# Are false implicatures lies? An empirical investigation

Benjamin Weissman & Marina Terkourafi
University of Illinois at Urbana-Champaign & Leiden University

### ABSTRACT

Lies are typically defined as asserted falsehoods uttered with the intention of deceiving the hearer. A particularly problematic case for this definition is that of false implicatures. These are prototypically cases where the proposition expressed by the speaker's utterance is true, yet an implicature conveyed by this proposition in context is false. However, implicature is a diverse category and whether a blanket statement such as "false implicatures are lies," as some have argued can account for all of them is open to investigation. We present an experimental investigation of whether naïve participants define different types of implicatures as lies. Our results show that only a couple of types of implicatures were strongly rated as lies by participants. These results suggest that participants distinguish between different types of communicated meanings on linguistic grounds, contributing both to the literature on lying, as well as to theoretical discussions of how different types of meaning are communicated.

**1 Introduction**

Many researchers of lying (e.g., Williams, 2002, p. 96; Mahon, 2008, p. 2; Arico & Fallis, 2013, p. 791) refer to a three-pronged definition that can be traced back to the writings of Saint Augustine (1952), a version of which is also found in dictionaries. This definition can be summarized as follows: a speaker S asserts a proposition *p* to a hearer H, S believes *p* is false, and S asserts *p* with an intention to deceive H into believing that *p*, or at least believing that S believes that *p*. Some versions include a criterion that *p* be actually false. Not all cases are straightforward, however, and this definition may fall short of truly and fully encompassing what speakers consider to be a lie on different occasions.

A particularly problematic case is that of false implicatures. These are instances where the proposition expressed by the speaker's utterance is true, yet an implicature carried by this utterance in context is false and knowingly so to the speaker; according to Meibauer (2005, 2011, 2014), these, too, can be lies. However, implicatures come in many shapes and forms and whether a blanket statement such as "false implicatures are lies" can account for all of them is open to investigation.[1] In this article, we present an experimental investigation of whether naïve participants define different types of false implicatures as lies. Our results show that participants are more readily prepared to call certain types of false implicatures lies than others. These results suggest that participants can distinguish between different types of implicated meanings, and do so more readily when real-world, moral consequences follow from this distinction.

We begin by presenting some challenges to the dictionary definition of lying (Section 2), focusing especially on false implicatures (Section 3). We then introduce the rationale for our study (Section 4), followed by our methodology (Section 5) and results (Section 6). A possible explanation for these results is presented next (Section 7), followed by discussion (Section 8), and our conclusions (Section 9).

**2 Lying between truth, morality, and culture**

While countless authors from various fields have assumed that the three-pronged definition of lying is, on its own, enough to get by (DePaulo et al., 1996; Lee, 2000; Williams, 2002; among others), others have challenged it theoretically (e.g., Carson, 2006; Sorensen, 2007; Fallis, 2009; Saul, 2012; Lackey, 2013; Meibauer, 2014), as well as empirically. Based on participant judgments of several stories, Coleman & Kay (1981) proposed that *lie* may work best as a prototype, with a "prototypical" lie instantiating all three of the following criteria: the proposition *p* is false, the speaker believes *p* to be false, the speaker utters *p* with the intention of deceiving the hearer – although utterances instantiating only one or two of these criteria may also be rated as lies. Their results suggested the following ranking of the three criteria in decreasing order of importance: "believe false," "intent to deceive," and "actually false." Using a similar methodology, Turri & Turri (2015) found that actual falsity, despite being the least important of the three criteria in the original Coleman & Kay study, is, in fact, required for an utterance to be considered a lie (so long as the other two are also met). However, a follow-up study by Wiegmann et al. (2016) argues convincingly that the Turri & Turri experiments guided

---

[1] Here and throughout this article, we use "implicature" in the classic Gricean sense of a part of speaker meaning (Horn 2004: 21-22). Therefore, the blanket statement "false implicatures are lies" should each time be understood as referring to propositions the speaker communicates through his/her utterance and which s/he knows to be false, not cases where the speaker is simply mistaken (s/he is implicating a proposition s/he believes to be true but which is actually false); though see Danziger's (2010) experiment with the Mopan Maya reported in the next section for some interesting cross-cultural complications with this characterization of lying.

participants towards a particular answer, and a re-worked series of experiments suggest that "objective falsity" (Turri & Turri, 2015, p. 162) is not, in fact, required for lying.

Culture has been found to play a role in defining lying as well. In a study by Hardin (2010), what seemed deceptive (and a lie) to English speakers was regarded as polite (not a lie) by Spanish speakers from eastern Ecuador.[2] Similarly, Danziger (2010) found that a story in which a speaker accidentally makes a false claim was treated as a lie by her Mopan Maya participants but not by American participants. Such evidence points to a different ranking or even absence of some of the lying criteria from definitions of the corresponding lexemes in different languages/cultures, highlighting the complexity of trying to access the relevant concepts through the corresponding lexemes available in each language (see also Sweetser 1987).

Despite this variability in judgements regarding precise criteria for lying and their eventual ranking, to call someone a liar, except jokingly, is to make a moral judgement about their character. It could indeed be argued that this moral dimension is what remains stable across cultures when something is characterized as an unequivocal lie, shifting the onus for deciding whether something is a lie toward the speaker's intention to deceive the listener and, perhaps even more so, their motivation for doing so (to cause harm to the listener or, conversely, to protect/benefit them). If that is correct, one might argue that an experimental investigation of people's intuitions about lying can yield little insight into whether different types of implicated meanings are susceptible to characterization as a lie, and therefore more or less readily integrated into an expression's truth-conditional content. However, this question can also be turned on its head: precisely *because* accusations of lying are consequential and not to be made lightly, people may be driven to sharpen their intuitions regarding whether some material is part of an utterance's truth-conditional content when asked whether a speaker who asserted this material has lied. This is the approach we take in this article, and it is corroborated by our experimental results.

**3 Previous research on false implicatures**

The following story presents an example of a false implicature:

(1)     *The Story of the Mate and the Captain*
        A captain and his mate have a long-term quarrel. The mate drinks more rum than is good for him, and the captain is determined not to tolerate this behaviour any longer. When the mate is drunk again, the captain writes in the logbook: *Today, October 11th, the mate is drunk*. When the mate reads this entry during his next watch, he is first getting angry, then, after a short moment of reflection, he writes into the logbook: *Today, October 14th, the captain is not drunk*.

        (Meibauer, 2005, p. 1380; based on Posner, 1980)

In this story, the mate's asserting that the captain is not drunk can generate the conversational implicature that the captain is almost always drunk, and the fact that he is not drunk is an unusual and noteworthy occurrence by virtue of Grice's (1975) first maxim of Quantity ("Be as informative as is required") and the maxim of Relation ("Be relevant").

---

[2] This discrepancy points to another important dimension that is largely unexplored, that of the liar's perceived blameworthiness. Lies uttered for the sake of constituting the speaker or the addressee's face may be defined as "white lies," definitions of which are also subject to variation cross-linguistically (see Terkourafi, forthcoming).

Focusing on the fact that these false propositions are not explicitly uttered, some authors have characterized such examples as "indirect lies" (Vincent & Castelfranchi, 1980) or lies (sometimes without using the term "implicature" to describe them; Shibles, 1985, p. 82). Others classify them as a type of deception other than lying (Adler, 1997; Dynel, 2011) or "merely misleading" (Saul, 2012, 37, p. 118; Stokke, 2013a, p. 354), denying that they constitute lies at all (Fallis, 2009, p. 40; Horn, forthcoming). Michaelson (2016) uses the idea that false implicatures are precisely *not* lies as the basis for his Lying Test, which draws on lie judgments to determine whether something is part of the semantic content of an utterance.

On the other hand, Meibauer (2005, 2011, 2014) argues that false implicatures *are* lies and should be accounted for in an extended definition of lying. If the captain in the story is, indeed, not drunk, but this is not such an unusual occurrence, then the assertion is true and the implicature is false. Meibauer claims that in this context, the mate has lied, even though technically his assertion is true; because implicatures "are intended by the speaker and are intended to be derived by the hearer … then it is of course possible to lie by intentionally using false implicatures" (2014, p. 114). This is captured in his extended definition of lying:

(2)     S lied at time *t* by uttering the declarative sentence σ iff:
        1. S asserted at *t* that *p.*
        2a. S actively believed at t that *not p.*
        OR
        2b. S thereby conversationally implicated that *q*, but actively believed that *not q.*
                                                                    (Meibauer, 2014, pp. 102–103)

Meibauer (2014, p. 1380) discusses *The Story of the Mate and the Captain* is as an example of particularized conversational implicature (PCI), a type of implicature that is context-specific and can only be derived with recourse to the speaker's intention. In addition to PCIs, Grice (1975) introduced generalized conversational implicatures (GCIs), a type of implicature that is calculable from the use of the words themselves, unless specific circumstances warrant against such an interpretation. An example of a false GCI (via Stokke, 2013b, p. 84) is given in (3) below:

(3)     Jasper's neighborhood recently put on a Community Week. People helped their neighbors
        out with various chores and tasks that needed doing. Selfishly, however, Jasper used
        Community Week to fix the roof on his own house, ignoring the neighbors. The
        following week Jasper is having dinner with Doris. Jasper is keen to give Doris a good
        impression of himself.
        Doris: So how did you help out during Community Week?
        Jasper: I fixed a roof.

Jasper's uttering of "a roof" in (3) carries the implicature that he did not fix *the* roof, which is in turn strengthened to mean the roof of his own house (Levinson, 2000, p. 92). Thus, the implicature that Jasper fixed the roof of someone else's house, under the pretense of Community Week, is licensed. Here the assertion is true, since he did fix a roof, but the implicature that the roof was not his own is false. According to Stokke, Jasper's utterance is a lie, and Meibauer claims explicitly that both GCIs and PCIs can be lies (2014, p. 135).

Others disagree with Meibauer's claim. Saul (2012) explores both philosophical and definitional aspects of lying, striving to elucidate a distinction between lying and misleading. According to her working definition of lying (she does not provide a separate one for misleading),[3] speakers lie if they are

(4)     "not the victim of linguistic error/malapropism or using metaphor, hyperbole, or irony … [and] (1) they say that P; (2) they believe P to be false; (3) they take themself to be in a warranting context" (Saul, 2012, p. 18).

