almost entirely on the the CCs of factive predicates, to the exclusion of the CCs of optionally factive and non-factive ones, our finding means that we are now tasked with investigating and formally analyzing the projectivity of the CCs of a much broader set of clause-embedding predicates than before. For instance, research on projective content has identified a number of factors that influence the projectivity of utterance content, including context, lexical content, prior content probability, at-issueness, information structure and the perceived degree of reliability and trustworthiness of the attitude holder (e.g., Gazdar 1979a,b, Beaver 2010, Schlenker 2010, Simons et al. 2010, 2017, Abrusán 2011, 2016, Anand and Hacquard 2014, Cummins and Rohde 2015, Djärv and Bacovcin 2017, Tonhauser 2016, in press, Tonhauser et al. 2018, 2019). An exciting question for future research is whether these factors are implicated in the projectivity of the CCs of all clause-embedding predicates, or only particular subsets.

On the theoretical side, our finding that the CCs of more clause-embedding predicates than previously assumed are projective calls for the development of formal analyses of the projectivity of these CCs. One approach to predicting the projectivity of the CCs of optionally factive and non-factive predicates has been to assume that such predicates are ambiguous: Spector and Egré (2015), for instance, proposed (p.1736) that predicates like *tell*, *predict* and *announce* have a factive lexical entry on which the CC is entailed and presupposed and a non-factive one on which it is not. The findings of our Exps. 1 point to some difficulties for this approach, given that the CC of all 20 predicates we investigated is at least mildly projective. First, this approach requires the assumption of widespread ambiguity among clause-embedding predicates. Second, it is not clear which predicates would only have a factive or only a non-factive lexical entry, and which ones would be ambiguous. And, finally, the approach cannot account for the observed by-predicate projection variability: for instance, the assumption that both *discover* and *announce* have a factive and a non-factive



lexical entry does not predict that the CC of *discover* more projective than the CC of *announce*.

We hypothesize that more fine-grained distinctions among the lexical meanings of clause-embedding predicates than the distinction between factive and non-factive predicates (or the three-way distinction between factive, optionally factive and non-factive predicates) matter for predicting speaker commitment to the CC, both in affirmative matrix clauses and under entailment-canceling operators. More specifically, we hypothesize that the CCs of classes of predicates that share lexical meanings and discourse uses are similar in projectivity. We are aware of two tentative proposals along those lines in the recent literature. First, Anand and Hacquard (2014) hypothesized that speakers may be taken 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). The authors briefly entertained an extension of this proposal to utterances of sentences with *inform* and *announce*, which convey that the CC is part of the reported common ground: “to the extent that we identify ourselves with that context” (p.76), the speaker and the addressee are committed to the truth of the CC. While this proposal would need to be fleshed out in more detail to be predictive, we are sympathetic to the idea that the lexical meaning of assertive predicates and the way such predicates are used in discourse plays a role in predicting the projection of the CC.

A second proposal was sketched in Karttunen 2016: this work gave up on the idea that there is a class of factive predicates with a uniform definition and instead proposed that more fine-grained distinctions among clause-embedding predicates play a role in the projection of the CC. Karttunen (2016) distinguished several subclasses of factive 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 Karttunen’s 2016 proposal, too, would 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 Conclusions

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

predicates, and also opens up several exciting new avenues for empirical and theoretical research on the projection of the contents of the complements of clause-embedding predicates.

## A 20 complement clauses

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

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

## B Details on the Bayesian models for Experiments 1 and 2

## C Experiments 1', 2a' and 3b

As mentioned in sections 2 and 3, we re-ran the experiments reported on in those sections with a two-alternative forced choice ratings (response options 'yes' or 'no') instead of a gradient response scales.

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

**Data exclusion** Prior to analysis, we excluded the data from participants who did not self-identify as native speakers of American English. If a participant took an experiment more than once, we only kept the data from the first time they took the experiment. We also excluded the data from participants who gave a wrong rating to at least one control: a rating was considered wrong if it was a 'yes' rating on a projectivity control, a non-entailing control or a non-contradictory control, or a 'no' rating on an entailing control or a contradictory control. The results presented below are based on data from 426 participants (ages 18-81; median: 37; 200 female, 220 male, 2 other) for Exp. 1b, from 341 participants (ages 18- 73; median: 38; 180 female, 161 male) for Exp. 2a' and from 302 participants (ages 18-73; median: 37; 150 female, 152 male) for Exp. 3a'.

**Results and discussion** For the main findings, see Figures 4 and ???. The three panels of Figure A1 plot the findings of Exps. 1, 2a and 3a against the findings of Exps. 1', 2a' and 3b, respectively.

- Experiment 1': The CCs of all predicates were projective compared to the projectivity of the CC of *be right*, which were the least projective of all CCs, the CCs of the other 19 predicates were more projective. **Describe model here, explain why we couldn't use the main clause controls as reference level: no variance.**

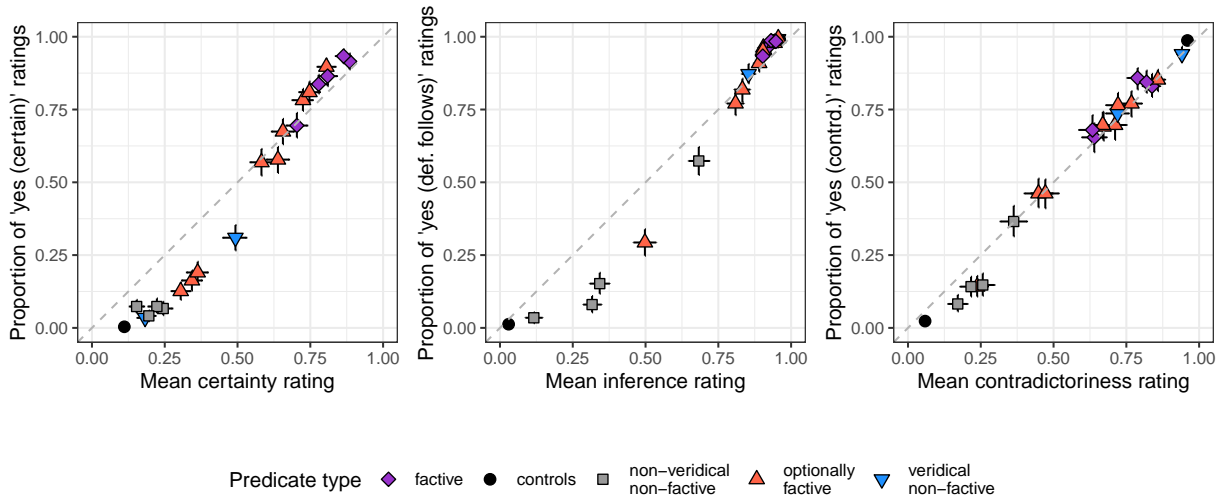


Figure A1: Proportion of ‘yes’ ratings in two-alternative forced choice task by mean ratings on gradient scale, with 95% bootstrapped confidence intervals, collapsing over complement clauses.

## D Model output of Exp. 1b

Formula: `model.brms.proj.b = brm(nResponse ~ verb + (1/workerid) + (1/item), data=cd, family=bernoulli(), cores = 4, control=list(max_treedepth = 15))`

## E Model output of Exp. 2b

Formula: `model.brms.proj.b = brm(nResponse ~ verb + (1/workerid) + (1/item), data=cd, family=bernoulli(), cores = 4, control=list(max_treedepth = 15, adapt_delta=.95))`

## F Model output of Exp. 3b

Formula: `model.brms.proj.b = brm(nResponse ~ verb + (1/workerid) + (1/item), data=cd, family=bernoulli(), cores = 4, control=list(max_treedepth = 15, adapt_delta=.95))`