Two aspects of this definition are especially interesting for us. First, a warranting context refers to a situation in which one "promises or guarantees, either explicitly or implicitly, that what one says is true" (Carson, 2006, p. 294, cited in Saul, 2012, p. 10). This additional criterion means that acting in a play or joke-telling do not count as lying. Second, the precise notion of "saying" that is at issue is important, since, according to Saul, the distinction between lying and misleading hinges on it. While Saul herself does not elaborate further on this distinction beyond maintaining that "lying" requires something to be said explicitly while misleading is the result of conversationally implicated disbelieved content, Stokke (2016) proposes that the line between said and implicated (and, therefore, also between lying and misleading) can be drawn with respect to the notion of QUD [Question Under Discussion]. According to this proposal, what is said by a sentence in context can be defined as "the weakest answer to a QUD that entails a minimal proposition expressed by the utterance in question, given the context" (Stokke, 2016: 18). On this view, misleading – a category distinct from lying – occurs when the speaker's utterance gives rise to disbelieved content that is not an answer to the QUD and is therefore not explicitly said.

Several definitions of what is said have been proposed since Grice used this term to draw a distinction between the proposition expressed by an utterance and any additional propositions (implicatures) communicated by it (Grice, 1975).[4] Saul presents her definition of what is said (henceforth WIS) as being testable following her "NTE" (Needed for Truth Evaluability) criterion:

(5)     "A putative contextual contribution to what is said is a part of what is said only if without this contextually supplied material, S[entence] would not have a truth-evaluable semantic content in C[ontext]."

(Saul, 2012, p. 57)

The NTE criterion captures the fact that, in order to distinguish between lying and misleading, we need a notion of WIS that has been minimally enriched to yield a truth-evaluable

---

[3] The version cited here in (4) serves as her working definition that she returns to throughout the manuscript and is adequate for our current purposes. Saul later also refers to a "complete" definition of lying, consisting of either (A) – the three conditions (1), (2), (3) listed here under (4), or (B) – comprised of "(1) They say something indeterminate across a range of acceptable complete propositions, CP1 … CPn; (2) for each complete proposition in the range CP1 … CPn, they believe that proposition to be false; (3) They take themselves to be in a warranting context" (Saul, 2012, p. 65).

[4] Definitions of what is said in the literature range from minimalist, subpropositional ones (Bach, 2001) to more inflationary accounts (Recanati, 2001), while other scholars (e.g., Carston, 1999) challenge the usefulness of this notion altogether. For an overview, see Terkourafi (2010).

proposition.[5] When speakers mislead, they communicate something above and beyond what they "say," hoping the audience will "[leap] straight to a false proposition" (i.e., the implicated one) (Saul, 2012, p. 68). Although Saul does not use the term "implicature" to describe this false proposition, her examples and definitions suggest that she considers false implicatures to be not lies but, rather, "merely misleading" (Saul, 2012, pp. 37, 68, 117). Conversely, for something to count as a lie, it must be part of WIS (as tested by the NTE).

      Stokke (2013b) and Fallis (2014) agree with Saul's appeal to WIS as it relates to lying but suggest reasons why a different definition of WIS is necessary to account for all lying and non-lying phenomena. Horn (2017) takes a similar theoretical stance, positing that what is implicated "never yields a lie" (p. 50) and maintaining that a lying-misleading distinction is a quite important one to keep. Using numerous courtroom and literary encounters to highlight the importance of this distinction in the real world, he furthermore argues that adopting a definition such as Meibauer's risks losing it. Finally, Dynel (2011) also argues that lies concern only WIS; if something untrue is communicated but it is not part of WIS, then, according to Dynel, this constitutes deception without lying. Regardless of the specific definition of WIS they each adopt, these authors agree that lying applies only to WIS, contrary to the account offered by Meibauer (2005, 2011, 2014).

      In her review of Meibauer (2014), Dynel (2015) hypothesizes that floutings of Quality (such as irony, metaphor, hyperbole, and meiosis/litotes) might be more likely to count as lies (Dynel, 2015, p. 328), because in these cases, the speaker's utterance is overtly untrue and therefore a literal WIS is not available for the speaker to fall back on. However, she makes no such claim about other types of false implicature. Instead, she mentions that the diversity of implicatures means that treating all of them as either lies or not lies would be "a sweeping generalization" and concludes that this issue deserves a more fine-grained analysis (Dynel, 2015, p. 327). In his response to her review, Meibauer echoes this sentiment: "it seems important that more empirical data [regarding lying] be consulted on the basis of, for instance, experimental pragmatics/semantics" (Meibauer, 2016a, p. 120). These recommendations furnish the impetus for the present study.

**4 Rationale**

To assess the claim that one can lie by implicature, empirical investigations such as those of Coleman & Kay (1981), Hardin (2010), Danziger (2010), Arico & Fallis (2013), Meibauer (2016b), and Rutschmann & Wiegmann (2017) can be of great benefit for two reasons. The first concerns the role of folk notions of lying, or the intuitions of naïve participants. As Arico & Fallis note, "it is not possible that competent speakers of the language are commonly and systematically incorrect about the ordinary usage of everyday terms like 'lying'" (2013, p. 808). In their study, participants overwhelmingly judged bald-faced lies[6] and proviso lies[7] as genuine instances of lying, providing support for theories that consider those cases to be lies. Similarly, if the views of a naïve population sampled indicate a consensus as to whether false implicatures are lies, this should inform our theories on this crucial point. In addition to this definitional pursuit, an experimental investigation of false implicatures can provide insights into the status of

---

[5] For others (e.g., Bach (2001)) WIS does not need to be propositional in this way.

[6] In bald-faced lies, it is mutually known by both speaker and hearer that the utterance is blatantly false.

[7] Proviso lies are lies accompanied by a proviso, such as a wink or an aside, to indicate that the speaker is not necessarily warranting the truth of what s/he says.

different types of implicatures as lies, and hence, the extent to which 'saying' is required for lying.

*4.1 Previous findings regarding the integration of GCIs into WIS*
In an experimental study intended to tease out whether various types of GCIs are part of WIS, and under what conditions, Doran et al. (2012) tested 11 types of GCIs, drawn from Levinson's (1995, 2000) GCI-typology based on three heuristics.[8] The Q-Heuristic ("What isn't said, isn't") deals with scalar and clausal implicatures, or, more generally, quantity implicatures (Geurts, 2010). For example, when Jasper says, in example (3) above, "I fixed a roof," the implicature that he did not fix 'the roof' arises because the stronger item ('the', which entails 'a') in the scale <the, a> was not used – and is therefore taken not to apply. The I-Heuristic ("What is expressed simply is stereotypically exemplified") licenses listeners to infer a rich stereotypical interpretation from a minimal description. Applying the I-heuristic to the same example adjusted accordingly, had Jasper said 'the roof' this would be strengthened to mean a contextually salient roof, viz. the roof of his own house. Finally, the M-Heuristic ("Marked message indicates marked situation") acts in a complementary way to the I-Heuristic, noting that if a speaker says something in an unusual manner, the hearer is licensed to assume that the stereotypical interpretation is not the case. Saying "Bill stopped the car" invokes the I-Heuristic (the hearer can assume that Bill did so in a stereotypical way, i.e., using the brake pedal); on the other hand, saying "Bill caused the car to stop" invokes the M-Heuristic, generating an implicature to the complement of the stereotype, that Bill did not stop the car in a stereotypical way (Levinson, 2000, pp. 31–39).

Doran et al. (2012) gave participants one-line facts and two-line dialogues, as in (6) below, an example of conjunction buttressing, and asked them to judge whether the underlined sentence was true or false as well as provide a 1–4 confidence rating in their answer:

(6)      Irene: I understand that George has had a really rough year.
Sam: Yeah. Last month, he lost his job and started drinking.
FACT: George started drinking on the 15th of last month and lost his job on the 20th of last month.

Each participant was instructed to respond in one of three conditions: baseline (simply asked to judge), literal (asked to judge *literally*), and Literal Lucy (asked to judge from the perspective of a fictitious character demonstrated during pre-experiment training to interpret non-literal uses of expressions literally). The higher the percentage of "false" responses, the more likely the GCI to be included in WIS. They found that the Literal Lucy condition, more than the other two, led participants to exclude GCIs from truth-conditional content. Within that condition, responses were averaged and then ordered to reveal a ranking of the different types of GCIs as to their likelihood of being included into WIS, based on how many "false" responses were given for each type of GCI. This analysis yielded two important findings: (1) that GCIs do not all behave the same in this regard (the results ranged from 63% to 15%, and covered much ground in between); and (2) that GCIs of the same type (corresponding to Levinson's Q-, I-, and M- heuristics) do not all behave the same. The Q-based GCIs did not pattern together, and neither did those identified as I-based or M-based (see Figure 1). These results lead us to hypothesize that lie judgments may display similar levels of variability across different implicature types.

---
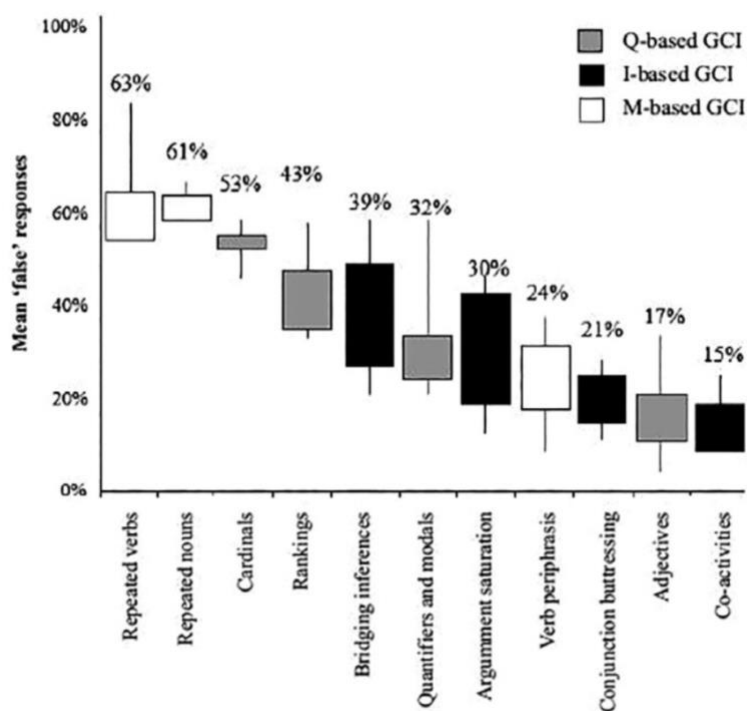
[8] PCIs were not tested in their study.

**FIGURE 1**   Boxplot of "false" responses in Literal Lucy condition reproduced by permission from Doran et al. (2012, p. 144)

It should be made clear from the outset that our adoption of the Levinsonian framework in this article does not reflect prior theoretical commitment but is rather motivated by the literature we are engaging with. Both Meibauer (2005, 2014) and Doran et al. (2012) use his tripartite GCI typology to organize their analyses. To enable a comparison between their results and ours, we do the same. Items called "implicature" under this framework may well not be considered implicatures in other frameworks,[9] and the GCI-PCI distinction itself is not built into all the theories. As such, everything we say below regarding implicatures/GCIs should be read as shorthand for "putative implicatures/GCIs;" in fact, our own experimental findings will lead us to consider this possibility for at least some of the items we examine. With these considerations in mind, we set out to address experimentally the following research questions:

   RQ1 – Do people consider various false implicatures to be lies?
   RQ2 – Do implicatures based on the same Levinsonian heuristic (Q, I or M) pattern together in terms of lie judgments?

## 5 The current study
### 5.1 Selection of stimuli
The GCI examples used in our study were based on the 11 types of GCIs in Doran et al. (2012), allowing for a direct comparison between our and their results. Their classification scheme, based on Levinson's heuristics, included four Q-based GCI types, four I-based GCI types, and three M-based GCI types, with four examples per type of implicature (shown in Table 1).

---

[9] See Horn (2004: 21-22) for insightful discussion of some of these differences.

**TABLE 1**   List of generalized conversational implicature (GCI) types used by Doran et al. (2012) with corresponding heuristics from Levinson (2000)

| GCI type name | Heuristic |
| --- | --- |
| Quantifiers and modals (e.g., "He ate most of the cake" +> not all of the cake) | Q |
| Cardinals (e.g., "She has three children" +> exactly three children) | Q |
| Gradable adjectives (e.g., "It was hot in Barcelona" +> not scorching) | Q |
| Rankings (e.g., "He tried to climb the mountain" +> he did not succeed in climbing) | Q |
| Argument saturation (e.g., "She bought a laptop and sent me an email" +> using laptop) | I |
| Bridging inferences (e.g., "The picnic was awful. The beer was warm" +> beer at picnic) | I |
| Coactivities (e.g., "Bob and Vicki are getting married" +> to each other) | I |
| Conjunction buttressing (e.g., "She got married and had a baby" +> in that order | I |
| Verbal periphrasis (e.g., "She made the car come to a halt" +> in an unconventional way) | M |
| Repeated verb conjuncts (e.g., "He slept and slept" +> for longer than is normal) | M |
| Repeated noun conjuncts (e.g., "He drank bottles and bottles of water" +> more than 2) | M |

*Note.* The examples are formatted: Implicature Type (e.g., "Utterance" +> implicature).

Of the 11 GCI examples in our study, seven were taken directly from their materials[10] and another three were closely based on them with modifications to more naturally generate false implicatures. The Quantifiers and Modals example was adapted from examples by Meibauer (2005, p. 1386; 2014, pp. 128–130) in order to test his prediction directly; this story is in line with those used by Doran and colleagues (2012) for this category. Meibauer (2014, pp. 129–135) states that false GCIs generated by all three heuristics should be able to count as lies.

In addition to the 11 types of GCIs, we tested four types of PCIs, corresponding to the four categories of Gricean maxims (Quality, Quantity, Relation, and Manner), since the resulting false implicatures are also claimed to be lies by Meibauer (2014, pp. 134–135). The Quantity-based stimulus was taken from Meibauer (2005, p. 1388), the Quality-based stimulus from Falkenberg (1982, p. 126, cited in Meibauer (2005, p. 1394), the Relation-based stimulus from Coleman & Kay (1981, p. 31), and the Manner-based stimulus from Levinson (1983, p. 108). These exact Quantity, Quality, and Relation implicatures are treated by Meibauer (2005, pp. 1388–1396) as lies.

*5.2 Norming*
The selected materials were first normed on a small number of naïve participants to ensure that they do indeed give rise to the predicted implicatures. For this, two-line dialogues were placed in a neutral context that didn't bias participant judgments about the existence or truthfulness of the implicature; the expected implicatures were always based on the second speaker's utterance. An example from the norming study (I-based GCI, Argument Saturation) is given in (7) below:

(7)      Rumors have spread about an incident in the art studio yesterday. Alex and Mark are discussing the incident.

---

[10] The stimuli taken directly are the Gradable Adjectives, Rankings, Argument Saturation, Bridging Inferences, Coactivities, Verbal Periphrasis, and Repeated Noun Conjuncts stories from this study. These correspond, respectively, to A9, A14, A17, A22, A26, A34, A42 from Doran et al. (2012).

Mark: I heard Sarah had a meltdown in the art studio yesterday! What happened?
Alex: You should've been there! In a fit of rage, Sarah picked up a hammer and broke a statue.
How likely is it that Sarah used the hammer to break the statue?

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| Extremely unlikely | Somewhat unlikely | Somewhat likely | Extremely likely |

The participants in the norming study (N= 14, age range = 20-26) were told to assume that the speakers in the dialogue were being completely truthful. They were then asked 'how likely is X?', where X was an action or fact following specifically from the implicature of the utterance (rather than from what the speaker had literally said), and directed to choose from among "Extremely unlikely," "Somewhat unlikely," "Somewhat likely," and "Extremely likely" on a 4-point Likert scale. A 4-point scale was chosen to allow participants to indicate the direction and strength of their judgment – if an item was rated in the wrong direction by more than four of the fourteen participants, it was flagged. Based on initial responses, four of the original stimuli were either slightly altered or totally rewritten (though still based on examples from the literature), and a survey containing just these four items was administered to a group of ten new participants. After two rounds of testing, all dialogues were judged to be giving rise to the implicature (over 70% of ratings in the expected direction), suggesting that each of the 15 implicatures is reliably derived from the corresponding utterance.

*5.3 Materials*
The 15 dialogues were each placed in three different context conditions: one biasing toward interpretation of the target line as not a lie (henceforth, condition T), one biasing toward interpretation of the target line as an unequivocal lie (condition L), and one in which the target line gives rise to an implicature which is false in context (condition FI).

In the previous paragraph, we describe these three conditions as "not a lie," "unequivocal lie" and "false implicature" for a couple of reasons. First, as the various components of the three-pronged definition suggest, whether something is considered a lie hinges on more than just factual falsity. Therefore, what we asked participants to provide were lie judgments (participants were asked "Did [name of speaker] lie?") rather than truth judgments or judgments about truth-speaking. Furthermore, each of these conditions reflects a different pairing of true and false propositions (in line with the pairings outlined by Meibauer 2005, p. 1380). In the T condition, both the proposition asserted and the implicature are true, in the L condition both the proposition asserted and the implicature are false,[11] while in the target FI condition, the proposition asserted is true but the implicature thereby generated is false (Table 2). A fourth possibility – that the proposition asserted is false while the implicature is true – is subsumed in the present set-up under the FI condition as a case of irony (see discussion in Section 6.2).

---

[11] The fact that in the L condition the proposition asserted is itself false can raise doubts about the extent to which the implicature, which is also false, is even derived in this case. In a framework such as Levinson's, where GCIs are licensed by particular lexical items all else being equal, a sentence containing such items (and the corresponding proposition) should in principle give rise to the false implicature even if the proposition licensing the implicature is itself false. In contextualist frameworks, however, the derivation of the implicature could be hampered by the falsity of the proposition asserted, if the false implicature is felt not to be relevant or activated in this context. While this observation may shed light on some of our findings, we do not pursue this line of explanation here, as a different experimental set-up is needed to test the relevant predictions.

**TABLE 2**  The three context conditions into which the stimuli dialogues were placed

| Context condition | Assertion $p$ | Implicature $q$ |
| --- | --- | --- |
| Not lie (T) | True | True |
| False implicature (FI) | True | False |
| Lie (L) | False | False |

To illustrate, we repeat the dialogue in (7) placed in the three context conditions to create the three stories in (8a–c) below:

(8a)   T: Rumors have spread about an incident in the art studio yesterday. Alex was in the studio all day and saw Sarah, frustrated with a project, pick up a hammer and use it to smash a statue to bits. The following day, Alex tells Mark about the incident.
Mark: I heard Sarah had a meltdown in the art studio yesterday! What happened?
Alex: You should've been there! In a fit of rage, Sarah picked up a hammer and broke a statue.

(8b)   FI: Rumors have spread about an incident in the art studio yesterday. Alex was in the studio all day and saw Sarah, frustrated with a project, pick up a hammer, walk over to a statue and kick the statue over with her foot, causing it to smash all over the floor. The following day, Alex talks about the incident.
Mark: I heard Sarah had a meltdown in the art studio yesterday! What happened?
Alex: You should've been there! In a fit of rage, Sarah picked up a hammer and broke a statue.

(8c)   L: Rumors have spread about an incident in the art studio yesterday. Alex was in the studio all day and saw Sarah, frustrated with a project, fling a paintbrush across the room, breaking a window. Later that night in the studio, Alex accidentally tripped over a statue, causing it to smash all over the floor. The following day, Alex talks about Sarah's incident.
Mark: I heard Sarah had a meltdown in the art studio yesterday! What happened?
Alex: You should've been there! In a fit of rage, Sarah picked up a hammer and broke a statue.

Three such contexts were developed for each of the 15 target dialogues (see Appendix A for a full list of stimuli) and divided up into three lists in such a way that each list contained only one version of each dialogue for a total of five stories in each of conditions T, L, and FI.

*5.4 Participants*
Sixty-six native English-speaking participants were recruited on Amazon Mechanical Turk (https://www.mturk.com/) and compensated for their participation. Data from six participants were excluded (see Section 5.5), leaving a final pool of 60 participants (34 M, 26 F; mean age = 37.22, sd = 11.08). The survey itself was hosted on Qualtrics (http://www.qualtrics.com/). Only survey takers with IP addresses in the United States were allowed to take the survey and the posting on the MTurk page called for native English speakers only. The average completion time of the survey was 7 minutes and 54 seconds.

*5.5 Experimental task*

Each participant saw only one version (T, FI, L) of each of the 15 dialogues, which were counterbalanced across lists. This resulted in each story being seen by a total of 20 participants in each context condition. The order of the stories within each list was randomized for each participant.

After seeing each story (e.g., one of 8a, or 8b, or 8c above), participants were instructed to answer the question "Did [name of speaker] lie?" (e.g., "Did Alex lie?" in 8 a-c above) by selecting a number along a 1–7 scale (e.g., Coleman & Kay, 1981; Arico & Fallis, 2013) where 1 was labeled "definitely not a lie," 4 was labeled "neither a lie nor not a lie," and 7 was labeled "definitely a lie." These three numbers were the only ones on the scale that were labeled.[12] The 7-point scale was adopted from previous studies that have used the Lie Judgment task.[13]

Though the dialogues in condition FI (false implicature) are the critical target items for the research questions, inclusion of the T and L conditions is important for several reasons. First, these can help us establish baselines for not lying (T) and lying (L) judgments for each story, such that judgments in the FI condition can then be compared with these baselines. They also serve as controls; that is, attention-checking stimuli to help monitor participants' engagement with the task -- only data by participants who made a genuine attempt at answering the questions were included in the final tallies. If a participant's answers appeared to be random, with three or more high ratings for stories in T condition or low ratings for stories in L condition, indicating that the participant had either not understood the task or was not attending to it, the data were excluded. This resulted in data from six participants being excluded from further analysis.