## References

- Abbott, Barbara. 2006. Where have some of the presuppositions gone? In B. J. Birner and G. Ward, eds., *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn*, pages 1–20. Philadelphia: John Benjamins.
- Abrusán, Márta. 2011. Predicting the presuppositions of soft triggers. *Linguistics & Philosophy* 34:491–535.
- Abrusán, Márta. 2014. *Weak island semantics*. New York: Oxford University Press.
- Abrusán, Márta. 2016. Presupposition cancellation: Explaining the ‘soft-hard’ trigger distinction. *Natural Language Semantics* 24:165–202.
- Anand, Pranav, Jane Grimshaw, and Valentine Hacquard. ms. Sentence embedding predicates, factivity and subjects. In C. Condoravdi, ed., *A Festschrift for Lauri Karttunen*. Stanford: CSLI Publications.
- Anand, Pranav and Valentine Hacquard. 2014. Factivity, belief and discourse. In *The Art and Craft of Semantics: A Festschrift for Irene Heim*, pages 69–90. MIT Working Papers in Linguistics.
- Anderson, Lloyd B. 1986. Evidentials, paths of change, and mental maps: Typologically regular asymme-

	Model 1
b_Intercept	−6.37* [−7.10; −5.72]
b_verbbe_right	2.33* [1.50; 3.21]
b_verbthink	2.54* [1.75; 3.40]
b_verbsay	3.10* [2.35; 3.93]
b_verbpretend	3.22* [2.50; 4.03]
b_verbsuggest	3.22* [2.49; 3.99]
b_verbprove	3.92* [3.21; 4.67]
b_verbconfirm	4.27* [3.60; 5.01]
b_verbestablish	4.51* [3.83; 5.26]
b_verbdemonstrate	5.31* [4.64; 6.06]
b_verbannounce	6.70* [6.04; 7.45]
b_verbconfess	6.75* [6.07; 7.50]
b_verbadmit	7.29* [6.60; 8.04]
b_verbreveal	7.41* [6.72; 8.16]
b_verbacknowledge	8.00* [7.31; 8.78]
b_verbhear	8.22* [7.50; 9.01]
b_verbdiscover	8.46* [7.75; 9.25]
b_verbsee	8.74* [8.03; 9.55]
b_verbinform	9.13* [8.38; 9.94]
b_verbbe_annoyed	9.41* [8.64; 10.22]
b_verbknow	9.72* [8.94; 10.57]

\* 0 outside 95% credible interval

Table A1: Model output of Exp. 1b

	Model 1
b_Intercept	6.20* [5.52; 6.98]
b_verbacknowledge	-1.58* [-2.47; -0.70]
b_verbadmit	-1.83* [-2.72; -0.97]
b_verbannounce	-4.39* [-5.17; -3.68]
b_verbbe_annoyed	-1.36* [-2.31; -0.46]
b_verbbe_right	0.36 [-0.93; 1.89]
b_verbconfess	-2.79* [-3.60; -2.04]
b_verbconfirm	-0.85 [-1.84; 0.14]
b_verbdemonstrate	-3.35* [-4.15; -2.63]
b_verbdiscover	-0.50 [-1.57; 0.64]
b_verbestablish	-1.92* [-2.76; -1.12]
b_verbhear	-7.57* [-8.41; -6.84]
b_verbinform	-3.93* [-4.71; -3.22]
b_verbknow	-0.27 [-1.39; 0.96]
b_verbnonMent.C	-11.92* [-12.94; -11.01]
b_verbpretend	-10.78* [-11.83; -9.85]
b_verbprove	0.92 [-0.57; 2.91]
b_verbreveal	-2.37* [-3.21; -1.54]
b_verbsay	-5.79* [-6.59; -5.09]
b_verbsee	-0.49 [-1.51; 0.62]
b_verbsuggest	-8.75* [-9.62; -7.95]
b_verbthink	-9.71* [-10.65; -8.87]