## 6 Results
### 6.1 Overview of Results
Average ratings of each FI (false implicature) condition story are shown in Figure 2 and in Table 3. A rating of 4, the midpoint on the Likert scale, represented "neither a lie nor not a lie" and is displayed as the horizontal black line across the graph. Bars are shaded according to type of implicature (PCIs and Q-, I-, and M- based GCIs).

---

[12] In a follow-up study, participants (N=25?) were presented with (xx?) of the original scenarios and a sliding scale on which to indicate their judgements, with only the end-points of the scale labelled as "not a lie" (0) and "definitely a lie" (7). Participants' judgements using a sliding scale closely mirrored the results obtained using an ordinal one in our main experiment, enhancing/buttressing/boosting (none of these) our confidence in the results presented here.

[13] Coleman & Kay decided on a 7-point scale in order to give participants enough room to indicate the strength of their judgment (a 5 and a 7 are in the same direction – both lies – but the difference between them reflects the strength of the participant's judgment) (Coleman & Kay, 1981, p. 30).
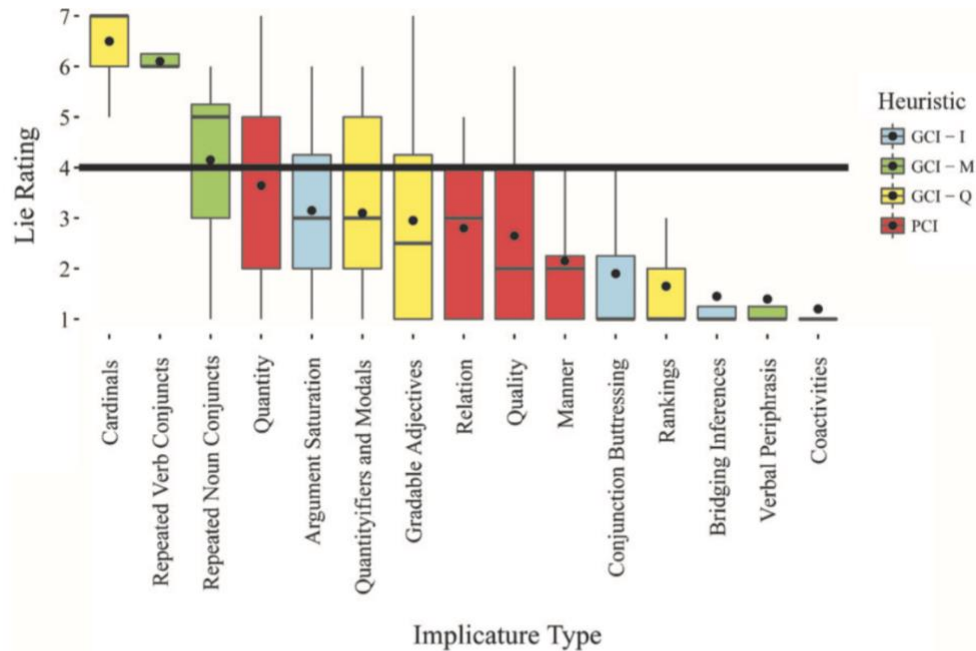
**FIGURE 2**    Rating for each story in condition FI (false implicature) averaged across 20 participants (between-participants analysis). Shading indicates to which heuristic each story belongs. Black dots indicate the mean rating for each type; the horizontal line in each column indicates median; box length extends to first and third quartile range; and whiskers demonstrate the range (excluding statistical outliers). The horizontal black line across entire graph indicates the midway cutoff between "lie" and "not a lie"

**TABLE 3**    Mean ratings (on a scale of 1–7) and standard deviations, in false implicature (FI) condition, for each implicature type tested

| Dialogue (type) | Mean rating | Standard deviation |
|---|---|---|
| Cardinals (GCI-Q) | 6.5 | 0.975 |
| Repeated verb conjuncts (GCI-M) | 6.1 | 0.624 |
| Repeated noun conjuncts (GCI-M) | 4.15 | 1.796 |
| Quantity (PCI) | 3.65 | 1.768 |
| Argument saturation (GCI-I) | 3.15 | 1.558 |
| Quantifiers and modals (GCI-Q) | 3.1 | 1.670 |
| Gradable adjectives (GCI-Q) | 2.95 | 1.910 |
| Relation (PCI) | 2.8 | 1.435 |
| Quality (PCI) | 2.65 | 1.796 |
| Manner (PCI) | 2.15 | 1.351 |
| Conjunction buttressing (GCI–I) | 1.9 | 1.338 |
| Rankings (GCI-Q) | 1.65 | 0.853 |
| Bridging inferences (GCI-I) | 1.45 | 0.973 |
| Verbal periphrasis (GCI-M) | 1.4 | 0.916 |
| Coactivities (GCI-I) | 1.2 | 0.400 |

The grand average rating for all T condition stories was 1.41 (sd = 0.44) and for all L condition stories was 6.18 (sd = 0.78). This suggests that these were correctly identified by participants as not lies and lies, respectively. For comparison, the grand average rating for all FI condition stories was 3.01 (sd = 1.60).[14] This gives some indication that the FI condition stories tended to fall in between the T and L condition stories, though closer to the "not lie" end of the spectrum. Significantly, there was less consensus about the FI condition stories than about the T or L condition stories. It is also evident from Figure 2 that implicatures falling under the same category, according to the classification used, do not pattern together in this task; instances of the same category are scattered across the scale, not clustered together. This finding mirrors the results of Doran et al. (2012), who similarly found a lack of clustering along Levinsonian-typology lines.

We fit a linear mixed-effects model using the lme4 package (Bates et al., 2014) in R (R Core Team, 2017) with Condition and Dialogue as fixed effects and random intercepts for Participant. Raw ratings were converted to Z-scores prior to running the model. The results of this model indicated that Condition, Dialogue, and interaction of the two were all significant effects (all p-values < .001, generated via the afex package (Singmann et al., 2015) in R). A post-hoc test was run to generate pair-wise comparisons among the three conditions of each dialogue, using the lsmeans package (Lenth, 2015). This allows a direct comparison between a false implicature (FI condition) and that exact same dialogue in the T or L condition. The p-values are corrected for multiple comparisons via Tukey method. Additionally, one-sample t tests were run for each story to assess if each story is significantly greater or less than 4, the midpoint of the scale. Table 4 shows a summary of the results of these statistical tests.

When comparing the FI (false implicature) condition stories to the T (not lie) condition stories,[15] more than half (8/14) of the stories did not achieve significance, meaning that for these stories, the average rating in the FI condition was not significantly different from the average rating of that same dialogue in the T condition. In other words, these stories were not judged as lies. For the remaining six stories, the comparison between the FI condition and the T condition reached statistical significance, meaning these stories were rated as significantly different from truth-speaking. In a comparison of the FI (false implicature) condition to the L (lie) condition, 11 out of 14 stories did achieve significance, meaning the dialogue in the FI condition was judged to be significantly different from the same dialogue in the L condition. Only three of the 14 stories were not significantly different across these two conditions.

---

[14] Ratings for the Quality implicature type are excluded due to considerations presented in Section 6.2.

[15] Quality has been excluded from these tallies as well, due to stimulus-based considerations explained in Section 6.2.

**TABLE 4** A summary of the results for each implicature type

| Implicature | Sig. diff. from scale midpoint | Sig. diff. from T condition | Sig. diff. from L condition |
|---|---|---|---|
| Cardinals (Q) | Greater*** | *** | — |
| Repeated verb conjuncts (M) | Greater*** | *** | — |
| Repeated noun conjuncts (M) | — | *** | — |
| Quantity (PCI) | — | *** | *** |
| Argument saturation (I) | Less* | * | *** |
| Quantifiers and modals (Q) | Less* | *** | *** |
| Gradable adjectives (Q) | Less* | — | *** |
| Relation (PCI) | Less*** | — | *** |
| Quality (PCI) | Less*** | N/A | N/A |
| Manner (PCI) | Less*** | — | *** |
| Conjunction buttressing (I) | Less*** | — | *** |
| Rankings (Q) | Less*** | — | *** |
| Bridging inferences (I) | Less*** | — | *** |
| Verbal periphrasis (M) | Less*** | — | *** |
| Coactivities (I) | Less*** | — | *** |

*Note.* The scale midpoint, 4, is the midway cutoff between a lie and not a lie. "T Condition" and "L Condition" refer to the T and L versions of each story.
*Significant level at $p < .05$.
**Significant level at $p < .01$.
***Significant level at $p < .001$.
— Significant level at $p > .05$.

Overall, results from these pairwise comparisons indicate that the stories in the false implicature (FI) condition were judged to be more different from the stories in the corresponding lie (L) condition than from those in the not lie (T) condition. Three of these — Cardinals, Repeated Verb Conjuncts, and Repeated Noun Conjuncts — were judged to be significantly different from their T condition counterparts and not significantly different from their L condition counterparts, suggesting that these false implicatures counted as lies to participants.[16] Eight (Gradable Adjectives, Rankings, Bridging Inferences, Coactivities, Conjunction Buttressing, Verbal Periphrasis, Relation, Manner) of the remaining 11 stories were judged to be significantly different from the L condition counterpart and not significantly different from the T condition counterpart. A further three stories (Quantity, Quantifiers & Modals, Argument Saturation) were judged to be significantly different from both the T and L condition counterparts.

*6.2 Flouting Quality: the case of irony*
Quality has been excluded from the tallies and averages above due to stimulus-based considerations. Specifically, for the other 14 stimuli, stories in the T condition are those in which both the assertion *p* and the implicature *q* are true (see Table 2 above). However, in order to have a Quality PCI that would be rated uniformly as not a lie, as required in our T condition, the assertion in the T condition must be false on the surface. A dialogue illustrating this kind of Quality PCI is shown in (9) below. This item is taken from Meibauer (2005), who in turn took it from Falkenberg (1982), who found it in an Agatha Christie novel. Charles crashes his car

---

[16] Doubts can be raised as to whether these categories are truly implicatures. See section 8.3 for further comment.

intentionally into a wall in order to pretend that he had an accident. When asked if he had an accident, he responds by stating that he crashed his car into the wall deliberately in order to implicate, falsely, that he had an accident.

(9)     Jim: What happened? Did you have an accident?
        Charles: Nooooo, I crashed my car deliberately. (*p*)
        +> I did not crash my car deliberately (*q*)

This example is treated as an instance of irony. However, contrary to typical cases of irony, where the speaker says what he believes to be false in order to implicate something true, in (9) Charles says something which is in fact true (although intended to be taken as ironic/false by the listener) in order to implicate something false (that he did not crash his car deliberately). This is the version of the story corresponding to the false implicature (FI) condition, which is explicitly stated by Meibauer (2005, p. 1394) to be a lie. On the other hand, if the car crash truly was an accident, then Charles's assertion is false and the implicature is true. This is the version of the dialogue that is a pure case of irony and should get rated as not being a lie, so this is the version we used in our T condition. To guide participants toward the ironic interpretation of the utterance in this condition, the non-conventional spelling "Nooooo" and a reference to Charles's proclivity for sarcasm were utilized in the stimulus. The examples in (10 a–c) below show the three context conditions in which the dialogue in (9) was embedded.

(10a)   **T**: Charles is driving home when his car is hit by another car, spinning his car out of control and landing him headfirst in the wall of a building. The other car drives away, and Charles gets out of his. Jim is the next person to drive by, and he pulls over to ask the ever-sarcastic Charles what happened.

        Jim: What happened? Did you have an accident?
        **Charles: Noooo, I totally crashed my car deliberately!**

(10b)   **FI:** Charles, eager for sympathy, decides to fake a car crash to make it appear as though he had an accident. When nobody is around, he drives his car headfirst into the wall of the building, and then gets out of the car. Jim is the next person to drive by, and he pulls over to ask the ever-sarcastic Charles what happened.

        Jim: What happened? Did you have an accident?
        **Charles: Noooo, I totally crashed my car deliberately!**

(10c)   **L**: Charles, walking home, rounds a corner and comes across a car smashed headfirst into the wall of a building. He looks inside and sees nobody in the car. While he is investigating, Jim drives by. Jim pulls over to ask the ever-sarcastic Charles what happened.

        Jim: What happened? Did you have an accident?
        **Charles: Noooo, I totally crashed my car deliberately!**

The FI condition story received a mean rating of 2.65 (sd = 1.80), falling short of the midway cutoff of 4 on the 7-point scale used and indicating that it was not judged as a lie by participants overall. The T and L conditions, unsurprisingly, did not pattern the same as the other dialogues in this study. Instead, both received mean ratings >4, (T mean rating = 4.35, sd = 1.74; L mean rating = 4.95, sd = 2.13), suggesting that they were both perceived as lies, despite one of them (the T condition) being merely ironic. Additionally, this L condition rating is lower than ratings for other stories in this condition. A lack of consensus on this item is indicated by the high standard deviations in each condition. Due to the complications introduced by the use of irony in this example, the Quality PCI is excluded from further discussion in the present results.[17]

*6.3 Summary of results*
The results of our experimental investigation suggest that people do not, on the whole, judge false implicatures to be lies in the same way as Meibauer (2005, 2011, 2014) does. The average of the 14 stories in FI condition lies between the averages for the T and L conditions and, importantly, sits closer to the T condition. That trend persists when each story is examined individually. To say that false implicatures are absolutely not lies, however, is an unfounded generalization. According to the metrics used here, only two stories in the study were strongly rated as lies, nine were strongly rated as not lies, and four landed somewhere in the middle. Overall, then, these results provide no evidence for a blanket statement that "false implicatures are lies." In the following sections, we examine in greater detail the role of the speaker's lying intention and of notions of what is said in our results.

**7 Monitoring for speaker intent**
A potential complicating factor with the above results is that participants may have been paying less attention to the phrasing of the assertion and instead monitoring the speaker's intent to lie in each context condition. If certain situations naturally led the character to lie or not to lie, and this attribution of a lying intention to the speaker is what drove participants' judgments in the experimental task, then the attribution of a lying intention could be a confounding factor in our results that we did not control for. In other words, despite our wish to test for false GCIs generated "based not on direct computations about speaker-intentions, but rather on general expectations about how language is normally used" (Levinson, 1995, p. 93), participants may have been focusing precisely on the presence of a contextually set up lying intention, effectively converting the task into one about false PCIs (cases where the presence of a false GCI is buttressed by the presence of an intention to deceive). This would turn our examples into cases of rhetorical cooperativity, that is, instances where the relevant implicatures, albeit natural and salient, are "unsafe" to draw in context (Asher & Lascarides, 2013, p. 17).

To examine this possibility, we carried out a post-hoc test designed to gauge if participants were attributing to the speaker an intention to lie based on the context alone (without the dialogue). Forty participants recruited through MTurk saw the 15 FI condition contexts (in one of two lists, such that each participant saw either seven or eight FI condition contexts and 20 participants saw each context once) with the final two lines of dialogue removed, and asked to

---

[17] Of course, irony can take many forms, and a prima facie explanation of our findings could hinge on the fact that speakers can have different grounds for calling something a lie, ranging from factual falsity (captured in our T condition here) to intention to deceive (captured in our L condition). See Dynel (2016) for a theoretical discussion of the many forms irony can take. Future experimental research is needed to investigate the possibilities of lying through irony.

judge on a 1–4 Likert scale[18] ("extremely unlikely," "somewhat unlikely," "somewhat likely," "extremely likely") how likely it is that [character] would try to lie in this situation. An example is given in (11) below.

(11)     John and Mary have recently started dating. Gabe is Mary's ex-boyfriend. Gabe has been sick with the flu for the past two weeks, but went on a date with Mary last night. **How likely is it that Mary will try to lie to John about it?**

Figure 3 shows the results of this post-test (black line) alongside the FI condition ratings obtained from the main lie judgment task (grey bars) in. Higher ratings indicate a stronger attributed intent to lie in the post-test and a higher lie rating in the main lie judgment task.
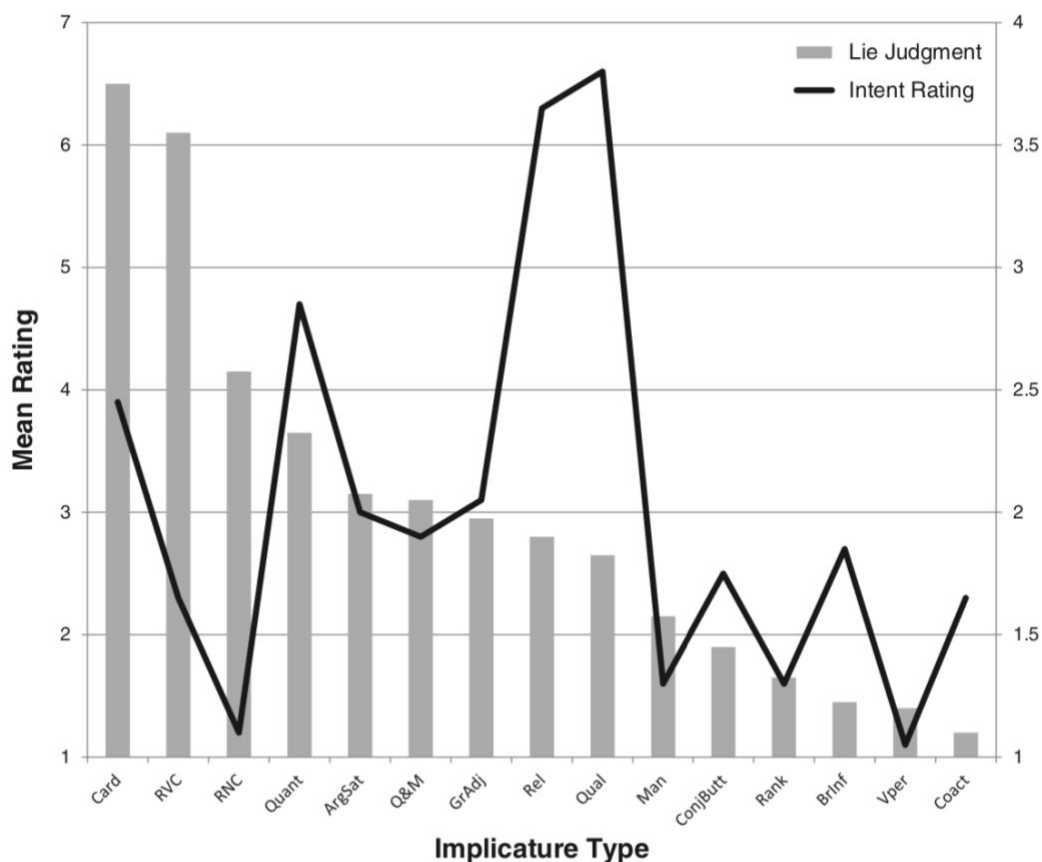


**FIGURE 3**     Lie judgment (main task) versus Intent rating (post-test) for each false implicature (FI) condition story

Kendall's Tau coefficient was computed for the two sets of data (from the post-test and the main lie judgement task) to test for the correlation between the two. If lie judgments were indeed driven by attributed lying intent in our previous results, the correlation coefficient should

---

[18] A 1-4 scale was used here to force participants to choose (there was no neutral option) and indicate the strength of their judgment. To recap, the post-test and norming study used 1-4 Likert scales, while the lie judgment task used a 1-7 scale (modeled after Coleman & Kay (1981)) – the scale for each survey was carefully selected with regards to the specific task within.

approach 1. For these two metrics, the correlation coefficient is 0.26 (p = 0.1804), with no evidence suggesting a significant correlation between the two. Thus, while attributed intent to lie clearly varied between our stories, that was not what was driving participants' judgments of the stimuli in the FI condition as lies or as not lies. For instance, Repeated Verb Conjuncts (one of the two stories rated consistently as a lie in the FI condition in the main lie judgement task) had a mean rating of only 1.65 out of 4 in the post-hoc test. On the other hand, the Relation story (mean rating = 3.65 out of 4 in the post-hoc test) was judged to be not a lie in condition FI in the main lie judgement task (mean rating = 2.8 out of 7). This is consistent with the results of Experiment 1 of Wiegmann et al. (2016), which demonstrated that blameworthiness for deceiving had no significant effect on lie judgment ratings.

These results indicate that participants' judgements in the main lie judgement task cannot be explained away by supposing that stimuli were accidentally embedded in contexts where a speaker was expected to lie. Since no consistent correlation can be established between the main and post-hoc test results, we conclude that participant judgments in our main study are not motivated by the speaker's lying intention.