\* 0 outside 95% credible interval

Table A2: Model output of Exp. 2b

	Model 1
b.Intercept	6.17* [5.55; 6.88]
b.verbacknowledge	-4.89* [-5.65; -4.21]
b.verbadmit	-4.86* [-5.59; -4.19]
b.verbannounce	-6.46* [-7.20; -5.80]
b.verbbe_annoyed	-5.18* [-5.91; -4.51]
b.verbbe_right	-1.90* [-2.70; -1.12]
b.verbconfess	-4.85* [-5.59; -4.18]
b.verbconfirm	-4.25* [-4.99; -3.57]
b.verbdemonstrate	-4.53* [-5.25; -3.84]
b.verbdiscover	-3.30* [-4.06; -2.61]
b.verbestablish	-4.29* [-5.06; -3.62]
b.verbhear	-8.98* [-9.79; -8.27]
b.verbinform	-6.46* [-7.20; -5.81]
b.verbknow	-3.64* [-4.39; -2.95]
b.verbnonMcontrd.C	-11.51* [-12.43; -10.71]
b.verbpretend	-8.97* [-9.78; -8.26]
b.verbprove	-3.36* [-4.10; -2.65]
b.verbreveal	-4.99* [-5.73; -4.32]
b.verbsay	-7.11* [-7.87; -6.42]
b.verbsee	-3.48* [-4.21; -2.78]
b.verbsuggest	-8.92* [-9.71; -8.21]
b.verbthink	-9.81* [-10.67; -9.06]

\* 0 outside 95% credible interval

Table A3: Model output of Exp. 3b

- tries. In W. Chafe and J. Nichols, eds., *Evidentiality: The Linguistic Coding of Epistemology*, pages 273–312. New Jersey: Ablex Publishing Corporation.
- Aravind, Athulya and Martin Hackl. 2017. Factivity and at-issueness in the acquisition of *forget* and *remember*. In *41st Annual Boston University Conference on Language Development*, pages 46–59.
- Beaver, David. 2001. *Presupposition and Assertion in Dynamic Semantics*. Stanford, CA: CSLI Publications.
- Beaver, David. 2010. Have you noticed that your belly button lint colour is related to the colour of your clothing? In R. Bäuerle, U. Reyle, and E. Zimmermann, eds., *Presuppositions and Discourse: Essays Offered to Hans Kamp*, pages 65–99. Oxford: Elsevier.
- Bürkner, Paul-Christian. 2017. brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1):1–28.
- Chierchia, Gennaro and Sally McConnell-Ginet. 1990. *Meaning and Grammar*. Cambridge, MA: MIT Press.
- Cremers, Alexandre. 2018. Plurality effects in an exhaustification-based theory of embedded questions. *Natural Language Semantics* 26:193–251.
- Cummins, Chris and Hannah Rohde. 2015. Evoking context with contrastive stress: Effects on pragmatic enrichment. *Frontiers in Psychology* 6:Article 1779.
- de Marneffe, Marie, Mandy Simons, and Judith Tonhauser. 2019. The CommitmentBank: Investigating projection in naturally occurring discourse. In *Sinn und Bedeutung* 23, pages 107–124.
- de Marneffe, Marie-Catherine, Christopher D. Manning, and Christopher Potts. 2012. Did it happen? The pragmatic complexity of veridicality assessment. *Computational Linguistics* 38:301–333.
- de Villiers, Jill G. 2005. Can language acquisition give children a point of view. In J. Astington and J. Baird, eds., *Why language acquisition matters for theory of mind*, pages 186–219. Oxford University Press.
- Djäv, Kajsa and Hezekiah Akiva Bacovcin. 2017. Prosodic effects on factive presupposition projection. In *Semantics and Linguistic Theory (SALT) XXVII*, pages 116–133. Ithaca, NY: CLC Publications.
- Djäv, Kajsa, Jérémy Zehr, and Florian Schwarz. 2016. Cognitive vs. emotive factives: An experimental differentiation. In *Sinn und Bedeutung* 21, pages 367–385.
- Egré, Paul. 2008. Question-embedding and factivity. In L. Lihoreau, ed., *Grazer Philosophische Studien*, pages 85–125. Amsterdam: Rodopi.
- Gazdar, Gerald. 1979a. *Pragmatics: Implicature, Presuppositions, and Logical Form*. New York: Academic Press.
- Gazdar, Gerald. 1979b. A solution to the projection problem. In C.-K. Oh and D. Dineed, eds., *Syntax and Semantics 11: Presupposition*, pages 57–89. New York: Academic Press.
- Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non)veridical dependency*. Dordrecht: John Benjamins.
- Giannakidou, Anastasia. 2001. The meaning of free choice. *Linguistics and Philosophy* 24:659–735.
- Givón, Talmy. 1995. *Functionalism and grammar*. Amsterdam: John Benjamins.
- Guerzoni, Elena and Yael Sharvit. 2007. A question of strength: On NPIs in interrogative clauses. *Linguistics & Philosophy* 30:361–391.
- Hazlett, Allan. 2010. The myth of factive verbs. *Philosophy and Phenomenological Research* 80:497–522.
- Heim, Irene. 1983. On the projection problem for presuppositions. In M. Barlow, D. Flickinger, and M. Westcoat, eds., *West Coast Conference on Formal Linguistics (WCCFL) 2*, pages 114–125.
- Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9:183–221.
- Heycock, Caroline. 2006. Embedded root phenomena. In M. Everaert and H. van Riemsdijk, eds., *The Blackwell Companion to Syntax*, page xxx. Wiley.
- Hintikka, Jaakko. 1975. Different constructions in terms of the basic epistemological verbs: A survey of some problems and proposals. In ??, ed., *The intentions of intentionality and other new models for*