## 8 Discussion
### 8.1 Lying and misleading
Our results advocate against a blanket statement that false implicatures are lies, as claimed by Meibauer (2014, pp. 134–135). Only two out of 15 false implicature stories garnered averages that were significantly higher than 4, the midway cutoff between a lie and not a lie, while 11 had average ratings significantly less than 4. Meibauer's predictions suggest that all should be above the cutoff; this is most certainly not borne out in our results.

Nonetheless, although a blanket statement in this regard is not possible, some false implicatures are considered lies. More specifically, our results suggest that false implicatures can be classified into three groups: those significantly different from truth-speaking, those significantly different from lies, and those significantly different from both truths and lies. Comparing each FI condition dialogue directly to that exact same dialogue in the other conditions, three cases of false implicature (FI) – Cardinals (Q), Repeated Verb Conjuncts (M), and Repeated Noun Conjuncts (M) – were judged to be significantly different from the truth (T) version and not significantly different from the lie (L) version. Two of these (Cardinals and Repeated Verb Conjuncts) were significantly higher than 4 as well. In other words, the three metrics corroborate that these two types of false implicatures are lies. On the other hand, eight types of false implicatures were judged to be significantly different from the lie version, not significantly different from the truth version, and significantly less than 4. These stories – Gradable Adjectives (Q), Relation (PCI), Manner (PCI), Conjunction Buttressing (I), Rankings (Q), Bridging Inferences (I), Verbal Periphrasis (M), and Coactivities (I) – can be considered, with statistical backing, not lies, according to our participants.

In addition to these unequivocal lies and straightforward truths, four types of false implicatures lie somewhere in the middle, making for an interesting case. The Repeated Noun Conjuncts story was rated significantly different from the T condition story and not significantly different from the L condition story, but not significantly greater than 4. The Argument Saturation and Quantifiers & Modals stories were rated as significantly lower than 4 and significantly different from the L condition story, but significantly different from the T condition story also. The Quantity story was rated as significantly different from both the T condition and L condition stories, and was not significantly different from 4, making it the prototypical "middle

ground" example in this study. One possible explanation for these results is that participants are distinguishing between "lying" and "misleading," as proposed, for instance, by Saul (2012, pp. 37, 68, 117). Saul maintains that any communicated false content must be part of WIS for the utterance to count as a lie. If the utterance meets the other criteria for lying, but through implicature as opposed to WIS, Saul would classify this as "misleading" as opposed to "lying;" such an explanation can account for these middle ground stories in this study.

For a statement such as "I won the lottery and bought a house" to consistently be judged not a lie (mean rating = 1.9, significantly different from L condition rating) when it is true that the house was bought first, the GCI —the stereotypical assumption that the events happened in the order said — cannot be what participants are assessing. In the norming study, all participants indicated that this utterance means that the events were more likely to have happened in this order, so there is no question that the desired implicature is being generated. To explain the finding that participants in the main study do not rate the utterance as a lie based on the falseness of the implicature, we are compelled to acknowledge that, when asked whether the speaker in this situation has lied, participants are judging only the literal meaning of the utterance: it is true that I won the lottery and it is true that I bought a house. Since both are, in fact, true, the utterance is judged not to be a lie. If the implicature were what was being judged, the ratings should have been higher (closer to 7, the lie endpoint of the scale).

These results are compatible with Michaelson's (2016) Lying Test, which uses lie judgments to determine the semantic content of an utterance. According to this test, if some content is believed to be false and delivered with the intent to deceive the listener, yet the overall utterance is judged not to be a lie, then that content is not part of the semantic content of the utterance (Michaelson, 2016, p. 482). If there is some utterance U with literal semantic content $p$ and implicated content $q$ which the speaker believes to be false, and the utterance is judged to not be a lie, it follows from Michaelson's argument that $q$ is, indeed, not part of the semantic content of U. In Michaelson's Lying Test, that false implicatures are not lies is assumed in order to determine differences between what is said and what is implicated; our results provide empirical support for this assumption.

Our results thus indicate that the literal meaning of an utterance is what people tend to consider when judging if an utterance is a lie or not. This is in line with Saul's (2012) proposal, according to which lying applies only to WIS, and WIS is the minimally enriched meaning, with contextual information filled in *only if* the literal WIS is not truth-evaluable (see Section 3). If a false proposition comes about as part of this definition of WIS, it can be deemed a lie; if a false proposition comes about as part of any further enrichment of the utterance, the utterance is not considered a lie but, rather, simply misleading.

With the current results trending toward discounting Meibauer's theory that false implicatures are lies, a follow-up analysis of the results can be conducted adopting Saul's theory. Operating under this assumption, it is possible to reverse direction and use the lie judgment task to analyze what is considered to be part of WIS.[19] Those stories that were rated consistently as not lies (most of the stories in the experiment) would also not count as part of WIS, according to Saul. Furthermore, this would mean that the types of false implicatures that were treated as lies

---

[19] We adopt the binary what is said/what is implicated distinction from Levinson (1995; 2000) and Doran et al. (2012) to preserve comparability with their data and discussion, on which we build, although we are aware that many examples of GCIs are regarded, in other frameworks, as cases of implicature (e.g., Bach, 1994; 2001) or enrichment (e.g., Carston 1995; 2002; Récanati, 2001).

(Cardinals and Repeated Verb Conjuncts) would actually be part of WIS. The following section discusses the implications of this analysis.

*8.2 Lying, what is said, and different types of implicatures*
Given the aforementioned discussion, our results can also be used as an alternate avenue to test what counts as part of WIS, a question tackled experimentally by Doran et al. (2012; see Section 4.1). The experimental task in that study was a true/false judgment task given a two-line dialogue and a fact about the situation, such that the judgment task implicitly measured if participants were able to separate implicated meaning from meaning said explicitly for different types of GCIs. In asking participants to determine if someone has lied for different types of GCIs (as well as for PCIs), we likewise tested different types of GCIs using an experimental task, and the results are consonant to a large extent (see Figure 4).
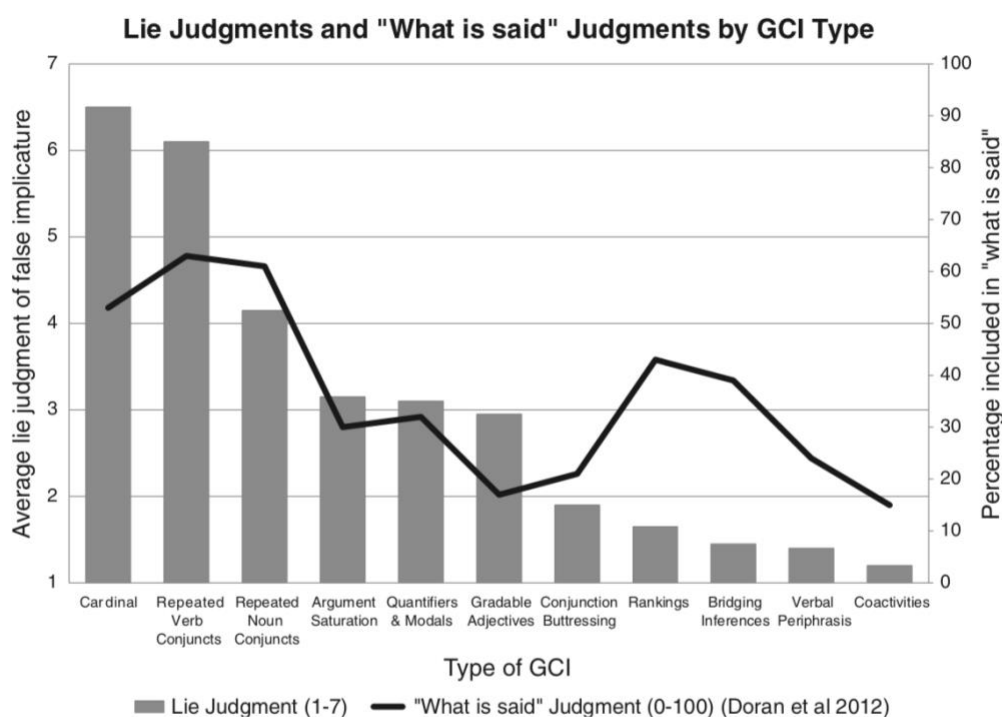


**FIGURE 4** Results of the current study with results from Doran et al. (2012) superimposed. Since we used a 1–7 scale and the data from Doran et al. (2012) were presented as percentages (their task was a binary truth value judgment task), these scales are shown as two different vertical axes to the left and to the right of this figure, respectively

The results from Doran et al. (2012) shown in Figure 4 are those in the Literal Lucy condition, in which participants were asked to judge from the perspective of a fictitious character who always interprets utterances literally. These results may be said to reflect people's intuitions when pushed to sharpen their judgments about what is (literally) said. Kendall's Tau coefficient was computed to test for a correlation between the results of Doran et al. (2012) and our own. The correlation coefficient was 0.49 (p = 0.04), indicating that there is a significant correlation

between the two sets of results.[20] The lie judgment task thus seems to provide an alternate avenue for testing where people draw the line between what is said and what is implicated *when asked to think explicitly about that boundary* (for instance, when it is morally consequential to do so, as in cases of lying). Comparing the false implicature lie ratings to two other metrics – contextually set up lying intention (in our post-hoc test) and how close an implicature is considered to WIS (in the Doran et al. (2012) study) – revealed that the latter is a much better correlation than the former, bolstering the idea that, in order to be considered a lie, something must be a part of WIS but not necessarily uttered with an intention to deceive;[21] in other words, aspects of the language used rather than the speaker's lying intention drive characterization of a false implicature as a lie.

Expanding on these findings, different types of GCIs (i.e., those based on the Q-, I-, or M-Heuristic from Levinson's framework) did not pattern together neither in the Doran et al. (2012) study nor in our own. This now-replicated finding can be taken to indicate that Levinson's GCI typology does not accurately capture the ways in which the said/implicated distinction is drawn by participants in experimental tasks in which that distinction is relevant. If one implicature based on a Levinsonian heuristic is consistently and confidently rated as a lie, and another, based on that exact same heuristic, is consistently and confidently rated not a lie, one might wonder if these two implicatures should indeed be accounted for by the same mechanism (simply because their semantics allows it; see the discussion about Cardinals below) or whether this move overshadows important differences between them that would warrant differential treatment. In our results, for example, the Cardinals example (mean rating = 6.5) behaves entirely differently from other Q-type implicatures, such as Rankings (mean rating = 1.65). If these are derived from the same theoretical foundation, why is one treated as definitively a lie and the other as definitively not a lie? This line of reasoning challenges the *raison d'être* of the Levinsonian typology and, in the case of Cardinals, calls for a redrawing of the line between WIS and GCIs itself.[22]

### 8.3 False implicature that are lies

In itself, the fact that the Cardinals story stood out and was rated so highly as a lie may not be so surprising. Cardinal terms (two, three, four, etc.) have been traditionally considered scalars similar to <all, some> or <hot, warm> (e.g., Horn, 1972; Levinson, 2000) but this idea has been challenged (e.g., Horn, 1992 and 2004, p. 20 for a summary of the linguistic arguments against this view), most recently by empirical evidence (e.g. Papafragou & Musolino, 2003; Bultinck,

---

[20] Raw data from the Doran et al. experiment (Doran, personal communication) were obtained and the same correlation test was run with the *literal* and *baseline* conditions as well. The correlation coefficient between the current results and the Doran et al. *literal* condition was 0.44 (p = .06) and between the current results and the Doran et al. *baseline* condition was 0.45 (p = .06).