- modalities. Dordrecht: Reidel.
- Hukari, Thomas and Robert Levine. 1995. Adjunct extraction. *Natural Language and Linguistic Theory* 31:195–226.
- Jaynes, Edwin and O. Kempthorne. 1976. Confidence intervals vs. Bayesian intervals. In W. L. Harper and C. A. Hooker, eds., *Foundations of probability theory, statistical inference, and statistical theories of science*, pages 175–257. Dordrecht: Springer Netherlands.
- Karttunen, Lauri. 1971a. Implicative verbs. *Language* 47:340–358.
- Karttunen, Lauri. 1971b. Some observations on factivity. *Papers in Linguistics* 4:55–69.
- Karttunen, Lauri. 2016. Presupposition: What went wrong? In *Proceedings of Semantics and Linguistic Theory (SALT) XXVI*, pages 705–731. Ithaca, NY: CLC Publications.
- Karttunen, Lauri and Stanley Peters. 1979. Conventional implicature. In C. Oh and D. Dineen, eds., *Presuppositions*, Syntax and Semantics Vol.11, pages 1–56. New York: Academic Press.
- Karttunen, Lauri and Annie Zaenen. 2005. Veridicity. In G. Katz, J. Pustejovsky, and F. Schilder, eds., *Annotating, Extracting and Reasoning about Time and Events*, no. 05151 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.
- Kastner, Itamar. 2015. Factivity mirrors interpretation: The selectional requirements of presuppositional verbs. *Lingua* 164:156–188.
- Kiparsky, Paul and Carol Kiparsky. 1970. Fact. *Progress in Linguistics* pages 143–173.
- Klein, Ewan. 1975. Two sorts of factive predicates. Pragmatics Microfiche it 1.1. frames B5C14.
- Landau, Idan. 2001. *Elements of control: Structure and meaning in infinitival constructions*. Springer.
- Lasersohn, Peter. 2009. Relative truth, speaker commitment, and control of implicit arguments. *Synthese* 166:359–374.
- Liu, Fang and Yunchuan Kong. 2015. Zoib: an r package for bayesian inference for beta regression and zero/one inflated beta regression. *The R Journal* 7(2):34–51.
- Mahler, Taylor. 2019. Does at-issueness predict projection? It’s complicated! In *49th Annual Meeting of the North East Linguistic Society*.
- Morey, Richard D., Rink Hoekstra, Jeffrey N. Rouder, Michael D. Lee, and Eric Jan Wagenmakers. 2016. The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin and Review* 23(1):103–123.
- Nicenboim, Bruno and Shravan Vasishth. 2016. Statistical methods for linguistic research: Foundational Ideas, Part II. *Linguistics and Language Compass* 10(11):591–613.
- Pavlick, Ellie and Tom Kwiatkowski. 2019. Inherent disagreements in human textual inferences. In *Transactions of the Association for Computational Linguistics*, pages 677–694. Association for Computational Linguistics.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Romoli, Jacopo. 2015. The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of Semantics* 32:173–291.
- Rooryck, Johan. 2000. *Configurations of sentential complementation*. London: Routledge.
- Schlenker, Phillippe. 2003a. The lazy Frenchman’s approach to the subjunctive. In *Proceedings of Romance Languages and Linguistic Theory*, pages 269–309.
- Schlenker, Phillippe. 2003b. A plea for monsters. *Linguistics & Philosophy* 26:29–120.
- Schlenker, Phillippe. 2010. Local contexts and local meanings. *Philosophical Studies* 151:115–142.
- Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117:1034–1056.
- Simons, Mandy, David Beaver, Craige Roberts, and Judith Tonhauser. 2017. The Best Question: Explaining the projection behavior of factive verbs. *Discourse Processes* 3:187–206.