[21] Although the two datasets follow the same linear trend, and there is a statistically significant correlation between the two, there are deviations (e.g. Cardinals, Rankings). This can potentially be attributed to the fact that the present story tested only one example of each implicature type – further testing could lead to even closer mapping of the two datasets.

[22] These considerations are of course only problematic for Levinson's theory if the theory is taken to address how *people* generate these interpretations when encountering these expressions in context. If, on the other hand, Levinson's theory is taken as an explanation of how *language* can generate implicatures, then participant judgements need not have a direct bearing on his typology. Levinson himself is somewhat ambivalent on this point when he writes: "Explicit processing considerations do not enter the framework offered here, but they do form part of the background of cognitive processing, for the character of the inferences in question as default inferences can, I think, be understood best against the background of cognitive processing" (1995, p. 5).

2005; Huang et al., 2013), and by theories positing that these terms are not scalar at all but instead possess underspecified or exact semantics (e.g., Carston, 1998; Breheny, 2008). According to the scalar explanation, *two* is stored with the truth conditional meaning "at least two," and "exactly two" is an interpretation arrived at through default pragmatic inferencing (by a combination of the lower-bound semantic meaning and the upper-bound implicature – no more than). The exact semantics explanation reverses that, stating that *two* means "exactly two," and the lower-bound meaning is what arises from pragmatic inferencing where appropriate. The underspecified explanation, on the other hand, considers that cardinals like *two* are inherently underspecified and their truth-conditional meaning is fixed locally as either "at least two", "at most two", or "exactly two" depending on context. The results of the present study can be taken to provide additional evidence against the scalar explanation, and are compatible with both the exact semantics and the underspecified proposals.

The false implicature Cardinals story, given in (12) below, was the highest-rated of all the false implicature types.

(12)    Jenn made a half dozen cupcakes in the morning and left them out to cool on the counter in her apartment. Her roommate, Molly, ate all of them. Jenn comes back in the evening to find that the cupcakes are gone, so she asks Molly about them.

Jenn: Did you eat my cupcakes?
**Molly: Yes, sorry, they were so good! I ate three**.

Although it can be taken as evidence against the scalar view of cardinal numbers, it is worth pointing out that there are a few other considerations that may have led to this story being rated so highly in this experiment. Firstly, the specific cardinal numbers in this story (*three* and *six*) are within the subitizing range, meaning the difference between three and six can matter quite a bit. If another scenario were crafted in which Molly worked at a commercial bakery that processed cupcakes in batches of hundreds every day and the question was about the batch currently in the oven, the exact number of cupcakes, being upwards of a hundred, may matter a lot less.[23] Another pertinent consideration is the fact that Molly's utterance is a direct response to Jenn's question, itself potentially an accusation, which may have prompted participants to treat Molly's utterance as a cover up for the truth, i.e. as a lie.[24]

The other two categories that received average ratings of >4, Repeated Noun Conjuncts ("boxes and boxes") and Repeated Verb Conjuncts ("drank and drank"), are not without controversy, either, although not to the same extent. Specifically, it is possible to treat these cases as constructions in English along the lines of Goldberg (1995), namely, an "X and X" construction, with subtypes "N and N" and "V and V," respectively. Notably, both of these examples involve reduplication, a quite common cross-linguistic phenomenon used to achieve a variety of effects, including plurality, continuous aspect, and augmentation and intensification (Moravcsik, 1978). Intuition suggests that verbal or nominal repetition in constructions such as "drank and drank" or "boxes and boxes" may be grammaticalized to the point where "an excessive or abnormal amount/number" could be considered part of these expressions' semantics, i.e., what they signify conventionally. Similar processes occur, for instance, in American Sign Language, in which repeating a signed lexical item indicates the repetition of the

---

[23] For a deeper look at the intricacies of numerical expressions, see Cummins et al. (2012) and Cummins (2015).
[24] *Acknowledgment omitted for anonymization purposes*

action that the sign refers to (Wilbur, 2005); in Estonian, in which reduplication of nouns and verbs indicates multiplicity and event repetition, respectively (Erelt, 2008); and in several pidgins and creoles (Bakker & Parkvall, 2005), among many others. All of these cases involve iconicity, which has also been discussed as a construction-driving property for Persian causatives (Salmani Nodoushan & Ghomshei, 2014) and cited as a force driving language change and grammaticalization (Haiman, 1983; Bisang, 2016). Further empirical testing and theorizing would be necessary to confirm the existence of a repeated conjuncts construction in English, but the groundwork exists.

Based on these results, it can be argued that the two types of "implicatures" discussed in this section are not implicatures at all. That is so also according to Michaelson's (2016) Lying Test, according to which if something is judged to be a lie, it must be part of the semantic content of the utterance (given that the conditions "believe false" and "intent to deceive" are satisfied; Michaelson, 2016, p. 482). The high degree of consensus on the Cardinals and Repeated Verb Conjuncts stories, then, according to Michaelson's line of argumentation, indicates that the relevant "implicatures" are part of the semantic content of these expressions, and therefore not implicatures in the first place.

## 9 Conclusions and directions for future research

In this study, we set out to investigate (1) if false implicatures are considered lies and (2) if implicatures based on the same Levinsonian heuristic pattern together in terms of lie judgements. To address these questions, we designed an experiment that asked participants to provide lie judgments of utterances that reliably induce false implicatures. In answer to the first question, our results provide evidence for Dynel's (2015) claim that implicatures are too diverse a class to be treated with a blanket statement, such as "false implicatures are lies." Although the same diversity prevents us from unequivocally claiming that "false implicatures are not lies," our results trend towards supporting that hypothesis rather than its opposite. The results of this lie judgment task correlate significantly with results from Doran et al.'s (2012) truth-value judgement task, suggesting that inclusion in what is said (as opposed to implicated) may be a crucial consideration in lie judgments. This is in line with the theory of lying proposed by Saul (2012) and runs counter to general claims by Meibauer (2005; 2011; 2015) that one can lie through implicature. In answer to the second question, our results indicate that different types of implicature behave differently in this lie judgment task; implicatures based on the same Levinsonian heuristic did not pattern together. Such a finding calls into question the Levinsonian GCI typology and the way the line between what is said and what is implicated is drawn therein.

In addition to the theoretical implications discussed above, our study opens up several avenues for future research. First, a replication of this study with more than one story per type of implicature[25] is called for to better and more comprehensively evaluate Meibauer's claim that all false implicatures are lies (see e.g., the discussion of scalar diversity in van Tiel et al. (2016). Assuming this kind of follow-up study reveals a level of consensus surrounding the robustness of different implicature types vis-à-vis the truth-lie distinction, the lie judgment task presented here can be used to explore alternative classifications of implicatures. Note that the finding that PCIs are interspersed with GCIs in our results raises the possibility that a distinction between the two types of implicatures is unwarranted in the first place, as posited, for instance, in Relevance Theory and more recent contextualist (e.g. Jaszczolt, 2005) and probabilistic (e.g. Degen & Tanenhaus, 2015) accounts. Another aspect calling for future study concerns the distinction

---

[25] Another important consideration is what is at stake in any particular story.

between lying and misleading (Saul, 2012). As already mentioned, Saul's claim that lying only concerns WIS is *prima facie* supported by our results: our participants tended to focus only on the lower-bound literal meaning of the utterance when determining whether a speaker had lied. Another part of her proposal, however, namely, that other "lies" communicated less explicitly are instances of "merely misleading," was not tested. In Weissman & Terkourafi (2017) we used a corpus methodology to provide preliminary empirical support for such a distinction. Future work could use these corpus results to design experiments that will tease out participants' intuitions regarding these two notions. It is likely that the exact phrasing of the options given to participants will play a prominent role in these future studies (Turri & Turri, 2015; Wiegmann et al., 2016).

Finally, deeper and more detailed research could turn to an enhanced sociolinguistic account of a definition of lying. Detailed information about ethnic affiliation, linguistic background, social upbringing, and dialectal differences, among other potential factors, could reveal subtle differences in how people judge lies and other types of deception – a point not unrelated to the different contents that moral behavior can assume in diverse contexts. Variation in lying judgements can be due to linguistic differences between the materials tested, but also to sociocultural differences between those providing the judgements. Therefore, both sources of variation must be tracked to disentangle their reciprocal effects on lying judgements.

## Supporting information
Additional supporting information may be found online at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/9BWDHB

## References
Adler, Jonathan E. 1997. Lying, deceiving, or falsely implicating. *The Journal of Philosophy* 94 (9), 435-452.

Arico, Adam and Don Fallis. 2013. Lies, damned lies, and statistics: An empirical investigation of the concept of lying. *Philosophical Psychology* 26 (6), 790-816.

Asher, Nicolas., & Lascarides, Alex. 2013. Strategic conversation. *Semantics and Pragmatics 6*, pp. 1-62.

Augustine. 1952. Lying (M. Muldowney, Trans.). In Roy J. Deferrari (ed.) *Treatises on various subjects*. New York: Catholic University of America. Pp. 53-120.

Bach, Kent. 1994. Conversational impliciture. *Mind & Language*, *9*(2), 124-162.

Bach, Kent. 2001. You don't say? *Synthese* 128.15–44.

Bakker, Peter & Mikael Parkvall. 2005. Reduplication in Pidgins and Creoles. In *Studies on Reduplication*, Bernhard Hurch (ed.), 511-532. Berlin: Mouton De Gruyter.

Bates, Douglas; Maechler, Martin; Bolker, Ben; and Walker, Steven. 2014. _lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, <URL: http://CRAN.R-project.org/package=lme4>.

Bisang, Walter. 2016. Linguistic change in grammar. In *The Routledge handbook of linguistics*, Keith Allan (ed.), 366-384. New York: Routledge.

Breheny, Richard. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25, 93-139.