- Simons, Mandy, Judith Tonhauser, David Beaver, and Craige Roberts. 2010. What projects and why. In *Semantics and Linguistic Theory (SALT) XXI*, pages 309–327. Ithaca, NY: CLC Publications.
- Smith, E. Allyn and Kathleen Currie Hall. 2011. Projection diversity: Experimental evidence. In *Proceedings of the 2011 ESSLLI Workshop on Projective Content*, pages 156–170.
- Spector, Benjamin and Paul Egré. 2015. A uniform semantics for embedding interrogatives: *An answer, not necessarily the answer*. *Synthese* 192:1729–1784.
- Stevens, Jon Scott, Marie-Catherine de Marneffe, Shari R. Speer, and Judith Tonhauser. 2017. Rational use of prosody predicts projectivity in manner adverb utterances. In *39th Annual Meeting of the Cognitive Science Society*, pages 1144–1149.
- Swanson, Eric. 2012. Propositional attitudes. In K. von Heusinger, C. Maienborn, and P. Portner, eds., *Semantics: An International Handbook of Natural Language Meaning*, vol. 2, pages 1538–1561. Berlin: Mouton de Gruyter.
- Tonhauser, Judith. 2016. Prosodic cues to speaker commitment. In *Semantics and Linguistic Theory (SALT) XXVI*, pages 934–960. Ithaca, NY: CLC Publications.
- Tonhauser, Judith. in press. Projection variability in Paraguayan Guaraní. *Natural Language and Linguistic Theory*.
- Tonhauser, Judith, David Beaver, and Judith Degen. 2018. How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics* 35:495–542.
- Tonhauser, Judith, David Beaver, Craige Roberts, and Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89:66–109.
- Tonhauser, Judith, Marie-Catherine de Marneffe, Shari R. Speer, and Jon Stevens. 2019. On the information structure sensitivity of projective content. In *Sinn und Bedeutung* 23, pages 363–389.
- van der Sandt, Rob. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9:333–377.
- van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins.
- White, Aaron S. and Kyle Rawlins. 2018. The role of veridicality and factivity in clause selection. In *48th Meeting of the North East Linguistic Society*, page ??
- White, Aaron Steven, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2018. Lexicosyntactic inference in neural models. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4717–4724. Brussels, Belgium: Association for Computational Linguistics.
- Wyse, Brendan. 2010. *Factive/non-factive predicate recognition within Question Generation systems*. Master’s thesis, Open University.
- Xue, Jingyang and Edgar Onea. 2011. Correlation between projective meaning and at-issueness: An empirical study. In *Proceedings of the 2011 ESSLLI Workshop on Projective Content*, pages 171–184.
- Zanuttini, Raffaella and Paul Portner. 2003. Exclamative clauses: At the syntax-semantics interface. *Language* 79:39–81.