Bultinck, Bert. 2005. Numerous meanings: The meaning of English cardinals and the legacy of Paul Grice. Amsterdam: Elsevier.

Carson, Thomas L. 2006. The definition of lying. *Nous* 40, 284-306.

Carston, Robyn. 1995. Quantity maxims and generalised implicature. *Lingua*, *96*(4), 213-244.

Carston, Robyn. 1998. Informativeness, relevance, and scalar implicature. In *Relevance theory: applications and implications*, Robyn Carston & Seiji Uchida (eds.), 179-236. Amsterdam: John Benjamins.

Carston, Robyn. 1999. The semantics/pragmatics distinction: a view from Relevance Theory. *The semantics/pragmatics interface from different points of view,* ed. by Ken Turner, 85 125. Oxford: Elsevier Science.

Carston, Robyn. 2002. Linguistic meaning, communicated meaning and cognitive pragmatics. *Mind & Language*, *17*(1-2), 127-148.

Coleman, Linda and Paul Kay. 1981. Prototype semantics: The English verb *lie*. *Language* 57 (1), 26-44.

Cummins, Chris. 2015. *Constraints on numerical expressions.* Oxford: Oxford University Press

Cummins, Chris, Uli Sauerland and Stephanie Solt. 2012. Granularity and scalar implicature in numerical expressions. Linguistics and Philosophy 35, 135-169.

Danziger, Eve. 2010. On trying and lying: Cultural configurations of Grice's Maxim of Quality. *Intercultural Pragmatics* 7 (2), 199-219.

Degen, Judith, & Tanenhaus, Michael. 2015. Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667-710.

DePaulo, Bella M.; Deborah A. Kashy; Susan E. Kirkendol; Melissa M. Wyer; and Jennifer A. Epstein. 1996. Lying in everyday life. *Journal of Personality and Social Psychology* 70 (5), 979-995.

Doran, Ryan; Gregory Ward; Meredith Larson; Yaron McNabb; and Rachel E. Baker. 2012. A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language* 88 (1), 124-154.

Dynel, Marta. 2011. Web of deceit: A Neo-Gricean view on types of verbal deception. *International Review of Pragmatics* 3, 139-167.

Dynel, Marta. 2015. Intention to deceive, bald-faced lies, and deceptive implicature: Insights into *Lying at the semantics-pragmatics interface* (Jörg Meibauer, 2014). *Intercultural Pragmatics* 12 (3), 309-332.

Dynel, Marta. 2016. Comparing and combining covert and overt truthfulness: On lying, deception, irony and metaphor. *Pragmatics & Cognition* 23 (1), 174-208.

Erelt, Mati. 2008. Intensifying reduplication in Estonian. *Linguistica Uralica* 44 (4), 268-277.

Falkenberg, Gabriel. 1982. *Lügen. Grundzüge einer Theorie sprachlicher Täuschung,* Tübingen: Niemeyer.

Fallis, Don. 2009. What is lying? *The Journal of Philosophy* 106 (1), 29-56.

Fallis, Don. 2014. Review of *Lying, Misleading, and What Is Said: An Exploration in Philosophy of Language and in Ethics*, by Jennifer M. Saul. *European Journal of Philosophy*, March 2014 Supplement, Vol. 21, e17-e22.

Geurts, Bart. 2010. *Quantity Implicatures*. Cambridge University Press.

Gibbs, Raymond W. and Jessica F. Moise. 1997. Pragmatics in understanding what is said. *Cognition 62* (1), 51-74.

Goldberg, Adele E. 1995. Constructions: A construction grammar approach to argument structure. Chicago: University of Chicago Press.

Grice, H. Paul. 1989. Logic and conversation. In *Studies in the way of words*, 22-40. Cambridge, MA: Harvard University Press. [Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.) *Syntax and semantics, Vol. 3: Speech acts*, 41-58. New York: Academic Press.

Hardin, Karol J. The Spanish notion of *lie*: Revisiting Coleman and Kay. *Journal of Pragmatics 42*, 3199-3213.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59, 781-819.

Horn, Laurence. 1972. *On the semantic properties of the logical operators in English*. (Doctoral dissertation). UCLA, Los Angeles, CA. Distributed by IULC, Indiana University, Bloomington, IN.

Horn, Laurence. 1992. The said and the unsaid. In *The Ohio State University Working Papers In Linguistics (40): Proceedings from the second conference on semantics and linguistic theory*, 163-192. Columbus, OH: The Ohio State University.

Horn, Laurence. 2004. Implicature. In Laurence Horn & Gregory Ward (eds.), *The Handbook of Pragmatics*, 3-28. Malden, MA: Blacwell.

Horn, Laurence. 2017. Telling it slant: Toward a taxonomy of deception. In Dieter Stein & Janet Giltrow (eds.), *The Pragmatic Turn in Law*, 23-55. Mouton Series in Pragmatics.

Huang, Yi Ting; Spelke, Elizabeth; & Snedeker, Jesse. 2013. What exactly do numbers mean? *Language Learning and Development 9*, 105-129.

Jaszczolt, Katarzyna M. 2005. Default semantics: Foundations of a compositional theory of acts of communication. Oxford: Oxford University Press.

Lackey, Jennifer. 2013. Lies and deception: An unhappy divorce. *Analysis 73* (2), 236-284.

Lee, Kang. 2000. Lying as doing deceptive things with words: A speech act theoretical perspective. In Astington, Janet W. (ed.) *Minds in the making: Essays in honor of David R. Olson*. Malden, MA: Blackwell. 177-196.

Lenth, Russell V. 2015. lsmeans: Least-Squares Means. R package version 2.20-2. <http://CRAN.R-project.org/package=lsmeans>.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, Stephen C. 1995. Three levels of meaning. In *Grammar and meaning: Essays in honour of Sir John Lyons*. Cambridge University Press. 90-115.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of the generalized conversational implicature.* Cambridge, MA: MIT Press.

Mahon, James E. 2008. The definition of lying and deception. *Stanford Encyclopedia of Philosophy*.

Meibauer, Jörg. 2005. Lying and falsely implicating. *Journal of Pragmatics 37*, 1373-1399.

Meibauer, Jörg. 2011. On lying: intentionality, implicature, and imprecision. *Intercultural Pragmatics 8* (2), 277-292.

Meibauer, Jörg. 2014. *Lying at the semantics-pragmatics interface*. Berlin: De Gruyter Mouton.

Meibauer, Jörg. 2016a. Topics in the linguistics of lying: A reply to Marta Dynel. *Intercultural Pragmatics 13* (1), 107-123.

Meibauer, Jörg. 2016b. Understanding bald-faced lies: An experimental approach. *International Review of Pragmatics 8* (2), 247-270.

Michaelson, Eliot. 2016. The lying test. *Mind & Language*, *31* (4), 470-499.

Moravcsik, Edith. 1978. 'Reduplicative Constructions', in Joseph H. Greenberg, (ed.), *Universals of Human Language*, Vol. 3: Word Structure, Stanford: Stanford University Press, 297-334.

Papafragou, Anna, and Julien Musolino. 2003. "Scalar implicatures: experiments at the semantics–pragmatics interface." *Cognition 86* (3): 253-282.

Posner, Roland. 1980. Semantics and pragmatics of sentence connectives in natural language. In *Speech act theory and pragmatics,* John R. Searle, Ferenc Kiefer, and Manfred Bierwisch (eds.), 168-203. Dordrecht, Netherlands: Reidel.

Recanati, François. 2001. What-is-said. *Synthese*, 128, 75–91.

Rutschmann, Ronja, and Wiegmann, Alex. 2017. No need for an intention to deceive? Challenging the traditional definition of lying. *Philosophical Psychology* 30 (4), 434-453.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <URL: http://www.R-project.org/>.

Salmani Nodoushan, Mohammed Ali, and Gholamreza Mohiyedin Ghomshei. 2014. Iconicity of cohesion in Persian causative constructions. *Linguistik Online 68* (6), 112-125.

Saul, Jennifer. 2012. *Lying, misleading, and what is said*. Oxford: Oxford University Press.

Shibles, Warren. 1985. *Lying: A critical analysis*. Whitewater, WI: The Language Press.

Singmann, Henrik, Ben Bolker, and Jake Westfall. 2015. afex: Analysis of Factorial Experiments. R package version 0.13-145. URL: http://CRAN.R-project.org/package=afex.

Sorensen, Roy. 2007. Bald-faced lies! Lying without the intent to deceive. *Pacific Philosophical Quarterly* 88, 251-264.

Stokke, Andreas. 2013a. Lying, deceiving, and misleading. *Philosophy Compass* 8, 348-359.

Stokke, Andreas. 2013b. Saying too little and saying too much. Critical Notice of *Lying, Misleading, and What is Said* by Jennifer Saul (2012). *Disputatio* V (35), 81-91.

Stokke, Andreas. 2016. Lying and misleading in discourse. *The Philosophical Review 125* (1). 83-134.

Sweetser, Eve. 1987. The definition of lie: An examination of the folk models underlying a semantic prototype. In Dorothy Holland and Naomi Quinn, eds.*, Cultural models in language and thought*, 43–66. Cambridge: Cambridge University Press.

Terkourafi, Marina. 2010. What-is-said from different points of view. *Language and Linguistics Compass 4* (8). 705-718.

Terkourafi, Marina. Forthcoming. Politeness and lying. In Jörg Meibauer (ed.), *The Oxford Handbook of Lying*. Oxford: Oxford University Press.

Turri, Angelo and John Turri. 2015. The truth about lying. *Cognition 138*, 161-168.

Van Tiel, Bob, Emiel Van Miltenburg, Natalia Zevakhina, and Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33 (1), 137-175.

Vincent, Jocelyne and Cristiano Castelfranchi. 1981. On the art of deception: How to lie while saying the truth. In *Possibilities and limitations of pragmatics. Proceedings of the conference on pragmatics, Urbino, July 8-14*. Herman Parred, Marina Sbisà, and Jef Verschueren (eds.), 749-777. Amsterdam: John Benjamins.

Weissman, Benajmin, and Marina Terkourafi. 2017. False implicatures and the lying/misleading distinction. Presented at The 15th International Pragmatics Conference (IPrA2017), Belfast, July 20, 2017.

Wiegmann, Alex, Jana Samland, and Michael R. Waldmann 2016. Lying despite telling the truth. *Cognition*, *150*, 37-42.

Wilbur, Ronnie B. 2005. A reanalysis of reduplication in American Sign Language. In *Studies on Reduplication*, Bernhard Hurch (ed.), 511-532. Berlin: Mouton De Gruyter.

Williams, Bernard. 2002. *Truth and truthfulness: An essay in genealogy*. Princeton: Princeton University Press.